

СХОДИМОСТЬ МЕТОДА НЬЮТОНА НА РАЗЛИЧНЫХ КЛАССАХ ФУНКЦИЙ

С. Е. МИХЕЕВ

Санкт-Петербургский государственный университет, Россия

e-mail miheev@apmath.spbu.ru

Convergences of the iterative Newton's method for systems of nonlinear equations $g(x) = 0$ with functions g having Lipschitz' constant L for its derivation: $g \in C^{1,1}$ versus functions g having local Lipschitz' constant $L(x)$ and estimation $L(x)\|(g'(x))^{-1}\|_2 \leq \sigma$ are compared. For the second class of functions the results similar to the Kantorovich (TK) theorems for $C^{1,1}(D)$ are obtained. It is shown that the second class has elements and initial approximations which do not guarantee convergence while the TK do. Vice versa, the class $C^{1,1}$ has elements and initial approximations that guarantee the convergence according to the TK, but they do not satisfy the conditions of the theorems presented here.

Введение

Обоснование метода Ньютона решения нелинейного уравнения

$$g(x) = \mathbf{0}, \quad x \in D \subset U, \quad g : D \longrightarrow W, \quad (1)$$

где U, W — банаховы пространства (B -пространства), было проведено в работах Л. В. Канторовича и И. П. Мысовских в середине 20-го века [1–9]. Во всех указанных трудах ограничение в некоторой области (открытого множества) D из B -пространства на скорость изменения производной $J = g'$ имело вид условия

$$\sup_{x \in D} \|J^{-1}(x_0)J'(x)\| \leq K, \quad (2)$$

где x_0 — начальная точка метода, либо $\|J'(x)\| \leq L$ для всех x из D . Во втором варианте непосредственно из доказательств в указанных трудах была видна возможность легкого расширения результатов и на случай, когда J всего лишь липшицева:

$$\|J(x + \Delta) - J(x)\| \leq L\|\Delta\| \quad \forall x, x + \Delta \in D. \quad (3)$$

Конкретные D, L определяют некоторое множество, которое здесь как обычно обозначается $C^{1,1}(D, L)$, при ненадобности список параметров (D, L) будет опускаться.

Здесь под нормой линейного оператора $A : U \rightarrow W$ подразумевается норма, подчиненная нормам пространств U и W : $\|A\| = \sup_{x \neq 0} \|Ax\|_W / \|x\|_U$.

И.П. Мысовских рассмотрел сужение множества $C^{1,1}(D, L)$, порождаемое ограничением $\|J^{-1}(x)\| \leq r$ для всех x из D и установил в нем сходимость метода Ньютона при более слабых условиях.

Для липшицевых J основная теорема о методе Ньютона, именуемая теоремой Канторовича, доказана новым способом в [10]. В [11] теорема Канторовича (ТК) приводится также для липшицевых J с утверждением, что доказательство следует оригиналу (однако во всех описанных там оригиналах использована J').

Сравнение приводимых здесь результатов будет произведено с теоремой Канторовича, применяемой к функциям из $C^{1,1}$.

Параметры теорем Канторовича и Мысовских не улучшаемы в $C^{1,1}$. Однако это не запрещает получать в иных классах результаты, не объемлющие их и не являющиеся их следствиями. Один из таких классов и будет здесь исследован на применимость метода Ньютона. Основным инструментом для этого послужит метод дифференцирования по итерации, суть которого состоит в дифференцировании уравнения, порождающего неявным образом следующую итеративную точку, по смещениям текущей итерации.

В обычном методе Ньютона порождающее уравнение является линейной аппроксимацией задачи (1):

$$g(x_k) + J(x_k)(x - x_k) = \mathbf{0}, \quad k = 0, 1, \dots, \quad (4)$$

где x_k — текущая итерация; решение системы (4) относительно x есть следующая итерация. (Заметим, что традиционная запись метода Ньютона в виде $x_{k+1} = x_k - J^{-1}(x_k)g(x_k)$ практически никогда не используется как основа расчетных формул и именно непосредственное решение уравнения (4) программно реализуется на каждом шаге.)

В конечномерном пространстве B можно после выбора базиса трактовать производную функции g как матрицу Якоби J (отсюда и обозначение).

Замкнутый шар $\{y \mid \|y - x\| \leq \delta\}$ будем обозначать здесь через $S(x, \delta)$.

1. Существование и оценка удаленности решения

Пусть U, W — B -пространства.

Определение 1. Назовем локальной константой Липшица функции $G : M \rightarrow W$ в x — изолированной точке множества $M \subset U$ — величину

$$L_G(x) = \overline{\lim}_{\Delta \rightarrow 0} \|G(x + \Delta) - G(x)\|_W / \|\Delta\|_U.$$

Здесь предел берется по таким Δ , что $x + \Delta$ принадлежит M .

Там, где возникнет потребность усилить внимание к тому, что речь идет об обычной константе Липшица, снабдим ее эпитетом “глобальная”.

Лемма 1. Пусть Z, U, W — B -пространства и множество M лежит в Z . Если определенное на M семейство линейных операторов $A(y) : U \rightarrow W$, $y \in M$, в x — изолированной точке множества M — имеет конечную локальную константу Липшица L и оператор $A(x)$ невырожден, то существует $\delta > 0$, такое, что для всех y из окрестности $V = S(x, \delta) \cap M$ операторы $A(y)$ невырождены, а семейство обратных операторов $A^{-1}(y) : W \rightarrow U$, $y \in V$, равномерно ограничено, в точке x тоже имеет локальную константу Липшица L^- и

$$L^- \leq \|A^{-1}(x)\|^2 L.$$

Доказательство. Из существования локальной константы в x следует непрерывность семейства в x . Поэтому согласно теореме Банаха и следствий из нее [8, теорема 5.4.4] существует для точки x число $\delta > 0$, такое, что в окрестности $V = S(x, \delta) \cap M$ существуют обратные операторы A^{-1} , которые равномерно ограничены:

$$\|A^{-1}(y)\| \leq \frac{\|A^{-1}(x)\|}{(1 - \|A^{-1}(x)\|\|A(y) - A(x)\|)} \leq 2\|A^{-1}(x)\| \quad \forall y \in V.$$

Пусть $x + \Delta$ принадлежит окрестности V , тогда

$$A^{-1}(x + \Delta) - A^{-1}(x) = A^{-1}(x)[A(x) - A(x + \Delta)]A^{-1}(x + \Delta).$$

Откуда

$$\|A^{-1}(x + \Delta) - A^{-1}(x)\| \leq L\|A^{-1}(x)\|\|\Delta\|\|A^{-1}(x + \Delta)\| + o(\Delta). \quad (5)$$

В силу ограниченности $\|A^{-1}(x + \Delta)\|$ из (5) следует непрерывность A^{-1} в x .

