Метод Ньютона и его приложения к решению уравнений и теории экстремума

Г. Г. Магарил-Ильяев В. М. Тихомиров

При изучении законов природы, в инженерных рассчетах, при решении разнообразных задач управления и экономики, необходимо исследовать и решать *уравнения*, которые описывают изучаемые процессы или явления. Здесь будет рассказано об одном универсальном методе решения уравнений, восходящем к Ньютону (1643 – 1727). Мы проделаем путь, идущий от древних (в частных случаях они пользовались методом, позже описанным Ньютоном) до середины двадцатого века.

1. Истоки: метод решения одного нелинейного уравнения с одним неизвестным

Задача 1.1. *Решить уравнение* $x^2 = 2$.

«А чего его решать, — может удивиться читатель, — его решение это ведь $\pm\sqrt{2}$, не так ли?» Но $\sqrt{2}$ — это просто символ, смысл которого надо еще расшифровать. Если спросить чему равен корень из двух, кто-то может сказать: «Корень из двух равен $1.41\ldots$». Это приближение $\sqrt{2}$ десятичной дробью с точностью до сотых. Еще в шестом веке до н.э. пифагорейцы установили, что $\sqrt{2}$ не дробь, и это стало одним из крупнейших завоеваний древней математики. Решить уравнение $x^2 = 2$ — значит указать способ (говорят еще *алгоритм*) приближения дробями к числу, квадрат которого равен двум, с любой точностью.

Считается, что первый алгоритм для решения уравнения $x^2 = 2$, принадлежит древнегреческому математику Герону, жившему в первом веке нашей эры. Алгоритм Герона описывается так: надо выбрать любую дробь x_0 и затем использовать такую итерационную процедуру $x_k = \frac{1}{2}(x_{k-1} + \frac{2}{x_{k-1}})$, $k = 1, 2\ldots$. Если начать с $x_0 = 1$, то будем последовательно получать $x_1 = \frac{3}{2} = 1.5$, $x_2 = \frac{17}{12} = 1.41(6)$, и если продолжать дальше, то числа x_k будут со все большей и большей точностью стремиться к числу, квадрат которого равен двум.

Когда говорится о решении уравнения с одним неизвестным, то обычно требуется указать целое число или дробь, являющуюся точным решением уравнения, либо, если это не так, то требуется указать алгоритм приближения к искомому числу дробями с любой степенью точности.

В этом пункте мы будем решать уравнения вида $F(x) = y$, где x и y — вещественные числа (совокупность которых обозначают $\mathbb{R}$), а F — функция, определенная, скажем, на некотором интервале (a, b), т. е. F каждому числу $x \in (a, b)$ ставит в соответствие число $F(x)$, что коротко записывают так: $x \mapsto F(x)$, $x \in (a, b)$ или $F : (a, b) \to \mathbb{R}$. Решить уравнение $F(x) = y$ — это значит для данного y найти такое $x \in (a, b)$, при котором уравнение становится верным равенством.

Простейшая функция одного переменного — линейная: $F(x) = Ax$, $A \neq 0$, в нее переменное входит в первой степени. В этом случае решение уравнения $Ax = y$ для любого y, очевидно, дается выражением $x = A^{-1}y$. В уравнение $x^2 = 2$ переменное входит во второй степени. Это уравнение нелинейное. Наша цель — научиться приближенно решать нелинейные уравнения.

Метод решения уравнений $F(x) = y$, о котором сейчас будет рассказано, основан на линеаризации, т. е. на решении по ходу дела линейных уравнений.

Пусть дано число y. Возьмем какое-нибудь число x_0 (так, чтобы $F(x_0)$ было, по возможности, близко к y) и попробуем «добраться» до y линейно. Точнее говоря, выберем отличное от нуля число A и в качестве приближенного решения уравнения $F(x) = y$ возьмем решение линейного уравнения $A(x - x_0) + F(x_0) = y$, которое обозначим x_1 и которое, очевидно, имеет вид $x_1 = x_0 + A^{-1}(y - F(x_0))$. Геометрически (см. рис. 1) x_1 — абсцисса точки пересечения горизонтальной прямой на уровне y с прямой, проходящей через точку $(x_0, F(x_0))$ с угловым коэффициентом A. Теперь вместо x_0 возьмем точку x_1 и поступая аналогично, найдем точку x_2. Продолжая

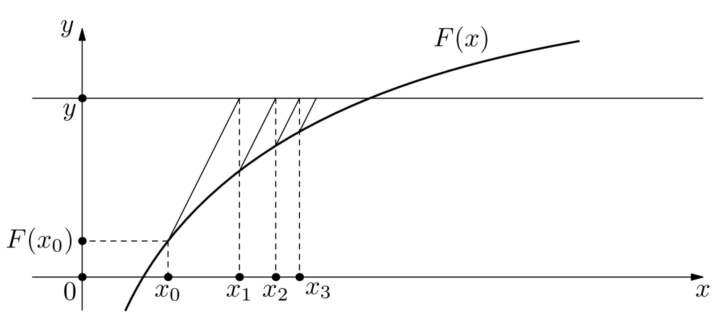

Рис. 1.

этот процесс, получим последовательность

$$x_k = x_{k-1} + A^{-1}(y - F(x_{k-1})), \quad k = 1, 2, \ldots, \tag{1.1}$$

которую назовем *модифицированным методом Ньютона* (решения уравнения $F(x) = y$).

Упражнение 1. *Сделайте три шага модифицированного метода Ньютона* (1.1) *для решения уравнения* $x^2 = 2$, *стартуя от точки* $x_0 = 1$ *и выбрав* $A = 2$. Эти вычисления естественно делать с помощью калькулятора. Сравните полученные результаты с числом $1.414213562\ldots$, представляющим корень из двух с точностью до одной миллиардной (это число появится на калькуляторе, если вы нажмете на $\sqrt{2}$).

Упражнение 2. *Попробуйте определить границы для числа* A, *при которых последовательность* (1.1) *модифицированного метода Ньютона, стартуя от точки* $x_0 = 1$, *будет стремиться к* $\sqrt{2}$.

Исторический комментарий. Секретарь Королевского общества Ольденбург в 1676 году обратился к Ньютону за некоторыми разъяснениями, в частности, касательно решения уравнений. Ньютон в ответном письме от 24.10.1676 г., на примере решения уравнения $F(x) = x^3 - 2x - 5 = 0$, изложил метод нахождения корня, который в современных обозначениях представляет собой такую итерационную процедуру

$$x_k = x_{k-1} - (F'(x_{k-1}))^{-1}F(x_{k-1}), \quad k = 1, 2, \ldots, \tag{1.2}$$

где F' — это производная F, а начальную точку x_0 предоставляется выбрать вычислителю, исходя из каких-то соображений целесообразности. Этот метод (его называют *методом Ньютона*) был изложен Ньютоном в сочинении «Анализ с помощью уравнений ...» (И. Ньютон. Математические работы. — ТТЛ, М. – Л., 1937, с. 9).

Помимо изложенного метода решения уравнений, Ньютон в те годы пользовался также методом приближения функций рядами. Он был не только создателем математического естествознания и одним из родоначальников математического анализа, но и замечательным вычислителем. Он составил для себя таблицы основных, как мы сейчас говорим «элементарных», функций. Он писал Ольденбургу: «Мне стыдно признаться, с какой точностью я проводил свои вычисления».

Производная $F'(x)$ функции $F(x) = x^2$ равняется $2x$, так что для корня из двух метод Ньютона приводит к такой последовательности: $x_k = x_{k-1} + \frac{1}{2x_{k-1}}(2 - (x_{k-1})^2) = \frac{1}{2}(x_{k-1} + \frac{2}{x_{k-1}})$. Именно эта последовательность встречается в сочинениях Герона.

Упражнение 3. *Сделайте три шага метода Ньютона* (1.2) *для решения уравнения* $x^2 = 2$, *стартуя от точки* $x_0 = 1$, *и сравните* *полученный результат с числом* 1.414213562, *дающим, как уже говорилось, приближение* $\sqrt{2}$ *с точностью до одной миллиардной.*

Сформулируем основной результат этого пункта. Отметим, что ниже мы пишем A^{-1} и $|A^{-1}|$ вместо, скажем, $1/A$ и $1/|A|$. Это связано с тем, что при обобщении данного результата A уже не будет числом, но форма записи останется прежней. Кроме того, чтобы сохранить единообразие обозначений при дальнейших обобщениях, интервал с центром в точке $\widehat{x}$ радиуса $\delta > 0$ обозначаем $U_{\mathbb{R}}(\widehat{x}, \delta)$, т.е. $U_{\mathbb{R}}(\widehat{x}, \delta) = \{x \mid |x - \widehat{x}| < \delta\}$.

Теорема 1 (об обратной функции для функций одного переменного). *Пусть заданы числа* x_0, $\delta > 0$, $A \neq 0$ *и функция* $F\colon U_{\mathbb{R}}(x_0, \delta) \to \mathbb{R}$. *Если существует такое число* $\theta \in (0, 1)$, *что для любых* $\xi, x \in U_{\mathbb{R}}(x_0, \delta)$ *выполняется неравенство*

$$|F(\xi) - F(x) - A(\xi - x)| \leqslant \frac{\theta}{|A^{-1}|}|\xi - x|, \tag{1.3}$$

то для каждого $y \in U_{\mathbb{R}}(F(x_0), \delta_0)$, *где* $\delta_0 = \delta(1-\theta)/|A^{-1}|$, *последовательность* (1.1) *сходится к такому числу* $\varphi(y) \in U_{\mathbb{R}}(x_0, \delta)$, *что* $F(\varphi(y)) = y$ *и при этом* $|\varphi(y) - x_0| \leqslant K|y - F(x_0)|$, *где* $K = |A^{-1}|/(1-\theta)$.