Поэтому, разделив на $\|\Delta\|$ неравенство (5) и переходя к верхнему пределу при $\Delta \rightarrow 0$, получаем утверждение леммы. \square

Если множество M — область задания отображения G — не имеет изолированных точек, то локальная константа Липшица, конечная или бесконечная, определена на всем M .

Определение 2. Пусть M не имеет изолированных точек. Функцию $L_G : M \rightarrow \mathbb{R}^1$, значение которой в каждой точке области задания есть локальная константа Липшица отображения G , назовем полупроизводной отображения G .

Отметим, что если отображение G дифференцируемо в x из D по Фреше, его полупроизводная есть $\|G'(x)\|$.

Определение 3. Обозначим через ЛЛ-класс (D, σ) класс функций g , удовлетворяющих следующим условиям:

1) линейный оператор J — производная функции g — в области D невырожден и имеет конечную полупроизводную L_J ;

2) для всех x из D выполняется $L_J(x)r(x) \leq \sigma$, где $r(x) \geq \|J^{-1}(x)\|$ и $\sigma > 0$.

В следующей теореме будет использован частный случай B -пространств — гильбертовы пространства, наделенные естественной нормой $\|x\| = \sqrt{(x, x)}$. Далее — H -пространства.

Теорема 1. Пусть U, W — H -пространства, $D \subset U$, $g : D \rightarrow W$, g принадлежит ЛЛ-классу (D, σ) и выполняются следующие условия:

1) $P_\sigma = r_0\sigma\|g(x_0)\| \leq 1 - \varepsilon$, где $r_0 = r(x_0)$ и ε принадлежит $(0, 1)$;

2) шар $S(x_0, d_0)$ принадлежит области D , где

$$d_0 = -\frac{\ln(1 - P_\sigma)}{\sigma} \leq -\frac{\ln \varepsilon}{\sigma}.$$

Тогда в шаре $S(x_0, d_0)$ существует решение α системы (1).

Доказательство. От противного. Пусть в шаре $S(x_0, d_0)$ нет решений системы (1). Тогда в силу непрерывности g существует $b > 0$, такое, что $\|g(x)\| > b$ для всех x из $S(x_0, d_0)$, кроме того, следующая задача Коши

$$\dot{x} = -J^*(x)g(x)/\|J^*(x)g(x)\|, \quad x(0) = x_0 \quad (6)$$

имеет в шаре $S(x_0, d_0)$ непрерывную правую часть системы дифференциальных уравнений. Здесь и далее $*$ — операция сопряжения.

Согласно [12, следствию 2.3.2] либо решение задачи (6) существует для всех $t \geq 0$, либо оно непродолжимо по t далее момента $T < +\infty$, и тогда $t \rightarrow T$ влечет $\|x(t) - x_0\| \rightarrow d_0$.

Рассмотрим поведение функции $g^2(x) = (g(x), g(x))$ на решении задачи (6). По построению

$$\begin{aligned} (g^2(x))'_t &= J(x)\dot{x} \cdot g(x) + g(x) \cdot J(x)\dot{x} = \\ &= -(J(x)J^*(x)g(x) \cdot g(x) + g(x) \cdot J(x)J^*(x)g(x)) / \|J^*(x)g(x)\| = -2\|J^*(x)g(x)\|. \end{aligned}$$

Оценим правую часть. Пусть $J^*(x)g(x) = y$. Тогда $g(x) = (J^*)^{-1}(x)y$ и

$$\|g(x)\| \leq \|(J^*)^{-1}(x)\| \|y\| = \|J^{-1}(x)\| \|y\| \implies \|g(x)\|/r \leq \|y\| = \|J^*(x)g(x)\|.$$

Следовательно,

$$(g^2)'_t \leq -2\|g\|/r. \quad (7)$$

Оценим приращение функции $\rho(t) \stackrel{\text{def}}{=} r(x(t))$ при изменении времени на $\delta > 0$. Как известно [13], норма интеграла от непрерывной абстрактной функции не более интеграла от нормы этой функции. А производная решения задачи (6) $\dot{x}(t)$ непрерывна. Поэтому

$$\|x(t + \delta) - x(t)\| = \left\| \int_0^\delta \dot{x}(t + \tau) d\tau \right\| \leq \int_0^\delta \|\dot{x}(t + \tau)\| d\tau.$$

Помимо этого из (6) видно, что $\|\dot{x}\| = 1$, следовательно, для всех t, δ , таких, что $x(t), x(t + \delta)$ из $S(x_0, d_0)$, справедливо

$$\begin{aligned} |\rho(t + \delta) - \rho(t)| &\leq \|J^{-1}(x(t + \delta)) - J^{-1}(x(t))\| \stackrel{\text{Лемма 1}}{\leq} \\ &\leq \|J^{-1}(x(t))\|^2 L_J(x(t)) \|x(t + \delta) - x(t)\| + o_1(\|x(t + \delta) - x(t)\|) \leq \\ &\leq \rho^2(t) L_J(x(t)) \int_0^\delta \|\dot{x}(t + \tau)\| d\tau + o_2(\delta) = \rho^2(t) L_J(x(t)) \delta + o_2(\delta) \leq \sigma \rho(t) \delta + o_2(\delta). \end{aligned} \quad (8)$$

Здесь o_i — бесконечно малые функции своих аргументов для любого индекса i . Таким образом, правые производные числа функции ρ в момент t ограничены по модулю величиной $\sigma \rho(t)$. Тогда согласно теореме 3.4.1 и замечаниям к ней в [12] функция $\rho(t)$ мажорируется решением задачи Коши

$$\dot{\eta} = \sigma \eta, \quad \eta(0) = r_0, \quad (9)$$

т. е. $\rho(t) \leq r_0 e^{\sigma t}$, $t \in [0, T)$.

Усилив (7) подстановкой вместо $r(x(t))$ функции $r_0 e^{\sigma t}$, имеем

$$(\|g\|^2)'_t \leq -2\|g\|e^{-\sigma t}/r_0 \iff (\|g\|)'_t \leq -e^{-\sigma t}/r_0.$$

Интегрируем последнее неравенство с начальным условием $g^2|_{t=0} = g^2(x_0)$:

$$\|g(x(t))\| \leq \|g(x_0)\| + (e^{-\sigma t} - 1) / (r_0 \sigma). \quad (10)$$

Его правая часть обнуляется в момент

$$d_0 \stackrel{\text{def}}{=} -\sigma^{-1} \ln(1 - r_0 \sigma \|g(x_0)\|),$$

а решение задачи Коши (6), как уже показано, определено при t из $[0, T)$. Оценим снизу T . Поскольку по построению $\|\dot{x}\| = 1$, для приращения переменной x справедлива оценка $\|\Delta x\| \leq \Delta t$. Значит, $T \geq d_0$, и решение задачи Коши (6) определено при t из $[0, d_0)$ и принадлежит $S(x_0, d_0)$.