Функция $\varphi\colon U_{\mathbb{R}}(F(x_0), \delta_0) \to U_{\mathbb{R}}(x_0, \delta)$, построенная в этой теореме, называется *обратной функцией* к F.

Доказательство опирается на понятия непрерывности функции одного переменного, сходимости последовательности чисел и фундаментальной последовательности чисел. Приведем соответствующие определения.

Пусть функция F определена на интервале U. Говорят, что F *непрерывна в точке* $x \in U$, если для любого числа $\varepsilon > 0$ найдется такое число $\delta > 0$, что $|F(x+h) - F(x)| < \varepsilon$, если только $|h| < \delta$. Функцию F называют *непрерывной на* U, если она непрерывна в каждой точке $x \in U$. Говорят, что числовая последовательность $\{x_1, x_2, \ldots, x_k, \ldots\}$ *сходится к числу* c или *имеет пределом число* c (и пишут $x_k \to c$ при $k \to \infty$ или $\lim\limits_{k\to\infty} x_k = c$), если для любого числа $\varepsilon > 0$ найдется такое число N, что $|x_k - c| < \varepsilon$, как только $k > N$. Последовательность $\{x_1, x_2, \ldots, x_k, \ldots\}$ называется *фундаментальной*, если для любого числа $\varepsilon > 0$ найдется такое число N, что $|x_k - x_m| < \varepsilon$, как только $k, m > N$.

В доказательстве будут использованы следующие факты: a) если функция F определена на интервале $(c-a, c+a)$ и непрерывна в точке c, а последовательность $\{x_1, x_2, \ldots, x_n, \ldots\}$ сходится к c, то последовательность $\{F(x_1), F(x_2), \ldots, F(x_n), \ldots\}$ сходится к числу $F(c)$ (этот факт легко следует из самого определения непрерывности F в точке c); b) всякая фундаментальная последовательность сходится (это одно из выражений свойства полноты вещественных чисел); c) предел суммы последовательностей равен сумме пределов и предел произведения последовательности на число равен произведению этого числа на предел последовательности (если $x_k \to c$, $x'_k \to c'$ при $k \to \infty$ и a — число, то $x_k + x'_k \to c + c'$

при $k \to \infty$, а $ax_k \to ac$; проверка этих свойств совсем проста); d) если $0 < \theta < 1$, то $1 + \theta + \ldots + \theta^k = \dfrac{1-\theta^{k+1}}{1-\theta}$ (формула для суммы геометрической прогрессии).

Доказательство теоремы 1. Пусть $\{x_1, \ldots, x_k\}$ — числа, построенные модифицированным методом Ньютона, и при этом все они принадлежат интервалу $U_{\mathbb{R}}(x_0, \delta)$. По определению x_k удовлетворяет уравнению

$$A(x_k - x_{k-1}) + F(x_{k-1}) = y \tag{1.1$'$}$$

и тогда

$$|x_{k+1} - x_k| = \\ = |A^{-1}(y - F(x_k))| = |A^{-1}||y - F(x_k) - y + F(x_{k-1}) + A(x_k - x_{k-1})| \leqslant \\ \leqslant \theta|x_k - x_{k-1}|. \tag{1.4}$$

Первое равенство следует из (1.1), второе — из (1.1$'$), неравенство — следствие (1.3). Тогда $|x_{k+1} - x_0| \leqslant |x_{k+1} - x_k| + |x_k - x_{k-1}| + \ldots + |x_1 - x_0| \leqslant \leqslant (\theta^k + \theta^{k-1} + \ldots + 1)|x_1 - x_0| < \dfrac{|A^{-1}|}{1-\theta}|y - F(x_0)| < \delta$ (первое неравенство — неравенство для модулей, второе является следствием неравенств типа (1.4) для $s \leqslant k$, третье следует из (1.1) при $k = 1$ и формулы d) для суммы геометрической прогрессии, последнее — следствие определения δ_0). Таким образом, в силу того, что x_0 принадлежит $U_{\mathbb{R}}(x_0, \delta)$ и из того, что x_s принадлежат $U_{\mathbb{R}}(x_0, \delta)$ при $0 \leqslant s \leqslant k$ следует, что x_{k+1} принадлежит $U_{\mathbb{R}}(x_0, \delta)$. Метод математической индукции позволяет сделать вывод, что все x_k, $k = 0, 1, 2, \ldots$, принадлежат $U_{\mathbb{R}}(x_0, \delta)$. Далее, используя те же соображения, что и при доказательстве предыдущего неравенства, получаем для любых k и s

$$|x_{k+s} - x_k| \leqslant \\ \leqslant |x_{k+s} - x_{k+s-1}| + \ldots + |x_{k+1} - x_k| \leqslant (\theta^{k+s-1} + \ldots + \theta^k) < \\ < \frac{|A^{-1}|\theta^k}{1-\theta}|y - F(x_0)|. \tag{1.5}$$

Отсюда следует, что последовательность $\{x_0, x_1, \ldots\}$ фундаментальна, и значит, в силу b) сходится к некоторому числу, которое обозначим $\varphi(y)$. Переходя к пределу по s в соотношении (1.5) при $k = 0$ и учитывая соотношение $|y - F(x_0)| < \delta_0$, получаем, что $\varphi(y) \in U_{\mathbb{R}}(x_0, \delta)$. Переходя (используя свойства предела из c) к пределу в равенстве (1.1) (или (1.1$'$)) и учитывая, что из (1.3) следует непрерывность F в $U_{\mathbb{R}}(x_0, \delta)$ (действительно, в силу (1.3) для любых $\xi, x \in U_{\mathbb{R}}(x_0, \delta)$), будем иметь $|F(\xi) - F(x)| = = |F(\xi) - F(x) - A(\xi - x) + A(\xi - x)| \leqslant ((\theta/|A^{-1}|)|\xi - x| + |A||\xi - x|) = = (\theta/|A^{-1}| + |A|)|\xi - x|)$, получаем, что $F(\varphi(y)) = y$. Снова переходя к пределу по s в неравенстве (1.5) при $k = 0$, приходим к неравенству $|\varphi(y) - x_0| \leqslant K|y - F(x_0)|$, где $K = |A^{-1}|/(1-\theta)$.

2. Метод решения двух уравнений с двумя неизвестными

В этом пункте мы сделаем один шаг от единицы к бесконечности, а именно, будем искать решение системы двух уравнений с двумя неизвестными:

$$\begin{cases} F_1(x_1, x_2) = y_1, \\ F_2(x_1, x_2) = y_2, \end{cases}$$

т. е. для данных чисел y_1 и y_2 требуется найти числа x_1 и x_2, для которых эти уравнения становятся верными равенствами.

Пару чисел (z_1, z_2) будем трактовать как вектор на плоскости (идущий из начала координат в точку с координатами (z_1, z_2)) и записывать его как столбец $\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$. Векторы можно складывать: $\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} + \begin{pmatrix} z_1' \\ z_2' \end{pmatrix} = \begin{pmatrix} z_1 + z_1' \\ z_2 + z_2' \end{pmatrix}$ и умножать на числа: $\alpha \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} \alpha z_1 \\ \alpha z_2 \end{pmatrix}$. Множество всех таких векторов обозначим $\mathbb{R}^2$.

Модулем или *длиной* вектора $z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$ называется число $|z| = \sqrt{z_1^2 + z_2^2}$ (согласно теореме Пифагора, это, действительно, длина вектора). Модуль вектора обладает следующими свойствами: (a) $|x| \geqslant 0$ и $|x| = 0$ тогда и только тогда $x = 0$, (b) $|\alpha x| = |\alpha|\,|x|$ для любого $\alpha \in \mathbb{R}$, (c) $|x+x'| \leqslant |x|+|x'|$ для любых $x, x' \in \mathbb{R}^2$. Первые два свойства очевидны. Последнее следует из известного неравенства Коши – Буняковского для векторов $a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ и $b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$: $(a_1 b_1 + a_2 b_2) \leqslant \sqrt{a_1^2 + a_2^2}\sqrt{b_1^2 + b_2^2}$ (которое сразу получается после возведения в квадрат и раскрытия скобок). Действительно, если обе части этого неравенства умножить на два и добавить слева и справа число $a_1^2 + a_2^2 + b_1^2 + b_2^2$, то придем к соотношению $|a + b|^2 \leqslant (|a| + |b|)^2$, равносильному неравенству треугольника.

Пусть $\widehat{z} \in \mathbb{R}^2$ и $\delta > 0$. Обозначим $U_{\mathbb{R}^2}(\widehat{z}, \delta) = \{z \in \mathbb{R}^2 \mid |z - \widehat{z}| < \delta\}$. Ясно, что это круг без границы с центром в точке $\widehat{z}$ радиуса δ.

Предположим, что функции F_1 и F_2 определены в круге $U_{\mathbb{R}^2}(x_0, \delta)$. Каждому вектору $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in U_{\mathbb{R}^2}(x_0, \delta)$ сопоставим вектор $F(x) = \begin{pmatrix} F_1(x) \\ F_2(x) \end{pmatrix}$ (где $F_i(x) = F_i(x_1, x_2)$, $i = 1, 2$). Тогда, как говорят, F есть отображение из $U_{\mathbb{R}^2}(x_0, \delta)$ в $\mathbb{R}^2$ и кратко пишут $x \mapsto F(x)$, $x \in U_{\mathbb{R}^2}(x_0, \delta)$ или $F\colon U_{\mathbb{R}^2}(x_0, \delta) \to \mathbb{R}^2$. Написанную выше систему двух уравнений теперь можно записать так: $F(x) = y$, где $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$. Эта запись, как мы видим, внешне не отличается от одномерной ситуации.

Важную роль среди систем уравнений занимают так называемые *линейные системы*, т. е. системы вида

$$\begin{cases} a_{11}x_1 + a_{12}x_2 = y_1, \\ a_{21}x_1 + a_{22}x_2 = y_2, \end{cases}$$

где a_{ij}, $1 \leqslant i, j \leqslant 2$ — заданные числа.