Следовательно, величина $\|g(x(t))\|$ определена на $[0, d_0)$ и согласно (10) должна либо обратиться в нуль в момент $\tau < d_0$, либо стремиться к нулю при $t \rightarrow d_0$. Но и то и другое невозможно, ибо она по построению больше b ($b > 0$). (Положительность аргумента под логарифмом гарантируется условием 1.)

Получено противоречие, которое завершает доказательство теоремы. \square

Замечание 1. Условия 1 и 2 теоремы 1 точны в том смысле, что при их нарушении можно указать функцию из ЛЛ-класса (D, σ) , не имеющую корней в D .

Действительно, рассмотрим скалярную функцию вида

$$g(t) = g_0 + (e^{-\sigma t} - 1) / (r_0 \sigma), \quad (11)$$

где t из D , $g_0, r_0 > 0$. Для нее $|\ddot{g}(t)/\dot{g}(t)| = \sigma$ для всех t из D , т.е. g принадлежит ЛЛ-классу (D, σ) . Поскольку $e^{-\sigma t}$ может принимать все значения из $[0, 1)$ и не может быть отрицательной или 0, наличие корня у функции g при каком угодно D влечет условие $g(0) - 1/r_0/\sigma < 0$, эквивалентное $P_\sigma < 1$. Это означает, что условие 1 теоремы 1 не может быть ослаблено. С другой стороны, корень функции g , расширенной на все R^1 согласно формуле (11), есть $d_0 = -(\sigma)^{-1} \ln(1 - P_\sigma)$, и нарушение условия 2 теоремы 1, такое, что $d_0 \notin D$, означало бы, что в D нет корней.

Замечание 2. Используя технику доказательства теоремы 1, можно получить для функций из $C^{1,1}(D, L)$ (т.е. с матрицей Якоби, обладающей обычной константой Липшица L в D) результат, являющийся частью одного из вариантов теоремы Канторовича, относящейся к оценке удаленности решения от начальной точки.

Для этого отметим, что из $L_J(x(t)) \leq L$ и (8) следует оценка

$$|\rho(t + \delta) - \rho(t)| \leq L\rho^2(t)|\delta| + o_4(\delta).$$

Согласно [12, теореме 3.4.1], составляем мажорирующую задачу Коши для нее: $\dot{\eta} = L\eta^2$, $\eta(0) = r_0$. Интегрируя, получаем $\rho(t) \leq r_0/(1 - Ltr_0)$. Усилив (7) подстановкой вместо $r(x(t))$ функции $\frac{r_0}{1 - Ltr_0}$, имеем $(\|g\|)'_t \leq Lt - 1/r_0$, что после интегрирования даст

$$\|g(x(t))\| - \|g(x_0)\| \leq \int_0^t (L\tau - 1/r_0) d\tau = Lt^2/2 - t/r_0. \quad (12)$$

Следовательно, $g(x(t))$ обнулится не позднее, чем правая часть (12) станет равной $-\|g(x_0)\|$, т.е. не позднее момента $d_1 = 1 - \sqrt{1 - 2r_0^2\|g(x_0)\|}/Lr_0$. Поэтому удаленность решения от начальной точки не более d_1 , что совпадает с оценкой замечания 1 к теореме 6 из [8], гл. XVIII, § 1.

Замечание 3. Открытость области задания D в доказательстве теоремы 1 потребовалась лишь для того, чтобы иметь производную J и ее полупроизводную во всех точках

шара $S(x_0, d_0)$. Вместо открытости было бы достаточно потребовать непосредственно их существования в этом шаре либо их существования в шаре максимального радиуса с центром в x_0 , который можно поместить в область задания. \square

Аналогичный характерному параметру P_σ параметр P_K теоремы Канторовича имеет вид $P_K = K \|J^{-1}(x_0)g(x_0)\|$, где K из (2). Когда J всего лишь Липшицева с константой L , в одном из вариантов теоремы Канторовича этот параметр принимает вид $P_L = r_0^2 L \|g(x_0)\|$ [11, теорема 11.3]. Оценка удаленности решения от начальной точки и сходимость к нему из этой начальной точки ньютоновских итераций гарантируются теоремой Канторовича при одном и том же условии: P_L (или P_K) $\leq 1/2$, причем это требование не чрезмерно, ибо существуют функции g , сколь угодно мало нарушающие это условие и не имеющие корня.

То, что правая часть в условии $P_\sigma < 1 - \varepsilon$ может быть близка к 1, не означает, что теорема 1 сильнее теоремы Канторовича, поскольку она применяется к функциям из ЛЛ-класса (D, σ) , не совпадающим с $C^{1,1}(D, L)$ ни при каких значениях параметров σ, L .

Соотношение между теоремой 1 и результатом из теоремы Канторовича, касающимся удаленности решения от начальной точки, хорошо видно в одномерном пространстве.

Пример 1, показывающий, что в $C^{1,1}(D, L)$ теорема 1 не лучше, чем теорема Канторовича.

Помимо параметров D, L теорема Канторовича использует информацию $f(x_0) = f_0$ и $J(x_0) = J_0$. Таким образом, ТК представляет собой суждения о классе функций $C^{1,1}(D, L, f_0, J_0)$, одно из которых гласит, что если $1/2 \geq P_L \equiv L|f_0|/\|J_0^{-1}\|^2$ и D достаточно велико, то каждая функция f из этого класса имеет корни.

Сравним это с тем, что дает применение теоремы 1 к $C^{1,1}(D, L, f_0, J_0)$ в одномерном пространстве: $D \subset R^1$. Скалярный аргумент будем обозначать через t и положим $f(0) = g_0, \dot{f}(0) = \dot{g}_0, \mathcal{G}_1 = C^{1,1}(D, L, g_0, \dot{g}_0)$ (все далее сказанное очевидным образом распространяется и на условия, задаваемые в произвольный момент t_0).