Свяжем с этой системой следующую таблицу (или говорят, матрицу)

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

(при этом числа a_{ij}, $1 \leqslant i, j \leqslant 2$ называются элементами матрицы A). Для каждого $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$ положим

$$Ax = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{pmatrix}.$$

Этим определено отображение $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ (которое мы обозначаем той же буквой A), сопоставляющее вектору x вектор Ax. Легко проверить, что для любых $x, x' \in \mathbb{R}^2$ и $\alpha, \beta \in \mathbb{R}$ справедливо равенство $A(\alpha x + \beta x') = \alpha Ax + \beta Ax'$, где αA и βA — матрицы, получаемые из матрицы A умножением всех ее элементов соответственно на α и β. Отображения, обладающие таким свойством, называются *линейными отображениями* или *линейными операторами*.

В терминах матрицы (отображения) A исходная линейная система записывается в виде $Ax = y$, где $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$. Такую запись называют матричной формой записи линейной системы и внешне она не отличается от одномерной ситуации.

Число $\det A = a_{11}a_{22} - a_{12}a_{21}$ называется *определителем* или *детерминантом* матрицы A. Если $\det A \neq 0$, то для любого $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ методом подстановки легко находится решение линейной системы, оно имеет вид

$$\begin{cases} x_1 = \dfrac{a_{11}}{\det A} y_1 - \dfrac{a_{21}}{\det A} y_2, \\ x_2 = -\dfrac{a_{12}}{\det A} y_1 + \dfrac{a_{22}}{\det A} y_2. \end{cases}$$

В матричной форме это можно записать как $x = By$, где B — соответствующая матрица. Ее называют *обратной матрицей к A* и обозначают A^{-1} (и говорят, что матрица A обратима). Итак, если матрица A такова, что

$\det A \neq 0$, то она обратима и решение линейной системы $Ax = y$ для любого y находится по формуле $x = A^{-1}y$ (что, снова, внешне не отличается от одномерного случая).

Наша цель — научиться находить решение системы двух нелинейных уравнений с двумя неизвестными. Для этого, как мы увидим, нужно, фактически, дословно повторить то, что было сказано про одно уравнение с одним неизвестным, заменяя лишь числа x_k векторами, модуль числа — длиной вектора, число A — матрицей A, а число A^{-1} — обратной матрицей к A. Но, помимо этого, в одномерном случае еще участвовало число $|A^{-1}|$. В двумерной ситуации — это «длина» (или обычно говорят *норма*) матрицы A^{-1}. Для каждой матрицы C с элементами c_{ij}, $1 \leqslant i, j \leqslant 2$, ее норма (обозначаемая $\|C\|$) определяется следующим образом: $\|C\| = \sqrt{c_{11}^2 + c_{12}^2 + c_{21}^2 + c_{22}^2}$. Снова, используя неравенство Коши – Буняковского, легко проверить, что $|Cx| \leqslant \|C\||x|$ для любого $x \in \mathbb{R}^2$.

Опишем теперь метод решения двух уравнений с двумя неизвестными, которые мы записали в виде $F(x) = y$. Пусть дан вектор y. Возьмем вектор x_0 (снова, так, чтобы $F(x_0)$ было, по возможности, близко к y) и обратимую матрицу A. В качестве приближенного решения уравнения $F(x) = y$ возьмем решение линейного уравнения $A(x - x_0) + F(x_0) = y$, которое обозначим x_1 и которое, очевидно, имеет вид $x_1 = x_0 + A^{-1}(y - F(x_0))$. Теперь вместо x_0 возьмем точку x_1 и, поступая аналогично, найдем точку x_2. Продолжая этот процесс, получим последовательность

$$x_k = x_{k-1} + A^{-1}(y - F(x_{k-1})), \quad k = 1, 2, \ldots. \tag{2.1}$$

Эту процедуру мы также называем *модифицированным методом Ньютона*.

Сформулируем основной результат этого пункта.

Теорема 2 (об обратном отображении для случая двух переменных). *Пусть заданы вектор x_0, число $\delta > 0$, обратимая матрица A и отображение $F\colon U_{\mathbb{R}^2}(x_0, \delta) \to \mathbb{R}^2$. Если существует такое число $\theta \in (0, 1)$, что для любых $\xi, x \in U_{\mathbb{R}^2}(x_0, \delta)$ выполняется неравенство*

$$|F(\xi) - F(x) - A(\xi - x)| \leqslant \frac{\theta}{\|A^{-1}\|}|\xi - x|, \tag{2.2}$$

то для каждого $y \in U_{\mathbb{R}^2}(F(x_0), \delta_0)$, где $\delta_0 = \delta(1 - \theta)/\|A^{-1}\|$, последовательность (2.1) сходится[1)] к такому вектору $\varphi(y) \in U_{\mathbb{R}^2}(x_0, \delta)$, что $F(\varphi(y)) = y$ и при этом $|\varphi(y) - x_0| \leqslant K|y - F(x_0)|$, где $K = \|A^{-1}\|/(1 - \theta)$.

Отображение $\varphi\colon U_{\mathbb{R}^2}(F(x_0), \delta_0) \to U_{\mathbb{R}^2}(x_0, \delta)$, построенное в этой теореме, называется *обратным отображением* к F.

[1)] Определение сходимости векторов будет дано чуть ниже.

Доказательство теоремы 2 опирается на те же понятия, что и теорема 1, но они касаются уже непрерывности функции не одного, а двух переменных, сходимости последовательности не чисел, а векторов и понятия фундаментальной последовательности не чисел, а векторов. Определения всех этих понятий требуют только замены слов «число» на «вектор из $\mathbb{R}^2$» и «функция» на «отображение». Но, тем не менее, повторим эти определения.

Говорят, что отображение $F\colon U_{\mathbb{R}^2}(x_0,r)\to\mathbb{R}^2$ *непрерывно в точке* $x\in U_{\mathbb{R}^2}(x_0,r)$, если для любого $\varepsilon>0$ найдется такое $\delta>0$, что $|F(x+h)-F(x)|<\varepsilon$, как только $|h|<\delta$. Отображение F называют *непрерывным на* $U_{\mathbb{R}^2}(x_0,r)$, если оно непрерывно в каждой точке $x\in U_{\mathbb{R}^2}(x_0,a)$. Говорят, что последовательность векторов $\{x_1,x_2,\dots,x_k,\dots\}$ из $\mathbb{R}^2$ *сходится к вектору* c или *имеет пределом вектор* c (и пишут $x_k\to c$ при $k\to\infty$ или $\lim\limits_{k\to\infty}x_k=c$), если для любого $\varepsilon>0$ найдется такое натуральное $N=N(\varepsilon)$, что $|x_k-c|<\varepsilon$, как только $k>N$. Последовательность $\{x_1,x_2,\dots,x_k,\dots\}$ векторов из $\mathbb{R}^2$ называется *фундаментальной*, если для любого $\varepsilon>0$ найдется такое натуральное $N=N(\varepsilon)$, что $|x_k-x_m|<\varepsilon$ как только $k,m>N$.

В доказательстве будут использованы аналоги фактов, описанных перед доказательством теоремы 1 (мы нумеруем их в том же порядке): a) если отображение F, определенное на $U_{\mathbb{R}^2}(c,r)$, непрерывно в точке c, то отсюда легко следует, что если последовательность $\{x_1,x_2,\dots,x_k,\dots\}$ векторов из $\mathbb{R}^2$ сходится к c, то последовательность $\{F(x_1),F(x_2),\dots,F(x_k),\dots\}$ сходится к вектору $F(c)$; b) всякая фундаментальная последовательность векторов из $\mathbb{R}^2$ сходится (это, снова, полнота вещественных чисел — сходимость векторов следует из сходимости координат этих векторов); c) предел суммы последовательностей равен сумме пределов и предел произведения последовательности на число равен произведению этого числа на предел последовательности (т. е., если $x_k\to c$ и $x'_k\to c'$ при $k\to\infty$, то $x_k+x'_k\to c+c'$ при $k\to\infty$ и если $x_k\to c$ при $k\to\infty$ и a — число, то $ax_k\to ac$ при $k\to\infty$; все эти свойства легко следуют из определения предела); d) как и в одномерном случае, используется формула для суммы геометрической прогрессии.

Доказательство теоремы 2 отличается от доказательства теоремы 1 только в одном месте. Второе равенство в соотношении (1.4) надо заменить на неравенство $|A^{-1}(y-F(x_k))|\leqslant\|A^{-1}\||y-F(x_k)|$, а далее уже ничего не меняется.

Исторический комментарий. Мы без особых затруднений преодолели дистанцию примерно в двести лет. Теоремы об обратных отображениях для многих переменных стали появляться во второй половине девятнадцатого века. Собственно говоря, все началось с теоремы о неявной функции — решении уравнения $F(x,y)=0$, т. е. нахождении функции $y=\varphi(x)$, для которой $F(x,\varphi(x))=0$. Первая теорема о неявной функции (такого рода теоремы будут нами изучаться в следующем пункте) была доказана Улиссом Дини (1845 – 1918) и прочитана им в курсе анализа в пизанском университете в 1877/78 учебном году. Чуть позже появились теоремы об обратных отображениях.

3. Метод решения m уравнений с n неизвестными

Сделаем еще один шаг и научимся решать систему m уравнений с n неизвестными, т. е. систему уравнений вида

$$\begin{cases} F_1(x_1,\ldots,x_n)=y_1,\\ F_2(x_1,\ldots,x_n)=y_2,\\ \quad\ldots\ldots\\ F_m(x_1,\ldots,x_n)=y_m. \end{cases}$$

Наш метод будет очень схож с тем, что привело нас к цели в одномерном и двумерном случаях.

Здесь мы имеем дело с m функциями n переменных, т. е. с функциями, определенными на подмножествах пространства $\mathbb{R}^n$. Напомним некоторые понятия и договоримся об обозначениях. Пространство $\mathbb{R}^n$ — это совокупность всех наборов из n действительных чисел, расположенных в столбец (называемых векторами или вектор-столбцами), которые, ради экономии места, мы записываем в строку $x=(x_1,\ldots,x_n)^T$, где ^{T} означает *транспонирование* — замену столбца на строчку. Векторы можно (покоординатно) складывать и умножать на числа. Пусть $x=(x_1,\ldots,x_n)^T\in\mathbb{R}^n$. Величина $|x|=\sqrt{x_1^2+\ldots+x_n^2}$ называется *модулем* или *длиной* вектора x. Эта величина обладает теми же свойствами, которые были отмечены в двумерном случае. Пусть $\widehat{x}\in\mathbb{R}^n$ и $\delta>0$. Множество $U_{\mathbb{R}^n}(\widehat{x},\delta)=\{x\mid |x-\widehat{x}|<\delta\}$ называется *открытым шаром* с центром в точке $\widehat{x}$ радиуса δ. Множество в $\mathbb{R}^n$ называется *открытым*, если с каждой своей точкой оно содержит и некоторый шар с центром в этой точке. Любое открытое множество, содержащее данную точку, называется *окрестностью* этой точки.

Предположим, что функции F_i, $1\leqslant i\leqslant m$, определены в шаре $U_{\mathbb{R}^n}(x_0,\delta)$. Тогда определено отображение $F\colon U_{\mathbb{R}^n}(x_0,\delta)\to\mathbb{R}^m$, сопоставляющее $x\in U_{\mathbb{R}^n}(x_0,\delta)$ вектор $(F_1(x),\ldots,F_m(x))^T$. Написанную систему уравнений теперь кратко можно записать так: $F(x)=y$, где $y=(y_1,\ldots,y_m)^T$.

Как и в предыдущих случаях, важную роль играют системы линейных уравнений:

$$\begin{cases} a_{11}x_1+\ldots+a_{1n}x_n=y_1,\\ a_{21}x_1+\ldots+a_{2n}x_n=y_2,\\ \quad\ldots\ldots\\ a_{m1}x_1+\ldots+a_{mn}x_n=y_m. \end{cases}$$

Обозначим через A матрицу этой системы, т. е.

$$A=\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n}\\ a_{21} & a_{22} & \ldots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}.$$

Тогда данная система кратко запишется так: $Ax = y$.[2)] Матрица A задает отображение (которое мы обозначаем той же буквой) $A\colon \mathbb{R}^n \to \mathbb{R}^m$, сопоставляющее вектору $x \in \mathbb{R}^n$ вектор $Ax \in \mathbb{R}^m$. Это отображение линейно (см. двумерный случай), и его называют также *линейным оператором* (из $\mathbb{R}^n$ в $\mathbb{R}^m$).

Если бы матрица A была обратима, то решение уравнения $Ax = y$ для любого y давалось бы формулой: $x = A^{-1}y$, где A^{-1} — обратная матрица к A, и дальнейшие построения, по сути, ничем бы не отличались от рассмотренных случаев. Но это возможно лишь тогда, когда $m = n$, а мы хотим решить систему в более общей ситуации, что важно для различных приложений.

Оказывается, что все рассуждения, связанные с применением модифицированного метода Ньютона, остаются без изменения, если A^{-1} заменить на отображение, которое является правым обратным к A. Дадим точные определения. Говорят, что линейное отображение $A\colon \mathbb{R}^n \to \mathbb{R}^m$ *сюръективно*, если для любого $y \in \mathbb{R}^m$ существует такое $x \in \mathbb{R}^n$, что $Ax = y$. На языке линейных систем это означает, что для любого y линейная система $Ax = y$ имеет решение, а это (как хорошо известно из линейной алгебры) равносильно тому, что ранг матрицы A равен m.

Лемма 1 (о правом обратном отображении в конечномерном случае). *Пусть $A\colon \mathbb{R}^n \to \mathbb{R}^m$ — линейный сюръективный оператор. Тогда существуют отображение $R : \mathbb{R}^m \to \mathbb{R}^n$ (правое обратное к A) и константа $\gamma > 0$ такие, что $AR(y) = y$ и $|R(y)| \leqslant \gamma|y|$ для любого $y \in \mathbb{R}^m$.*

Доказательство Рассмотрим систему векторов $e_1 = (1, 0, \ldots 0)^T, \ldots, e_m = (0, \ldots 0, 1)^T$ в $\mathbb{R}^m$. По условию существуют векторы $f_k \in \mathbb{R}^n$, $1 \leqslant k \leqslant m$, такие, что $Af_k = e_k$. Для каждого $y = \sum_{k=1}^m y_k e_k \in \mathbb{R}^m$ положим $R(y) = \sum_{k=1}^m y_k f_k$. Тогда равенство $AR(y) = y$ следует из линейности оператора A, а второе свойство следует из очевидных оценок (и того, что длина вектора не меньше модуля любой его координаты): $|R(y)| \leqslant \sum_{k=1}^m |y_k||f_k| \leqslant \max_{1\leqslant k\leqslant m} |y_k| \sum_{k=1}^m |f_k| \leqslant \gamma|y|$, где $\gamma = \sum_{k=1}^m |f_k|$.

Из леммы видно, что правых обратных к A «много» и константы γ тоже можно варьировать.

Опишем теперь метод решения системы m уравнений с n неизвестными, записанной в виде $F(x) = y$. Пусть дан вектор y. Возьмем вектор x_0 и пусть $A\colon \mathbb{R}^n \to \mathbb{R}^m$ — сюръективный оператор с правым обратным R. В качестве приближенного решения уравнения $F(x) = y$ возьмем какое-нибудь решение линейного уравнения $A(x - x_0) + F(x_0) = y$. Вектор $x_1 = x_0 + R(y - F(x_0))$, очевидно, будет таким решением (для проверки

[2)]Мы предполагаем, что читатель этого пункта знаком с понятием умножения матриц, в частности, с умножением матрицы на вектор-столбец.

надо применить к обеим частям оператор A). Теперь вместо x_0 берем точку x_1 и, поступая аналогично, найдем точку x_2. Продолжая этот процесс, получим последовательность

$$x_k = x_{k-1} + R(y - F(x_{k-1})), \quad k = 1, 2, \dots\ . \tag{3.1}$$

Эту процедуру мы по-прежнему называем *модифицированным методом Ньютона.*

Теорема 3 (об обратном отображении в конечномерном случае). *Пусть заданы вектор $x_0 \in \mathbb{R}^n$, число $\delta > 0$, сюръективный линейный оператор $A\colon \mathbb{R}^n \to \mathbb{R}^m$ с правым обратным R и отображение $F\colon U_{\mathbb{R}^n}(x_0, \delta) \to \mathbb{R}^m$. Если существует такое число $\theta \in (0, 1)$, что для любых $\xi, x \in U_{\mathbb{R}^n}(x_0, \delta)$ выполняется неравенство*

$$|F(\xi) - F(x) - A(\xi - x)| \leqslant \frac{\theta}{\gamma}|\xi - x|, \tag{3.2}$$

где γ из оценки для R, то для каждого $y \in U_{\mathbb{R}^m}(F(x_0), \delta_0)$, где $\delta_0 = \delta(1-\theta)/\gamma$, последовательность (3.1) *сходится к такому вектору $\varphi(y) \in U_{\mathbb{R}^n}(x_0, \delta)$, что $F(\varphi(y)) = y$ и при этом $|\varphi(y) - x_0| \leqslant K|y - F(x_0)|$, где $K = \gamma/(1 - \theta)$.*

Отображение $\varphi\colon U_{\mathbb{R}^m}(F(x_0), \delta_0) \to U_{\mathbb{R}^n}(x_0, \delta)$, построенное в этой теореме, называется *обратным отображением* к F.

Доказательство фактически ничем не отличается от доказательства теоремы 1, но тем не менее, мы его проведем.

Доказательство Докажем, что a) векторы x_k для всех $k \geqslant 0$ лежат в $U_{\mathbb{R}^n}(x_0, \delta)$ и b) что последовательность $\{x_k\}_{k\geqslant 0}$ фундаментальна. Утверждение a) докажем по индукции. Элемент x_0, очевидно, принадлежит $U_{\mathbb{R}^n}(x_0, \delta)$. Пусть $x_s \in U_{\mathbb{R}^n}(x_0, \delta)$, $1 \leqslant s \leqslant k$. Используя последовательно (3.1), оценку для правого обратного, равенство

$$A(x_k - x_{k-1}) = y - F(x_k) \tag{3.3}$$

(которому удовлетворяет x_k и которое можно получить применяя A к обеим частям (3.1)), (3.2) и затем итерируя процедуру, получим:

$$\begin{aligned}|x_{k+1} - x_k| = |R(y - F(x_k)| \leqslant \gamma|y - F(x_k) - y + F(x_{k-1}) + \Lambda(x_k - x_{k-1})| \leqslant \\ \leqslant \theta|x_k - x_{k-1}| \leqslant \theta^2|x_{k-1} - x_{k-2}| \leqslant \ldots \leqslant \theta^k|x_1 - x_0|.\end{aligned}$$

Применяя теперь неравенство треугольника, предыдущее неравенство, формулу для суммы геометрической прогрессии, (3.1) при $k = 1$ (с оценкой для правого обратного) и то, что $|y - F(x_0)| < \delta_0$, будем иметь

$$\begin{aligned}|x_{k+1} - x_0| \leqslant |x_{k+1} - x_k| + \ldots + |x_1 - x_0| \leqslant \\ \leqslant (\theta^k + \theta^{k-1} + \ldots + 1)|x_1 - x_0| < \frac{\gamma}{1-\theta}|y - F(x_0)| < \delta\end{aligned} \tag{3.4}$$

т. е. $x_{k+1} \in U_{\mathbb{R}^n}(x_0, \delta)$ и значит, все x_k, $k \geqslant 0$, принадлежат $U_{\mathbb{R}^n}(x_0, \delta)$.