Относительно \mathcal{G}_1 ТК утверждает, что если $1/2 \geq P_L \equiv L|g_0|/\dot{g}_0^2$ и интервал задания D достаточно велик, то каждая функция f из \mathcal{G}_1 имеет корни, среди которых существует единственный корень α_f с оценкой удаленности от нуля:

$$|\alpha_f| \leq |\dot{g}_0| \frac{1 - \sqrt{1 - 2P_L}}{L}. \quad (13)$$

Функция

$$g(t) = g_0 + \dot{g}_0 t + \frac{Lt^2}{2} \operatorname{sgn} g_0 \quad (14)$$

имеет корень α_g , на котором реализуется равенство в оценке (13). А так как g из \mathcal{G}_1 , то g является максимайзером удаленности корней функций из \mathcal{G}_1 и утверждение ТК об удаленности корней на множестве \mathcal{G}_1 точно.

Если $D = R^1$, то непосредственно к \mathcal{G}_1 теорему 1 применить нельзя, ибо тот же максимайзер доставляет при некотором t нулевое значение производной, т. е. для \mathcal{G}_1 не существует конечного параметра σ . Попробуем ограничить область задания функций из \mathcal{G}_1 благоприятным для теоремы 1 образом: определим D как сегмент $[-d, d]$ с d , подлежащий выбору (см. замечание 3 к теореме 1). Очевидно, должно выполняться $d < \hat{d} = |\dot{g}_0|/L$, поскольку g из \mathcal{G}_1 и либо $\dot{g}(\hat{d}) = 0$, либо $\dot{g}(-\hat{d}) = 0$.

Оценим на D для f из \mathcal{G}_1 снизу абсолютные величины их производных. Из липшицевости \dot{f} следует

$$|\dot{f}(t)| \geq |\dot{g}_0| - L|t|.$$

Поэтому, учитывая $d < \widehat{d}$, имеем

$$M = \max_{f \in \mathcal{G}_1} \max_{t \in [-d, d]} |\dot{f}(t)|^{-1} \leq \frac{1}{|\dot{g}_0| - Ld}. \quad (15)$$

С одной стороны, в качестве $\sigma(d)$ можно назначить любое число, не меньшее ML , с другой — правая часть (15) есть $\max(|\dot{g}(d)|^{-1}, |\dot{g}(-d)|^{-1})$. Таким образом, наименьшим (т. е. наилучшим для теоремы 1) параметром σ является $\sigma(d) = L/(|\dot{g}_0| - Ld)$. Обозначим величину $d\sigma(d)$ через y . Тогда

$$y = \frac{Ld}{|\dot{g}_0| - Ld} \iff 1 + y = \frac{|\dot{g}_0|}{|\dot{g}_0| - Ld} = \frac{\sigma(d)|\dot{g}_0|}{L}. \quad (16)$$

Величина d должна быть не менее параметра d_0 из теоремы 1, что соответствует неравенству

$$d \geq -\frac{\ln(1 - \sigma(d)|g_0/\dot{g}_0|)}{\sigma(d)} \iff e^{-d\sigma(d)} \leq 1 - \sigma(d)|g_0/\dot{g}_0|.$$

Это совместно с (16) дает

$$e^{-y} \leq 1 - \frac{(1+y)L|g_0|}{\dot{g}_0^2} = 1 - (y+1)P_L \implies P_L \leq \frac{1 - e^{-y}}{1+y}. \quad (17)$$

Правая часть (17) является унимодальной функцией с единственным максимумом μ , лежащим в интервале $(0.317, 0.318)$.

Итак:

1) для всех $P_L > \mu \approx 0.317$ теорема 1 не гарантирует существования решения y всех функций из \mathcal{G}_1 ни при каком сужении области задания, а ТК их дает при всех P_L из $[0, 0.5]$;

2) при $P_L = \mu$ теорема 1 дает такую гарантию только в области задания $D_\mu = (-d_\mu, d_\mu)$, где $d_\mu = \dot{g}_0/(Ly_\mu + L)$ — корень уравнения $d\sigma(d) = y_\mu$, а y_μ — корень уравнения $\mu = \frac{1 - e^{-y}}{1+y}$;

отметим, что $d_\mu < \widehat{d}$;

3) для всех P_L из $[0, \mu)$ существует семейство областей D , в которых теорема 1 дает гарантии существования решения, причем семейство для меньших P_L целиком содержит семейство для больших P_L , в частности D_μ . \square

Исследуем обратную связь.

Пример 2, показывающий, что в ЛЛ-классе (D, σ) теорема Канторовича не лучше, чем теорема 1.

Рассмотрим одномерное пространство: $D = R^1$. Обозначим через \mathcal{G}_2 подмножество скалярных функций f из ЛЛ-класса (R^1, σ) , выделяемых двумя условиями $f(0) = g_0$ и $\dot{f}(0) = \dot{g}_0$ (все далее сказанное очевидным образом распространяется и на условия, задаваемые в произвольный момент t_0). Относительно \mathcal{G}_2 теорема 1 утверждает, что если $1 > P_\sigma \equiv \sigma|g_0/\dot{g}_0|$, то каждая функция f из \mathcal{G}_2 имеет корни, среди которых существует единственный корень α_f с оценкой удаленности от нуля:

$$|\alpha_f| \leq -\frac{\ln(1 - P_\sigma)}{\sigma}. \quad (18)$$

Функция

$$g(t) = g_0 - (e^{-\sigma t} - 1) \dot{g}_0/\sigma$$

имеет корень α_g , на котором реализуется равенство в оценке (18). А так как g из \mathcal{G}_2 , то g является максимайзером удаленности корней функций из \mathcal{G}_2 и утверждение теоремы 1 об удаленности корней на множестве \mathcal{G}_2 точно.

Непосредственно к \mathcal{G}_2 ТК применить нельзя, ибо тот же максимайзер доставляет при $t \rightarrow -\infty$ сколь угодно большие значения полупроизводной от \dot{g} , что соответствует отсутствию константы Липшица. Благоприятным для ТК образом сузим множество \mathcal{G}_2 до \mathcal{G}_3 так: $\mathcal{G}_3 := \mathcal{G}_2(\sigma) \cap \mathcal{G}_1(L)$, где L — глобальная константа Липшица, равная наибольшему значению локальных констант Липшица, которые позволено иметь в нуле функциям из $\mathcal{G}_2(\sigma)$, т. е. $L = \sigma|\dot{g}_0|$ и $P_L = P_\sigma$.