Докажем b). Для любых $k, s \in \mathbb{N}$ имеем: $|x_{k+s} - x_k| \leqslant |x_{k+s} - x_{k+s-1}| + \ldots + |x_{k+1} - x_k| \leqslant (\theta^{k+s-1} + \ldots + \theta^k)|x_1 - x_0| \leqslant \dfrac{\gamma\theta^k}{1-\theta}|y - F(x_0)| < \delta\theta^k$, откуда вытекает, что последовательность $\{x_k\}_{n\in\mathbb{N}}$ фундаментальна и тем самым сходится. Обозначим ее предел $\varphi(y)$. Переход к пределу в (3.3) (который существует из-за непрерывности F в $U_{\mathbb{R}^n}(x_0, \delta)$, что доказывается также как в одномерном случае) приводит к равенству $F(\varphi(y)) = y$, а переход к пределу в (3.4) обеспечивает неравенство $|\varphi(y) - x_0| \leqslant K|y - F(x_0)|$ с $K = \gamma/(1-\theta)$.

4. Бесконечномерный случай

В двадцатые годы прошлого века было введено понятие *нормированного пространства*, содержащее в себе многие важнейшие свойства как прямой $\mathbb{R}$, так и пространств $\mathbb{R}^n$.

Множество X называется (вещественным) *векторным пространством*, если для его элементов определена операция сложения и умножения элемента на вещественное число. Эти операции удовлетворяют естественным свойствам, к которым мы привыкли, оперируя, например, с векторами на плоскости: коммутативность и ассоциативность сложения, существование нулевого элемента, противоположного элемента и т. п. Элементы векторного пространства (независимо от их природы) часто называют векторами.

Векторное пространство X называется *нормированным пространством*, если на нем определена функция $\|\cdot\|_X$, сопоставляющая вектору x число $\|x\|_X$, называемое *нормой вектора* x, удовлетворяющее свойствам: 1) $\|x\|_X \geqslant 0$ для любого $x \in X$ и $\|x\|_X = 0$ тогда и только тогда, когда $x = 0$, 2) $\|\alpha x\|_X = |\alpha|\|x\|_X$ для любых $x \in X$, $\alpha \in \mathbb{R}$, 3) $\|x + x'\|_X \leqslant \|x\|_X + \|x'\|_X$ для любых $x, x' \in X$.

Отметим, что модуль числа, длина вектора в $\mathbb{R}^2$ и в $\mathbb{R}^n$ удовлетворяют перечисленным свойствам и тем самым эти пространства являются нормированными.

Пусть $\widehat{x} \in X$ и $r > 0$. Множества $U_X(\widehat{x}, r) = \{x \in X \mid \|x - \widehat{x}\|_X < r\}$ и $B_X(\widehat{x}, r) = \{x \in X \mid \|x - \widehat{x}\|_X \leqslant r\}$ называются соответственно открытым и замкнутым шаром в X с центром в $\widehat{x}$ радиуса r.

Подмножество нормированного пространства называется *открытым*, если с любой своей точкой оно содержит и некоторый шар с центром в этой точке.

Любое открытое множество, содержащее данную точку, называется *окрестностью* этой точки.

Подмножество нормированного пространства называется *замкнутым*, если его дополнение открыто.

Упражнение. Докажите, что открытый (замкнутый) шар является открытым (замкнутым) множеством.

В нормированном пространстве естественным образом определяется понятия сходимости последовательности и непрерывности отображения.

Говорят, что последовательность векторов $\{x_k\}_{k\in\mathbb{N}}$ в X сходится к вектору $\widehat{x}$ и пишут $\lim_{n\to\infty} x_k = \widehat{x}$, если для любого $\varepsilon > 0$ найдется натуральное число $N = N(\varepsilon)$ такое, что неравенство $\|x_k - a\|_X < \varepsilon$ выполняется для всех $k > N$.

Пусть U — открытое подмножество X, Y — другое нормированное пространство и $F\colon U \to Y$. Говорят, что отображение F непрерывно в точке $\widehat{x} \in U$, если для любого $\varepsilon > 0$ найдется такое число $\delta = \delta(\varepsilon) > 0$, что $\|F(\widehat{x} + x) - F(\widehat{x})\|_Y < \varepsilon$ как только $\|x\|_X < \delta$.

Легко проверить, что если отображение F непрерывно в точке $\widehat{x}$, то для любой последовательности $\{x_k\}_{k\in\mathbb{N}}$, сходящейся к $\widehat{x}$, последовательность $\{F(x_k)\}_{k\in\mathbb{N}}$ сходится к $F(\widehat{x})$.

Если F непрерывно в каждой точке U, то говорят, что F непрерывно на U.

Последовательность $\{x_k\}_{k\in\mathbb{N}}$ векторов из X называется *фундаментальной* или *последовательностью Коши*, если для любого $\varepsilon > 0$ найдется такое натуральное число N, что $\|x_{k+s} - x_k\| < \varepsilon$ для всех $k > N$ и $s \in \mathbb{N}$.

Нормированное пространство называется *полным нормированным пространством* или *банаховым пространство*, если в нем каждая фундаментальная последовательность является сходящейся.

Пространство $X = \mathbb{R}^n$ с введенной выше нормой $\|x\|_X = |x| = (\sum_{k=1}^n x_k^2)^{1/2}$, является банаховым пространством. Это есть непосредственное следствие полноты вещественных чисел $\mathbb{R}$. Важное во многих вопросах пространство $C([a,b])$ непрерывных функций на $[a,b]$ с нормой $\|x(\cdot)\|_{C([a,b])} = \max_{t\in[a,b]} |x(t)|$ также является банаховыми пространствами. Для доказательства надо взять фундаментальную последовательность $\{x_k(\cdot)\}$ в $C([a,b])$. Тогда для каждого $t \in [a,b]$ числовая последовательность $\{x_k(t)\}$ будет фундаментальной и значит, сходящейся. Если обозначить через $x(\cdot)$ функцию, для которой $x(t) = \lim_{k\to\infty} x_k(t)$, то уже нетрудно проверить, что $x(\cdot)$ — непрерывная функция и последовательность $\{x_k(\cdot)\}$ сходится к ней в $C([a,b])$ (проделайте это).

Вернемся к вопросу о решении уравнений, но уже в нормированных пространствах, т. е. к вопросу о нахождении решения уравнения $F(x) = y$, где F — отображение из нормированного пространства X в нормированное пространство Y. Как и раньше, важную роль играют линейные уравнения $Ax = y$, где A — линейный оператор из X в Y. Мы сейчас покажем, что вопрос о нахождении решения уравнения $F(x) = y$ совершенно аналогичен конечномерному случаю.

Лемма 2 (о правом обратном отображении в бесконечномерном случае). *Пусть X и Y — банаховы пространства $A\colon X \to Y$ — линейный непрерывный сюръективный оператор*[3]. *Тогда существуют отображение $R\colon Y \to X$ (правое обратное к A) и константа $\gamma > 0$ такие, что $AR(y) = y$ и $\|R(y)\|_X \leqslant \gamma\|y\|_Y$ для всех $y \in Y$.*

Доказательство Эта лемма моментально следует из одного из основных принципов линейного анализа — принципа Банаха об открытости. Действительно, согласно этому принципу (см. учебник Колмогорова и Фомина, гл. IV, п. 5), образ единичного открытого шара в X с центром в нуле при отображении A содержит некоторый открытый шар в Y с центром в нуле. Пусть радиус этого шара равен r. Тогда для любого элемента $y \in Y$, по норме меньшего r, найдется элемент $x(y) \in X$, по норме меньший единицы, такой, что $Ax(y) = y$. Положим $R(0) = 0$ и $R(y) = (2\|y\|_Y/r)x((r/2\|y\|_Y)y)$, если $y \in Y$ и $y \neq 0$. Ясно, что $(r/2\|y\|_Y)y$ принадлежит шару радиуса r с центром в нуле и мы имеем $AR(y) = y$ и $\|R(y)\|_X \leqslant \gamma\|y\|_Y$, где $\gamma = 2/r$.

Теорема 4 (об обратном отображении в бесконечномерном случае). *Пусть X и Y — банаховы пространства и заданы вектор $x_0 \in X$, число $\delta > 0$, сюръективный линейный непрерывный оператор $A\colon X \to Y$ с правым обратным R и отображение $F\colon U_X(x_0,\delta) \to Y$. Если существует такое число $\theta \in (0,1)$, что для любых $\xi, x \in U_X(x_0,\delta)$ выполняется неравенство*

$$\|F(\xi) - F(x) - A(\xi - x)\|_Y \leqslant \frac{\theta}{\gamma}\|\xi - x\|_X, \tag{4.1}$$

где γ из оценки для R, то для каждого $y \in U_Y(F(x_0),\delta_0)$, где $\delta_0 = \delta(1-\theta)/\gamma$, последовательность

$$x_k = x_{k-1} + R(y - F(x_{k-1})), \quad k \in \mathbb{N},$$

сходится к такому вектору $\varphi(y) \in U_X(x_0,\delta)$, что $F(\varphi(y)) = y$ и при этом $\|\varphi(y) - x_0\|_X \leqslant K\|y - F(x_0)\|_Y$, где $K = \gamma/(1-\theta)$.

Отображение $\varphi\colon U_Y(F(x_0),\delta_0) \to U_X(x_0,\delta)$, построенное в этой теореме, называется *обратным отображением* к F.

Отметим, что если оператор A обратим, то обратное отображение φ единственно в следующем смысле: если x такое, что $F(x) = y$, то $x = \varphi(y)$. Действительно, поскольку оператор A обратим, то $R = A^{-1}$ и считаем, что

[3] Это означает, как и в конечномерном случае, что уравнение $Ax = y$ имеет решение для любого $y \in Y$.

$\gamma = \|A^{-1}\|$. Тогда согласно (4.1) имеем

$$\|x - \varphi(y)\|_X = \|\Lambda^{-1}(\Lambda(x - \varphi(y)))\|_X \leqslant \|\Lambda^{-1}\|\|\Lambda(x - \varphi(y))\|_Y = \\ = \|\Lambda^{-1}\|\|F(x) - F(\varphi(y)) - \Lambda(x - \varphi(y))\|_Y \leqslant \theta\|x - \varphi(y)\|_X,$$

т. е. $x = \varphi(y)$.

Доказательство теоремы 4 является дословным повторением доказательства теоремы 3 с заменой $|x|$ при $x \in X$ на $\|x\|_X$, $|y|$ при $y \in Y$ на $\|y\|_Y$ и учетом того, что построенная фундаментальная последовательность $\{x_k\}$ сходится, поскольку X — банахово пространство.

Исторический комментарий. Доказанная нами заключительная теорема принадлежит американскому математику Л. Грейвсу (1896 – 1973). Она была опубликована в 1951 г. Этим мы завершили наш экскурс в вопросы, связанные с решением нелинейных уравнений от Герона, жившего в первом веке нашей эры, до середины прошлого столетия.

Приложения

Доказанные нами теоремы о существовании решений нелинейных уравнений или, иначе говоря, о существовании обратных функций и отображений, играют в математике огромную роль. В следующих двух разделах будет рассказано о двух фундаментальных приложениях этих результатов к доказательству правила множителей Лагранжа и к доказательству теоремы существования и единственности решения задачи Коши дифференциального уравнения. Первая теорема дает ключ к решению задач на максимум и минимум от истоков до нашего времени, вторая вскрывает фундаментальное свойство детерминизма процессов, описываемых дифференциальными уравнениями.

5. Правило множителей Лагранжа.

В первой части этого раздела мы как бы низвергнемся из бесконечности к единице и из университета снова поступаем в школу.[4] Будем учиться решать задачи на максимум и минимум. Термины «максимум» и «минимум» объединяются общим термином «экстремум», и задачи на максимум и минимум называют *экстремальными задачами*.

Экстремальные задачи стали исследовать на заре развития математики. В «Началах» Евклида (третий век до нашей эры) содержится такая

[4] Правда во второй части этого раздела мы поднимимся от единицы к двум и далее к любому n, а в следующем разделе снова вознесемся к бесконечности.

Задача 5.1. *На основании AC треугольника ABC требуется найти такую точку F, что параллелограмм $ADEF$, у которого вершины D и E лежат соответственно на сторонах AB и BC, имел бы наибольшую площадь.*

Переведем эту задачу на формальный язык. Пусть $b = |AC|$ — длина основания, H — высота треугольника ABC, $x = |EF|$, $h(x)$ — высота треугольника BDE. Из подобия треугольников DBE и ABC получаем: $\dfrac{h(x)}{H} = \dfrac{x}{b}$. Площадь параллелограмма равна $\dfrac{H(b-x)x}{b}$. Откуда приходим к такой формулировке: *найти максимум функции* $x \mapsto f(x) = \dfrac{H(b-x)x}{b}$ *на отрезке* $[0, b]$.

Задачи на максимум и минимум встречаются в трудах и других математиков древности, например, у Зенона, Архимеда, Аполлония, Герона. И в новое время (в 16 и 17 веках) было решено множество экстремальных задач. В качестве примера приведем задачу Тартальи (1500 – 1557), который прославился тем, что научился решать уравнения третьей степени.

Задача 5.2. *Разделить число восемь на две такие части, чтобы произведение их произведения на разность было максимальным.*

Вплоть до 17 века каждая экстремальная задача решалась «индивидуально», приемом специально придуманным именно для нее. Первый общий метод нахождения экстремумов функций одного переменного был предложен Ферма (1600 – 1665). Для формулировки теоремы Ферма надо напомнить понятие производной функции одного переменного.

Пусть функция f определена на некотором интервале U. Говорят, что f *дифференцируема в точке* $\widehat{x} \in U$, если существует такое число a, что для всех x, для которых $\widehat{x} + x \in U$ справедливо представление $f(\widehat{x} + x) = f(\widehat{x}) + ax + r(x)$, где $r(x)/x \to 0$ при $x \to 0$. Число a (определяемое этим представлением однозначно) называется *производной функции f в точке* $\widehat{x}$ и обозначается $f'(\widehat{x})$.

Точка $\widehat{x} \in U$ называется *локальным максимумом* (*минимумом*) функции f, если найдется такой интервал $(\widehat{x}-\delta, \widehat{x}+\delta) \subset U$, что для любой точки x из этого интервала выполнено неравенство $f(x) \leqslant f(\widehat{x})$ ($f(x) \geqslant f(\widehat{x})$). Точки локального максимума и локального минимума функции f называются локальными экстремумами этой функции.

Теорема Ферма для функций одного переменного. *Если в точке локального экстремума функция одного переменного дифференцируема, то производная в этой точке равна нулю.*

Действительно, пусть $\widehat{x}$ — локальный экстремум f. Функция f дифференцируема в $\widehat{x}$ и поэтому $f(\widehat{x}+x) = f(\widehat{x}) + f'(\widehat{x})x + r(x) = f(\widehat{x}) + x(f'(\widehat{x}) + r(x)/x)$. Если, скажем, $f'(\widehat{x}) > 0$ и $\widehat{x}$ — локальный максимум, то для всех достаточно малых $x > 0$ будем иметь $f(\widehat{x} + x) > f(\widehat{x})$ в противоречии с тем, что $\widehat{x}$ — локальный максимум. Остальные случаи рассматриваются аналогично.

Задачи 5.1.и 5.2 легко решаются с помощью теоремы Ферма. В задаче 5.1 функция $f(x) = \dfrac{H(b-x)x}{b}$ определена на отрезке $[0, b]$ и непрерывна на нем. По теореме Вейерштрасса эта функция достигает на этом отрезке своего абсолютного максимума, причем достигает его внутри отрезка (поскольку она неотрицательна и на концах отрезка равна нулю), а значит, абсолютный максимум должен совпадать с одним из локальных максимумов. Из теоремы Ферма следует, что локальный максимум $\widehat{x}$ один, причем он равен $\dfrac{b}{2}$. Значит, искомая точка F в задаче Евклида — середина основания AC.

Решить задачу Тартальи предоставим читателю. Для квадрата большего из чисел ответ Тарталья сформулировал в такой форме: *число восемь надо разделить пополам и квадрат этой половины увеличить на одну треть этого квадрата.*

Теорема Ферма дает возможность найти экстремум в задаче без ограничений. Большинство же интересных задач на экстремум — с ограничениями. Приведем два примера.

Задача 5.3. *Найти максимум произведения n положительных чисел, если их сумма равна единице.*

Задача 5.4. *Найти максимум линейной функции на единичном шаре.*

Доказательство неравенства между средним геометрическим и средним арифметическим сводится к решению задачи 5.3, а неравенства Коши – Буняковского — к решению задачи 5.4.

Поставим проблему в общей форме. Пусть на открытом подмножестве $U \subset \mathbb{R}^n$ определены функции f_i, $i = 0, 1, \ldots, m$. *Требуется найти такие точки $x \in U$, которые удовлетворяют условиям $f_i(x) = 0$, $1 \leqslant i \leqslant m$, и в которых $f_0(x)$ принимает свое максимальное (минимальное) значение.* Эту проблему будем записывать как

$$f_0(x) \to \max(\min), \quad f_1(x) = 0, \ldots, f_m(x) = 0. \tag{P}$$

Нас, в основном, будут интересовать локальные максимумы и минимумы. Точка $x \in U$ называется *допустимой в задаче* (P), если $f_i(x) = 0$, $1 \leqslant i \leqslant m$. Точка $\widehat{x} \in U$ называется *локальным максимумом (минимумом) в задаче* (P), если найдется такое $\delta > 0$, что $U_{\mathbb{R}^n}(\widehat{x}, \delta) \subset U$ и для любой допустимой точки $x \in U_{\mathbb{R}^n}(\widehat{x}, \delta)$ выполнено неравенство $f_0(x) \leqslant f_0(\widehat{x})$ $(f_0(x) \geqslant f_0(\widehat{x}))$.

Задачи 5.3 и 5.4 записываются в форме задачи (P) следующим образом:

$$x_1x_2\cdots x_n \to \max, \quad x_1+x_2+\cdots+x_n=1, \quad x_i>0, \quad 1\leqslant i\leqslant n,$$

и

$$y_1x_1+y_2x_2+\cdots+y_nx_n \to \max, \quad x_1^2+x_2^2+\cdots+x_n^2=1.$$

Для того, чтобы сформулировать прием решения задачи (P), описанный Лагранжем, необходимо дать определение дифференцируемости функции многих переменных (а также некоторое усиление этого понятия).

Дифференцируемость и строгая дифференцируемость. Удобно привести соответствующие определения сразу для отображений конечномерных пространств. Пусть U — окрестность точки $\widehat{x}\in\mathbb{R}^n$. Говорят, что отображение $F\colon U\to\mathbb{R}^m$ *дифференцируемо в точке* $\widehat{x}$, если существует такая матрица A размера m на n, что для всех $x\in\mathbb{R}^n$, для которых $\widehat{x}+x\in U$ справедливо представление $F(\widehat{x}+x)=F(\widehat{x})+Ax+r(x)$, где $|r(x)|/|x|\to 0$ при $x\to 0$. Матрица A (определяемая этим представлением однозначно) называется *производной отображения* F *в точке* $\widehat{x}$ и обозначается $F'(\widehat{x})$.

Если $m=1$, т. е. F — функция, то производная есть вектор-строка, элементы которой суть частные производные данной функции по $x_1,\ldots,x_n$ в точке $\widehat{x}$.