Итак, интервал для параметра P_σ , на котором теорема 1 гарантирует существование решения для функций из \mathcal{G}_3 , есть $[0,1)$, что в два раза больше допустимого диапазона для ТК ($P_\sigma = P_L \in [0, 1/2]$). \square

2. Метод дифференцирования по итерации

В анализе сходимости однократных итеративных методов удобным оказался следующий прием. Соединим отрезком текущую итерацию x_k с решением α исходной задачи и параметризуем его:

$$x_k(t) = \alpha + \vec{d}t, \quad \vec{d} \doteq x_k - \alpha, \quad t \in [0, 1]. \quad (19)$$

Если уравнение, порождающее следующую итерацию x_{k+1} , имело вид $G(x - x_k, x_k) = \mathbf{0}$, то после подстановки в него $x_k(t)$ на место x_k получится семейство уравнений с параметром t . Обозначим решение такого уравнения относительно x через $x(t)$ и введем $y \doteq x(t) - \alpha$. Поскольку $G(\mathbf{0}, \alpha) = \mathbf{0}$, справедливо $y(\mathbf{0}) = \mathbf{0}$.

Продифференцируем тождество $G(y - \vec{d}t, x_k(t)) \equiv \mathbf{0}$ по t :

$$G'_1(y - \vec{d}t, x_k(t))(\dot{y} - \vec{d}) + G'_2(y - \vec{d}t, x_k(t))\vec{d} = \mathbf{0} \quad (20)$$

(G'_1 — матрица из частных производных по первой группе переменных, G'_2 — матрица из частных производных по второй группе переменных).

Если $G'_1(y - \vec{d}t, x_k(t))$ не вырождена, (20) можно разрешить относительно \dot{y} :

$$\dot{y} = (I - (G'_1(y - \vec{d}t, x_k(t)))^{-1}G'_2(y - \vec{d}t, x_k(t)))\vec{d}. \quad (21)$$

Отсюда получаем дифференциальное неравенство

$$\|\dot{y}\| \leq \|I - (G'_1(y - \vec{d}t, x_k(t)))^{-1}G'_2(y - \vec{d}t, x_k(t))\| \|\vec{d}\|. \quad (22)$$

Если получить оценку

$$\|I - (G'_1(y - \vec{d}t, x_k(t)))^{-1}G'_2(y - \vec{d}t, x_k(t))\| \leq f(\|y\|, t),$$

то (22) совместно с очевидным $\|y\|'_t \leq \|\dot{y}\|$ и с $y(0) = \mathbf{0}$ даст систему

$$\|y\|'_t \leq f(\|y\|, t)\|\vec{d}\|, \quad \|y(0)\| = 0. \quad (23)$$

Если f непрерывна в некотором открытом множестве M , то [12, теорема 3.4.1] максимальное решение задачи Коши

$$\dot{z} = f(z, t)\|\vec{d}\|, \quad z(0) = \|y(0)\| = 0, \quad (24)$$

мажорирует любое решение системы (23) на общем интервале существования $[0, a]$.

Согласно построению, $\|x_{k+1} - \alpha\| = \|y(1)\|$. Следовательно, нужно получить оценку $\|y(1)\| \leq z(1)$. Поэтому общий интервал существования должен содержать $[0, 1]$, т. е. должно быть обеспечено $a \geq 1$. Путь к этому указывает следствие из теоремы Пеано:

Лемма 2. *Определим для положительных b прямоугольник $R_b = [-b, b] \times [0, 1]$. Пусть существует b такое, что*

$$\sup_{(z,t) \in R_b} |f(z,t)| \leq b.$$

Пусть f непрерывна в R_b . Тогда задача Коши (24) имеет на сегменте $[0, 1]$ хотя бы одно решение, мажорирующее любое решение системы (23) на общем интервале существования.

В тех случаях, когда функция G всего лишь липшицева и не дифференцируема, существенное удобство в выкладках предоставляет несколько иной взгляд на понятие разностного отношения $h(\tau, t) := \frac{h(t+\tau) - h(t)}{\tau}$ функции $h(\eta)$ скалярного аргумента η из сегмента D . Обычно второй аргумент разностного отношения считается фиксированным, а в оценках, которые предстоит сделать, он будет переменным. В описании такой смысловой нагрузки на параметр целесообразно ввести дополнительное определение.

Определение 4. *Разностной производной в точке t заданной на сегменте $D \subset R^1$ функции h будем называть зависящую от параметра t функцию $h^\nabla(t, \cdot)$, заданную в $\widehat{D}(t, D) = (D - t) \setminus \{0\}$ согласно формуле*

$$h^\nabla(t, \varepsilon) = \frac{h(t+\varepsilon) - h(t)}{\varepsilon} \quad \forall \varepsilon \in \widehat{D}(t, D). \quad (25)$$

Величину ε будем называть отклонением, а операцию нахождения разностной производной — разностным дифференцированием.

Допуская вольность речи, будем опускать второй параметр разностной производной тогда, когда это не исказит смысл формулы.

У разностной производной есть важное достоинство в сравнении с обычной производной — она существует и конечна всегда, когда исходная функция всего лишь определена и принимает конечные значения. Вместе с тем она обладает свойствами, сходными со свойствами обычной производной.

1. $(f \pm g)^\nabla = f^\nabla \pm g^\nabla$.
2. $(\tau f)^\nabla = \tau f^\nabla$, τ — число.
3. Если f дифференцируема в точке t , то

$$f^\nabla(t, \varepsilon) = f'(t) + \omega(\varepsilon) \cong f'(t),$$

где $\lim_{\varepsilon \rightarrow 0} \omega(\varepsilon) = \mathbf{0}$. Здесь и далее под знаками \cong и \lesssim будут пониматься соответственно равенство и неравенство, приближенные с точностью до величин, стремящихся к $\mathbf{0}$, когда к $\mathbf{0}$ стремится отклонение в разностной производной.

Замечание 4. Когда f в системе (23) не зависит от z : $f(z, t) \|\vec{d}\| = \varphi(t)$, справедливо более сильное, чем лемма 2, утверждение. Достаточно измеримости по Лебегу и конечности функции φ на $[0, 1]$ для того, чтобы интеграл $\mathcal{I}(t) \doteq \int_0^t \varphi(\tau) d\tau$, $t \in [0, 1]$, мажорировал [14] все решения систем

$$\|y\|_t^\nabla \lesssim \varphi(t), \quad \|y(0)\| = 0,$$

$$\|y\|'_t \leq \varphi(t), \quad \|y(0)\| = 0.$$

(Интеграл $\mathcal{I}(t)$ является решением задачи Коши (24): $z(t) = \mathcal{I}(t)$ для всех t из $[0, 1]$. [15, п. 11.8.2.]). \square

Вернемся к свойствам разностной производной.

4. Пусть f и g — матрицы размерности $(m \times n)$ и $(n \times k)$ соответственно и g имеет конечную локальную константу Липшица в t . Тогда

$$g(t + \varepsilon) = g^\nabla(t, \varepsilon)\varepsilon + g(t) \cong g(t). \quad (26)$$

С другой стороны,

$$\begin{aligned} (fg)^\nabla(t, \varepsilon) &= [f(t + \varepsilon)g(t + \varepsilon) - f(t)g(t)] / \varepsilon = \\ &= [(f(t + \varepsilon) - f(t))g(t + \varepsilon) + f(t)(g(t + \varepsilon) - g(t))] / \varepsilon = f^\nabla(t)g(t + \varepsilon) + f(t)g^\nabla(t). \end{aligned}$$

Это совместно с (26) дает $(fg)^\nabla(t) \cong f^\nabla(t)g(t) + f(t)g^\nabla(t)$.

Заметим, что это свойство имеет место и при липшицевости f вместо g .

5. Пусть f дифференцируема в точке $g(t)$ и g имеет конечную локальную константу Липшица в t . Тогда

$$(f(g))^\nabla(t, \varepsilon) = [f(g(t + \varepsilon)) - f(g(t))] / \varepsilon = f'(g(t))g^\nabla(t) + o(g(t + \varepsilon) - g(t)) / \varepsilon \cong f'(g(t))g^\nabla(t).$$

6. Пусть f и g имеют локальные константы Липшица L_f и L_g в точках $g(t)$ и t соответственно. Тогда $\|(f(g))^\nabla(t)\| \lesssim L_f \|g^\nabla(t)\| \lesssim L_f L_g$.

Действительно,

$$\|(f(g))^\nabla(t, \varepsilon)\| \leq L_f (\|g(t + \varepsilon) - g(t)\| - o(g(t + \varepsilon) - g(t))) / \varepsilon$$

и

$$\frac{o(g(t + \varepsilon) - g(t))}{\varepsilon} = \frac{o(g(t + \varepsilon) - g(t))}{\|g(t + \varepsilon) - g(t)\|} \|g^\nabla(t)\| \longrightarrow 0 \text{ при } \varepsilon \longrightarrow 0.$$

7. Пусть функция f имеет производные f'_g и f'_h , а функции g и h имеют конечные локальные константы Липшица L_g и L_h в точке t . Тогда

$$(f(g, h))^\nabla(t) \cong f'_g(g(t), h(t))g^\nabla(t) + f'_h(g(t), h(t))h^\nabla(t).$$

Это, очевидно, следует из разложения $(f(g, h))^\nabla(t, \varepsilon) =$

$$= (f'_g(g(t), h(t)), f'_h(g(t), h(t))) \begin{pmatrix} g(t + \varepsilon) - g(t) \\ h(t + \varepsilon) - h(t) \end{pmatrix} \varepsilon^{-1} + o(g(t + \varepsilon) - g(t), h(t + \varepsilon) - h(t)) \varepsilon^{-1}.$$

8. Пусть функция f имеет конечные локальные константы Липшица по группам аргументов: L_{fg} и L_{fh} при их значениях $g(t)$, $h(t)$, а функции g и h имеют конечные локальные константы Липшица L_g и L_h в точке t . Тогда

$$\begin{aligned} \|(f(g, h))^\nabla(t, \varepsilon)\| &\leq \\ &\leq L_{fg} \|g^\nabla(t, \varepsilon)\| + L_{fh} \|h^\nabla(t, \varepsilon)\| + o(\|g(t + \varepsilon) - g(t)\| + \|h(t + \varepsilon) - h(t)\|) / \varepsilon \cong \\ &\cong L_{fg} \|g^\nabla(t)\| + L_{fh} \|h^\nabla(t)\| \cong L_{fg} L_g + L_{fh} L_h. \end{aligned}$$

Теорема 3.4.1 из [12] позволяет применить метод дифференцирования по итерации при отсутствии G'_2 — производной функции G по второй группе переменных и уйти от вопроса о существовании производной у $x(t)$. Для этого следует произвести разностное дифференцирование по t тождества $G(y - \vec{d}t, x_k(t)) \equiv \mathbf{0}$ и начать исследовать вместо (22) приближенное дифференциальное неравенство

$$\|y^\nabla\| \lesssim \|I - (G'_1(y - \vec{d}t, x_k(t)))^{-1} G'_2(y - \vec{d}t, x_k(t))\| \|\vec{d}\| \quad (27)$$

с использованием вышеприведенных свойств разностной производной.

3. Сходимость

При известной оценке d удаленности решения от начальной точки для метода Ньютона в $C^{1,1}$ были получены [9] гарантии сходимости при условии $P_L d \leq 2/3$. Аналогичный результат в ЛЛ-классе интересен появлением трансцендентной константы, причем проверяется, что она точна.

Теорема 2. (Локальная сходимость метода Ньютона). Пусть область D принадлежит B -пространству, пусть функция g принадлежит ЛЛ-классу (D, σ) и некоторая оценка \hat{d} удаленности решения α от начальной точки удовлетворяет следующим условиям:

1) $\sigma \hat{d} \doteq c_0 < \bar{c}$, где \bar{c} есть решение уравнения

$$C(c) \doteq (e^c - 1 - c)/c = 1 \quad (28)$$

(расчеты показывают, что \bar{c} принадлежит $(1.25, 1.26)$);

2) $S(x_0, p\hat{d}) \subset D$, где $p = 1 + C(\sigma \hat{d})$.

Тогда:

а) метод Ньютона, начатый в точке x_0 , корректно определен и сходится к решению α , все итерации лежат в шаре $S(x_0, p\hat{d})$;

б) скорость сходимости оценивается неравенством

$$\|x_{k+1} - \alpha\| \leq C(\sigma \|x_k - \alpha\|) \|x_k - \alpha\|,$$

причем при $s \leq 1$ справедливо $C(\sigma s) \leq C(\sigma)s$;

в) решение α системы $g(x) = \mathbf{0}$ единственно в шаре $S(x_0, p\hat{d})$.

Доказательство. Корректность метода Ньютона на первом шаге (т. е. существование $J^{-1}(x_0)$) следует из того, что функция g принадлежит ЛЛ-классу (D, σ) , и из того, что x_0 из D (условие 2).

Пусть метод Ньютона был корректен и не породил итераций, удаленных от решения более чем на \hat{d} , вплоть до k -го шага. Исследуем $(k+1)$ -й шаг с помощью метода дифференцирования по итерации.

Определим $d \doteq \|x_k - \alpha\|$, $\vec{d} \doteq (x_k - \alpha)/d$. Затем введем параметр t и зададим $x_k(t)$ как $\alpha + \vec{d}t$ (очевидно, $x_k(d) = x_k$).