Говорят, что отображение $F\colon U\to\mathbb{R}^n$ *строго дифференцируема в точке* $\widehat{x}$, если оно дифференцируемо в этой точке и для любого $\varepsilon>0$ найдется такое $\delta>0$, что для любых $\xi,x\in U_{\mathbb{R}^n}(\widehat{x},\delta)$ справедливо неравенство

$$|F(\xi)-F(x)-F'(\widehat{x})(\xi-x)|\leqslant\varepsilon|\xi-x|. \tag{5.1}$$

Снова, если $m=1$, то мы получаем определение строгой дифференцируемости для функции.

Если функция определена на некотором открытом подмножестве пространства $\mathbb{R}^n$, то аналогично одномерному случаю доказывается

Теорема Ферма для функций многих переменных. *Если в точке локального экстремума функция дифференцируема, то ее производная в этой точке равна нулю.*

А теперь мы готовы предоставить слово самому Лагранжу. В своей книге «Теория аналитических функций» (1797) он писал:

> «Можно высказать следующий общий принцип. Если ищется максимум или минимум некоторой функции многих переменных при условии, что между этими переменными имеется связь, задаваемая одним или несколькими уравнениями, нужно прибавить к функции, о которой говорилось, функции, задающие уравнения связи, умноженные на неопределенные множители, и

> искать затем максимум или минимум построенной суммы, как если бы переменные были независимы. Полученные уравнения, присоединенные к уравнениям связи, послужат для определения всех неизвестных.»

Мы позволим себе лишь незначительно изменить рецепт Лагранжа. Во-первых, оказалось удобным умножить на неопределенный множитель и функцию, экстремум которой ищется, т. е. (имея в виду задачу (P)) составлять выражение $\mathcal{L}(x, \lambda_0, \ldots, \lambda_m) = \sum\limits_{i=0}^{m} \lambda_i f_i(x)$; которое стали называть *функцией Лагранжа* (а числа λ_i — *множителями Лагранжа*). Во-вторых, мы не будем «искать минимум или максимум» функции Лагранжа, а просто будем применять к задаче на экстремум функции Лагранжа необходимое условие экстремума «как если бы переменные были независимы», т. е. теорему Ферма. Сформулируем точный результат.

Правило множителей Лагранжа. *Пусть в задаче (P) все функции f_i, $i = 0, 1, \ldots, m$, строго дифференцируемы в точке $\widehat{x} \in U$. Тогда, если $\widehat{x}$ — локальный экстремум в этой задаче, то существует такие множители Лагранжа $\lambda_0, \lambda_1, \ldots, \lambda_m$, не равные одновременно нулю, что производная функции Лагранжа по x в этой точке равна нулю, т. е.*

$$\lambda_0 f_0'(\widehat{x}) + \lambda_1 f_1'(\widehat{x}) + \ldots + \lambda_m f_m'(\widehat{x}) = 0.$$

Доказательство Теорема утверждает, что если $\widehat{x}$ — локальный экстремум, то векторы $f_i'(\widehat{x})$, $0 \leqslant i \leqslant m$, линейно зависимы. Докажем теорему от противного: предположим, что эти векторы линейно независимы и придем к противоречию с тем, что $\widehat{x}$ — локальный экстремум. Рассмотрим отображение $F\colon U \to \mathbb{R}^{m+1}$, определенное по правилу $F(x) = (f_0(x), f_1(x), \ldots, f_m(x))^T$. По условию функции f_i, $0 \leqslant i \leqslant m$, строго дифференцируемы в $\widehat{x}$. Отсюда легко следует, что и отображение F строго дифференцируемо в $\widehat{x}$. Векторы $f_i'(\widehat{x})$, $i = 0, 1, \ldots, m$, образуют строки матрицы $F'(\widehat{x})$. Следовательно, ранг этой матрицы равен $m + 1$ и значит, $F'(\widehat{x})$ — сюръективный оператор. Согласно лемме 1 существует правый обратный R к $F'(\widehat{x})$. В силу строгой дифференцируемости F в точке $\widehat{x}$, по $\varepsilon = 1/2\gamma$ (где γ из оценки для правого обратного) можно выбрать такое $\delta > 0$, что будет выполняться неравенство (5.1), которое есть в точности неравенство (3.2) в теореме 3 с $\theta = 1/2$ и $A = F'(\widehat{x})$. Согласно этой теореме (применительно к нашему случаю) существует отображение $\varphi\colon U_{\mathbb{R}^{m+1}}(F(\widehat{x}), \delta_0) \to U_{\mathbb{R}^n}(\widehat{x}, \delta)$, где $\delta_0 = \delta/2\gamma$, и константа $K = 2\gamma$ такие, что $F(\varphi(y)) = y$ и $|\varphi(y) - \widehat{x}| \leqslant K|y - F(\widehat{x})|$ для всех $y \in U_{\mathbb{R}^{m+1}}(F(\widehat{x}), \delta_0)$. Так как $F(\widehat{x}) = (f_0(\widehat{x}), 0, \ldots, 0)^T$, то для всех достаточно малых по модулю ν вектор $y_\nu = (f_0(\widehat{x}) + \nu, 0, \ldots, 0)^T$ принадлежит $U_{\mathbb{R}^{m+1}}(F(\widehat{x}), \delta_0)$. Положим $x_\nu = \varphi(y_\nu)$. Тогда $F(x_\nu) = F(\varphi(y_\nu)) = y_\nu$, или

$(f_0(x_\nu), f_1(x_\nu), \ldots, f_m(x_\nu))^T = (f_0(\widehat{x}) + \nu, 0, \ldots, 0)^T$, т. е. $f_0(x_\nu) = f_0(\widehat{x}) + \nu$, $f_i(x_\nu) = 0$, $1 \leqslant i \leqslant m$, и $|x_\nu - \widehat{x}| \leqslant K|y_\nu - F(\widehat{x})| = K|\nu|$. Это означает, что в любой окрестности точки $\widehat{x}$ существует допустимый в задаче (P) вектор x_ν такой, что $f_0(x_\nu) < f_0(\widehat{x})$ $(f_0(x_\nu) > f_0(\widehat{x}))$, если $\nu < 0$ $(\nu > 0)$, в противоречии с тем, что $\widehat{x}$ — локальный экстремум. Таким образом, векторы $f_i'(\widehat{x})$, $0 \leqslant i \leqslant m$, линейно зависимы и теорема доказана.

Решим теперь, используя данную теорему, задачи 5.3 и 5.4.

Рассмотрим задачу 5.3, но с заменой $x_i > 0$ на $x_i \geqslant 0$, $1 \leqslant i \leqslant n$. В этой задаче множество ограничений компактно (т. е. ограничено и замкнуто), максимизируемая функция непрерывна и, следовательно, по теореме Вейрштрасса решение $(\widehat{x}_1, \ldots, \widehat{x}_n)$ такой задачи существует. Очевидно, что $\widehat{x}_i \neq 0$, $1 \leqslant i \leqslant n$. Но тогда эта точка является решением и исходной задачи и, в частности, локальным максимумом в задаче 5.3 с единственным ограничением: $x_1 + \ldots + x_n = 1$. Можно применить правило множителей Лагранжа: существуют такие числа λ_0 и λ_1, не равные одновременно нулю, что производная функции Лагранжа $(x_1, \ldots, x_n) \mapsto \lambda_0 x_1 \cdot \ldots \cdot x_n + \lambda_1(x_1 + \ldots + x_n - 1)$ в этой точке равна нулю: $\lambda_0 \widehat{x}_2 \cdot \ldots \cdot \widehat{x}_n + \lambda_1 = \ldots = \lambda_0 \widehat{x}_1 \cdot \ldots \cdot \widehat{x}_{n-1} + \lambda_1 = 0$. Отметим, что $\lambda_0 \neq 0$, ибо иначе $\lambda_1 = 0$, что невозможно. Отсюда и из выписанных равенств легко следует, что $\widehat{x}_i/\widehat{x}_j = 1$ для любых $i \neq j$, т. е. $\widehat{x}_1 = \ldots = \widehat{x}_n$ и так как $\widehat{x}_1 + \ldots + \widehat{x}_n = 1$, то $\widehat{x}_i = 1/n$, $1 \leqslant i \leqslant n$. Итак, необходимым условиям удовлетворяет единственная точка $(1/n, \ldots, 1/n)$ и значит, она решение исходной задачи 5.3 и значение этой задачи равно $1/n^n$.