Решение относительно x уравнения

$$J(x_k(t))(x - x_k(t)) + g(x_k(t)) = \mathbf{0} \quad (29)$$

при $t = d$ есть x_{k+1} , при $t = 0$ есть α . Дифференцируя разностно (29) по t и заменяя g^∇ на J , с точностью до погрешностей, возникших из-за этой замены и бесконечно малых

относительно размеров окрестности разностного дифференцирования, имеем

$$(J)_t^\nabla(x - x_k) + J(x^\nabla - \vec{d}) + J\vec{d} \cong \mathbf{0}.$$

(Здесь опущены аргумент t у x_k и аргумент $x_k(t)$ у J, J_t^∇). После преобразования: $x^\nabla \cong J^{-1}J_t^\nabla(x_k - x)$. Оценивая полупроизводной норму разностной производной, имеем

$$\|x^\nabla\| \lesssim \|J^{-1}\| L_J(x) \|x_k - x\| \leq \sigma \|x_k - x\|.$$

Произведем подстановку $x = \alpha + y$. Тогда

$$\|y^\nabla\| \lesssim \sigma(t + \|y\|), \quad y(0) = \mathbf{0}. \quad (30)$$

Так как $\| \|y\|^\nabla \| \leq \|y^\nabla\|$, согласно теореме 3.4.1 из [12] мажорирующая норма любого решения этой системы задача Коши имеет следующий вид:

$$\dot{z} = (t + z)\sigma, \quad z(0) = 0.$$

Решив ее, придем к оценке

$$\|x_{k+1} - \alpha\| = \|y(d)\| \leq z(d) = \left. \frac{e^{\sigma t} - \sigma t - 1}{\sigma} \right|_{t=d} = d \frac{e^{\sigma d} - \sigma d - 1}{\sigma d} = \|x_k - \alpha\| C(\sigma d). \quad (31)$$

Нетрудно проверить, что функция C монотонно возрастает. А по условию 1 и индуктивному предположению выполняется $\sigma d \leq \sigma \hat{d} = c_0 < \bar{c}$, $C(\bar{c}) = 1$. В совокупности с монотонностью функции C это дает

$$C(\sigma d) \leq C(\sigma \hat{d}) = C(c_0) < C(\bar{c}) = 1,$$

с учетом (31) обеспечивающее выполнение неравенства $\|x_{k+1} - \alpha\| < \|x_k - \alpha\|$ и по индукции — справедливость (31) для всех k . Так как α находится не далее \hat{d} от x_0 , а итерации x_1, x_2, \dots находятся не далее от α , чем $C(\sigma \hat{d})\|\hat{d}\|$, то все итерации лежат в шаре $S(x_0, \hat{d} + \hat{d}C(\sigma \hat{d}))$. Поэтому можно применить оценку (31) k раз: $\|x_k - \alpha\| \leq C^k(c_0)\|x_0 - \alpha\|$, что означает сходимость итеративного процесса.

При наличии сходимости п. б есть оценка (31).

Единственность решения системы $g(x) = \mathbf{0}$ в шаре $S(x_0, p\hat{d})$ следует из того, что при наличии еще одного решения α' уже сходящаяся к α ньютоновская итеративная последовательность должна также сходиться и к α' , что возможно только при $\alpha' = \alpha$.

Если $d < 1$, то из разложения C в ряд следует $C(\sigma d) < C(\sigma)d$. Поэтому, если верны неравенства $\|x_k - \alpha\|C(\sigma) < 1$ и $\|x_k - \alpha\| < 1$, то справедлива следующая оценка скорости сходимости:

$$\|x_{k+1} - \alpha\| < C(\sigma)\|x_k - \alpha\|^2. \quad \square$$

Теорема 3. *Условия 1, 2 теоремы 2 являются также и необходимыми в том смысле, что при их нарушении можно найти функцию и начальную точку, для которых метод Ньютона не сходится, а для начальных точек, обеспечивающих сходимость, оценка скорости сходимости точна.*

Доказательством этой теоремы является следующий

Пример 3. Рассмотрим скалярную функцию $g(t) = c(-e^{-|t|\sigma} + 1) \operatorname{sgn} t$, где $c > 0$, $t \in D \subset R^1$. Для нее $|\ddot{g}(t)/\dot{g}(t)| = \sigma$ для всех t . Следовательно, g принадлежит ЛЛ-классу (D, σ) .

Вычислим удаленность приближения $T(t)$ после итерации по методу Ньютона, начатого из произвольного момента t , от начала координат — решения уравнения (1):

$$T(t) = t - \frac{g(t)}{\dot{g}(t)} = t + \frac{e^{-t\sigma} - 1}{\sigma e^{-t\sigma}} = t + \frac{1}{\sigma} - \frac{e^{t\sigma}}{\sigma} \equiv -C(t\sigma)t. \quad (32)$$

Это означает точность оценки п. б) теоремы 2 (независимо от того, сходится метод Ньютона или нет). Поэтому в силу монотонного возрастания функции C и того, что $C(\bar{c}) = 1$, удаленность первой итерации от решения уравнения (1) строго уменьшается при $\sigma|t_0| < \bar{c}$, строго увеличивается при $\sigma|t_0| > \bar{c}$ и неизменно при равенстве. Используя рекурсию, приходим к выводу, что в первом случае метод Ньютона будет сходиться, во втором — расходиться к бесконечности, в третьем — иметь неустойчивый цикл. Другими словами, условие 1 теоремы 2 не чрезмерно, и ее оценка скорости сходимости для этого примера точна.

Выясним ситуацию с условием 2 теоремы 2. Вычтем t из начального и конечного звеньев (32):

$$T(t) - t = -t \left(1 + C(t\sigma) \right) \Big|_{t=\hat{d}} = -\hat{d}p.$$

Таким образом, расстояние между начальным приближением и первой итерацией может оказаться равным $\hat{d}p$, и для принадлежности первой итерации области задания D необходимость условия 2 очевидна. \square

Простым следствием теорем 1 и 2 является

Теорема 4. Пусть U, W — H -пространства, $D \subset U$, $g : D \rightarrow W$, g принадлежит ЛЛ-классу (D, σ) и выполнены также условия:

$$1) P_\sigma \equiv r_0 \sigma \sqrt{g^2(x_0)} < \frac{2\bar{c}}{1 + 2\bar{c}}, \text{ где } \bar{c} - \text{корень уравнения (28);}$$

$$2) S(x_0, pd_0) \subset D, \text{ где } p = 1 + C(\sigma d_0), \quad d_0 = -\frac{\ln(1 - P_\sigma)}{\sigma}.$$

Тогда:

- а) существует решение α системы $g(x) = \mathbf{0}$, единственное в шаре $S(x_0, d_0)$;
- б) метод Ньютона, начатый в точке x_0 , корректно определен и сходится к конечному решению α ;
- в) скорость сходимости оценивается неравенством

$$\|x_{k+1} - \alpha\| \leq C(\sigma \|x_k - \alpha\|) \|x_k - \alpha\|.$$

Доказательство. Поскольку $2\bar{c} < 1 + 2\bar{c}$, выполняется условие 1 теоремы 1. Условие 2 доказываемой теоремы сильнее условия 2 теоремы 1. Применение последней показывает, что d_0 есть оценка удаленности решения α от точки x_0 .