Выведем отсюда неравенство для средних. Пусть $x_1, \ldots, x_n$ — произвольные положительные числа. Тогда числа $x_1/\sum_{i=1}^n x_i, \ldots, x_n/\sum_{i=1}^n x_i$ удовлетворяют ограничениям задачи 5.3 и значит, $x_1 \cdot \ldots \cdot x_n/(\sum_{i=1}^n x_i)^n \leqslant 1/n^n$. Извлекая из обеих частей корень n-й степени, получаем неравенство
$$\sqrt[n]{x_1 \cdot \ldots \cdot x_n} \leqslant \frac{x_1 + \ldots + x_n}{n}\,.$$

Перейдем к задаче 5.4. Естественно, мы предполагаем, что не все y_i, $1 \leqslant i \leqslant n$, равны нулю. Решение этой задачи $(\widehat{x}_1, \ldots, \widehat{x}_n)$ существует по теореме Вейрштрасса (множество ограничений компактно, а максимизируемая функция непрерывна). Согласно правилу множителей Лагранжа существуют такие числа λ_0 и λ_1, не равные одновременно нулю, что производная функции Лагранжа $(x_1, \ldots, x_n) \mapsto \lambda_0(y_1 x_1 + \ldots + y_n x_n) + \lambda_1(x_1^2 + \ldots + x_n^2 - 1)$ в точке $(\widehat{x}_1, \ldots, \widehat{x}_n)$ равна нулю: $\lambda_0 y_1 + 2\lambda_1 \widehat{x}_1 = \ldots = \lambda_0 y_n + 2\lambda_1 \widehat{x}_n = 0$. Множитель $\lambda_0 \neq 0$, так как в противном случае $\lambda_1 = 0$ (если бы $\lambda_1 \neq 0$, то $\widehat{x}_1 = \ldots = \widehat{x}_n = 0$ в противоречие с тем, что $\widehat{x}_1 + \ldots + \widehat{x}_n = 1$), что невозможно. Отметим, что и $\lambda_1 \neq 0$, ибо в противном случае мы получили бы, что $y_1 = \ldots = y_n = 0$. Из выписанных равенств следует, что $\widehat{x}_i = (-\lambda_0/2\lambda_1) y_i$, $i = 1, \ldots, n$. Так как $\sum_{i=1}^n y_i \widehat{x}_i = (-\lambda_0/2\lambda_1) \sum_{i=1}^n y_i^2$

и у нас задача на максимум, то необходимо $\lambda_0/2\lambda_1 < 0$. Подставляя $\widehat{x}_i$, $1 \leqslant i \leqslant n$, в ограничения, получаем, что $\lambda_0/2\lambda_1 = -1/\sqrt{\sum_{i=1}^n y_i^2}$. Таким образом, $\widehat{x}_i = y_i/\sqrt{\sum_{j=1}^n y_j^2}$, $1 \leqslant i \leqslant n$, и значение задачи равно $\sqrt{\sum_{j=1}^n y_j^2}$.

Выведем отсюда неравенство неравенство Коши – Буняковского. Пусть $(x_1, \ldots, x_n)$ и $(y_1, \ldots, y_n)$ — произвольные ненулевые наборы чисел. Набор $(x_1/\sqrt{\sum_{i=1}^n x_i^2}, \ldots, x_n/\sqrt{\sum_{i=1}^n x_i^2})$ удовлетворяет ограничениям задачи 5.4 (с данными набором $(y_1, \ldots, y_n)$). Тогда по доказанному

$$\sum_{i=1}^n y_i \left(\frac{x_i}{\sqrt{\sum_{j=1}^n x_j^2}} \right) \leqslant \sqrt{\sum_{j=1}^n y_j^2}.$$

Откуда сразу следует, что

$$\sum_{i=1}^n y_i x_i \leqslant \left(\sum_{i=1}^n y_i^2 \right)^{1/2} \left(\sum_{i=1}^n x_i^2 \right)^{1/2}.$$

Отметим, что подобным образом можно доказывается огромное число точных неравенств, возникающих в анализе и геометрии.

Упражнение. Докажите неравенство Гёльдера

$$\sum_{i=1}^n y_i x_i \leqslant \left(\sum_{i=1}^n |y_i|^p \right)^{1/p} \left(\sum_{i=1}^n |x_i|^q \right)^{1/q},$$

где $1 < p, q < \infty$ и $1/p + 1/q = 1$.

Задача (P) и правило множителей Лагранжа естественным образом обобщаются на бесконечномерный случай. Пусть X и Y — нормированные пространства, U — открытое подмножество X, $f_0 \colon U \to \mathbb{R}$ и $F \colon U \to Y$. Рассмотрим задачу

$$f_0(x) \to \max(\min), \quad F(x) = 0.$$

Аналогично предыдущему определяется понятие локального экстремума. Теорема 4 (вместе с теоремой отделимости для выпуклых множеств) позволяет доказать аналог правила множителе Лагранжа для данного случая (сам этот факт был установлен Л. А. Люстерником в 1934 году). Мы не будем на этом останавливаться, но отметим, что из этой теоремы, в частности, вытекают необходимые условия экстремума в различных задачах классического вариационного исчисления.

6. Существование решений дифференциальных уравнений

Пусть $(t, x) \to f(t, x)$ — функция $(n+1)$-го переменного ($t \in \mathbb{R}$, $x \in \mathbb{R}^n$) со значениями в $\mathbb{R}^n$, непрерывная в окрестности точки $(t_0, x_0) \in \mathbb{R} \times \mathbb{R}^n$. Если непрерывно дифференцируемая функция $\widehat{x}(\cdot)$ со значениями в $\mathbb{R}^n$, определенная на некотором отрезке $\Delta = [t_0 - \alpha, t_0 + \alpha]$ такова, что $\dot{\widehat{x}}(t) = f(t, \widehat{x}(t))$ для всех $t \in \Delta$ и $\widehat{x}(t_0) = x_0$, то говорят, что $\widehat{x}(\cdot)$ — решение задачи задачи Коши

$$\dot{x} = f(t, x), \quad x(t_0) = x_0. \tag{6.1}$$

на отрезке Δ.

Следующий результат есть непосредственное следствие теоремы 4 об обратном отображении.

Выше было определено банахово пространство непрерывных функций $C([a, b])$. Нам понадобится аналогичное пространство, но для вектор-функций, т. е. функций со значениями в $\mathbb{R}^n$. Совокупность всех непрерывных на отрезке Δ функций $x(\cdot)$ со значениями в $\mathbb{R}^n$ обозначим через $C(\Delta, \mathbb{R}^n)$. Это банахово пространство с нормой $\|x(\cdot)\|_{C(\Delta,\mathbb{R}^n)} = \max_{t\in\Delta} |x(t)|$, где $x(t) = (x_1(t), \dots, x_n(t))^T$.

Локальная теорема существования и единственности. Пусть $t_0 \in \mathbb{R}$, $x_0 \in \mathbb{R}^n$, $a > 0$, $b > 0$, $D = [t_0 - a, t_0 + a] \times B_{\mathbb{R}^n}(x_0, b)$, функция $f\colon D \to \mathbb{R}^n$ непрерывна на D и удовлетворяет на этом множестве условию Липшица по x, т. е. существует такая константа $L > 0$, что для всех $(t, x_i) \in D$, $i = 1, 2$, справедливо неравенство

$$|f(t, x_1) - f(t, x_2)| \leqslant L|x_1 - x_2|. \tag{6.2}$$

Тогда найдется такое число $0 < \alpha \leqslant a$, что на отрезке $\Delta = [t_0 - \alpha, t_0 + \alpha]$ определено единственное решение $\widehat{x}(\cdot)$ задачи Коши (6.1) и при этом

$$\max_{t\in\Delta} |\widehat{x}(t) - x_0| \leqslant 2 \max_{t\in\Delta} \left| \int_{t_0}^{t} f(\tau, x_0)\, d\tau \right|.$$

Доказательство Пусть $M = M(x_0) = \max_{t\in[t_0-a,t_0+a]} |f(t, x_0)|$ и $\alpha = \min(a, 1/2L, b/2M)$. Применим теорему 4, обозначив $X = Y = C(\Delta, \mathbb{R}^n)$, $x_0 = x_0(\cdot)$ (функция, равная тождественно x_0 на Δ), $\delta = b$, $F\colon U_X(x_0(\cdot), b) \to Y$ задается формулой ($t \in \Delta$)

$$F(x(\cdot))(t) = x(t) - x_0 - \int_{t_0}^{t} f(\tau, x(\tau))\, d\tau, \tag{i}$$

A — тождественный оператор ($\gamma = \|A^{-1}\| = 1$) и $\theta = 1/2$ (тем самым $\delta_0 = b/2$).

Для всех $\xi(\cdot), x(\cdot) \in U_X(x_0(\cdot), b)$ и $t \in \Delta$ имеем, используя последовательно (i), (6.2) и определение Δ:

$$|F(\xi(\cdot))(t) - F(x(\cdot))(t) - (\xi(t) - x(t))| = \left|\int_{t_0}^{t} (f(\tau, \xi(\tau)) - f(\tau, x(\tau)))\, d\tau\right| \leqslant$$

$$\leqslant |t - t_0| L \|\xi(\cdot) - x(\cdot)\|_{C(\Delta, \mathbb{R}^n)} \leqslant \frac{1}{2} \|\xi(\cdot) - x(\cdot)\|_{C(\Delta, \mathbb{R}^n)},$$

т. е. справедливо неравенство (4.1). Далее, снова, в силу (i) и определения Δ

$$|F(x_0(\cdot))(t)| = \left|\int_{t_0}^{t} f(\tau, x_0)\, d\tau\right| \leqslant |t - t_0| M \leqslant \frac{b}{2}\,.$$

Это значит, что $0 \in U_X(F(x_0(\cdot))(\cdot), b/2)$ и тем самым существует функция $\widehat{x}(\cdot) = \varphi(0)$, удовлетворяющая условию: $F(\widehat{x}(\cdot))(\cdot) = 0$, т. е.

$$\widehat{x}(t) - x_0 - \int_{t_0}^{t} f(\tau, \widehat{x}(\tau))\, d\tau = 0, \quad \text{для любого } t \in \Delta.$$

Таким образом, $\widehat{x}(\cdot)$ — решение задачи Коши (6.1) на отрезке Δ и оно единственно в силу замечания после формулировки теоремы 4.

Последнее утверждение сразу следует из соответствующей оценки в теореме.

Отметим, что глобальная теорема существования и единственности для линейной системы уравнений также есть непосредственное следствие теоремы 4.

В непосредственной близости от тех вопросов, которым посвящена данная работа, лежит множество других фундаментальных результатов алгебры, геометрии, анализа, теории дифференциальных уравнений, которые являются фактически следствиями доказанных утверждений. Таковы, например, критерий Сильвестра, приведение симметрических матриц к каноническому виду (и тем самым существенная часть аналитической геометрии), вопросы непрерывной и дифференцируемой зависимости решений дифференциальных уравнений от начальных данных и параметров, представление функций рядами Фурье, решение задач математической физики методом Фурье и многое другое.

Г. Г. Магарил-Ильяев, Московский государственный институт радиотехники, электроники и автоматики (технический университет)
В. М. Тихомиров, механико-математический факультет МГУ