Из условия 1 теоремы 4 и определения функции C следует, что

$$1 - P_\sigma > 1 - \frac{2\bar{c}}{1 + 2\bar{c}} = \frac{1}{1 + 2\bar{c}} \equiv e^{-\bar{c}}.$$

Потенцируя это соотношение и используя определение d_0 , получаем

$$\bar{c} > -\ln(1 - P_\sigma) = \sigma d_0.$$

Таким образом, выполнено условие 1 теоремы 2 при $\hat{d} = d_0$. Очевидно, что условие 2 теоремы 2 содержится в условии 2 теоремы 4 при $\hat{d} = d_0$. Применение теоремы 2 завершает доказательство теоремы 4. \square

В теореме 1, гарантирующей существование решения, условие на характерный параметр ($P_\sigma < 1$) слабее условия на него $\left(P_\sigma < \frac{2\bar{c}}{1+2\bar{c}} \approx 0.71\right)$ в теореме 4, гарантирующей сходимость метода Ньютона. Возможно, это является не атрибутом ЛЛ-класса, а следствием упрощенного подхода к доказательству теоремы 4¹. Тем не менее для некоторых функций, принимаемых за элементы ЛЛ-класса, она гарантирует сходимость метода Ньютона, в то время как зачисление этих функций в $C^{1,1}$ дает параметр P_L , больший $1/2$, и теорема Канторовича не гарантирует сходимости метода Ньютона.

Примером такой функции может служить рассмотренная в примере 2

$$g(t) = \begin{cases} g_0 - (e^{-\sigma t} - 1)\dot{g}_0/\sigma, & t \geq 0 \\ g_0 + \dot{g}_0 t, & t < 0 \end{cases}.$$

Для нее $P_\sigma = P_L = P_K = \sigma g_0/\dot{g}_0$. При любом соотношении величин $g_0 > 0$, $\dot{g}_0 < 0$, обеспечивающих $P_\sigma \in \left(\frac{1}{2}, \frac{2\bar{c}}{1+2\bar{c}}\right)$, теорема 4 гарантирует сходимость ньютоновского процесса, теорема Канторовича — нет.

Обратную ситуацию: гарантии от ТК сходимости метода Ньютона для функции, принимаемой за элемент из $C^{1,1}$, и неприменимость к ней как к элементу ЛЛ-класса теоремы 4, можно наблюдать в примере 1 для функции (14). Там $P_L = Lg_0/\dot{g}_0^2$, и при удовлетворяющем теореме Канторовича $P_L \in [-1 + \sqrt{2}, 1/2]$ будет, какую бы область D не взять, $P_\sigma \geq 1$: не выполнено условие теоремы 4. Выбрать соответствующие параметры g_0 , \dot{g}_0 не составляет труда.

Список литературы

- [1] КАНТОРОВИЧ Л. В. Функциональный анализ и прикладная математика // Успехи мат. наук. 1948. № 3, вып. 6. С. 89–185.
- [2] КАНТОРОВИЧ Л. В. О методе Ньютона // Тр. Мат. ин-та им. В.А. Стеклова. 1949. № 28. С. 104–144.
- [3] МЫСОВСКИХ И.П. К вопросу о сходимости метода Ньютона // Там же. С. 145–147.
- [4] КАНТОРОВИЧ Л. В. Принцип мажорант и метод Ньютона // Докл. АН СССР. 1951. Т. 76, № 1. С. 17–20.
- [5] КАНТОРОВИЧ Л. В. Некоторые дальнейшие применения метода мажорант // Докл. АН СССР. 1951. Т. 80, № 6. С. 849–852.

¹Если этим же способом получить оценки для $C^{1,1}$, т. е. использовать теорему Мысовских [9] с оценкой удаленности из ТК, то получим

$$Lr_0 \frac{1 - \sqrt{1 - 2P_L}}{Lr_0} < \frac{2}{3} \iff 1 - 2P_L > \frac{1}{9} \iff P_L < \frac{4}{9},$$

что несколько сильнее условия из ТК: $P_L \leq 1/2$.

- [6] КАНТОРОВИЧ Л. В. Приближенное решение функциональных уравнений // Успехи мат. наук. 1956. № 11, вып. 6. С. 99–116.
- [7] КАНТОРОВИЧ Л. В. Некоторые дальнейшие применения метода Ньютона // Вест. ЛГУ. Сер. Математика, механика и астрономия. 1957. Вып. 2. С. 68–103.
- [8] КАНТОРОВИЧ Л. В., АКИЛОВ Г. П. Функциональный анализ. М., 1977. 744 с.
- [9] МЫСОВСКИХ И. П. О сходимости метода Л. В. Канторовича решения функциональных уравнений и его применениях // Докл. АН СССР. 1950. Т. 70, № 4. С. 565–568.
- [10] ОРТЕГА ДЖ., РЕЙНБОЛДТ В. Итерационные методы решения нелинейных систем уравнений со многими неизвестными. М., 1975. 558 с. (J. Ortega and W. Rheinboldt. Iterative Solution of Nonlinear Equations in Several Variables).
- [11] КРАСНОСЕЛЬСКИЙ М. А. и др. Приближенное решение операторных уравнений. М., 1969. 455 с.
- [12] ХАРТМАН Ф. Обыкновенные дифференциальные уравнения. М., 1970. 720 с.
- [13] КОЛМОГОРОВ А. Н., ФОМИН С. В. Элементы теории функций и функционального анализа. М., 1972. 496 с.
- [14] МИХЕЕВ С. Е. Нелинейные методы в оптимизации. СПб., 2001. 276 с.
- [15] ТИТЧМАРШ Е. Теория функций. М., 1982. 463 с.

*Поступила в редакцию 31 августа 2004 г.,
в переработанном виде — 30 декабря 2004 г.*