TARTU RIIKLIKU ÜLIKOOLI

TOIMETISED

УЧЕНЫЕ ЗАПИСКИ
ТАРТУСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА
ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS

633

ARVUTUSMEETODID RAJAÜLESANNETE JA OPERAATORVÕRRANDITE LAHENDAMISEKS

МЕТОДЫ ВЫЧИСЛЕНИЙ ДЛЯ КРАЕВЫХ ЗАДАЧ И ОПЕРАТОРНЫХ УРАВНЕНИЙ

Matemaatika- ja mehhaanikaalaseid töid

Труды по математике и механике

TARTU RIIKLIKU ÜLIKOOLI TOIMETISED
УЧЕНЫЕ ЗАПИСКИ
ТАРТУСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА
ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS
ALUSTATUD 1893.a. VIHIK 633 ВЫПУСК ОСНОВАНЫ В 1893.г.

ARVUTUSMEETODID RAJAÜLESANNETE JA OPERAATORVÕRRANDITE LAHENDAMISEKS

МЕТОДЫ ВЫЧИСЛЕНИЙ ДЛЯ КРАЕВЫХ ЗАДАЧ И ОПЕРАТОРНЫХ УРАВНЕНИЙ

Matemaatika- ja mehhaanikaalaseid töid
Труды по математике и механике

ТАРТУ 1983

Ученые записки Тартуского государственного университета.
Выпуск 633.
МЕТОДЫ ВЫЧИСЛЕНИЙ ДЛЯ КРАЕВЫХ ЗАДАЧ И ОПЕРАТОРНЫХ УПРАВНЕНИЙ.
Труды по математике и механике.
На русском языке.
Резюме на английском и немецком языках.
Тартуский государственный университет.
ЭССР, 202400, г.Тарту, ул.Юликооли, I8.
Ответотвенный редактор Э. Тамме.
Корректоры Б. Шифрин, Э. Яйгма, С. Райтар.
Подписано к печати I5.I2.I982.
МВ I2930.
Формат 60x90/I6.
Бумага писчая.
Машинопись. Ротапринт.
Учетно-издательских листов 4,69.
Печатных листов 6;0.
Тираж 350.
Заказ № I33I.
Цена 70 коп.
Типография ТГУ, ЭССР, 202400, г.Тарту, ул.Пялсона, I4.
2 - 2

Уч.зап.Тартуск.ун-та,1982,633,3-10

ОБ ИНВАРИАНТНОСТИ ВРАЩЕНИЯ ВЕКТОРНЫХ ПОЛЕЙ ПРИ АППРОКСИМАЦИИ МНОГОЗНАЧНЫХ ОТОБРАЖЕНИЙ

Г. Вайникко

Указываются условия, в которых аппроксимирующие и аппроксимируемое многозначные отображения порождают векторные поля одинакового вращения. Основной результат (теорема 2) касается случая, когда одновременно с отображением аппроксимируется и пространство, в котором оно действует.

§ 1. Элементарная теорема

Пусть E – банахово пространство, $\Omega \subset E$ – непустое открытое множество; границу и замыкание будем обозначать через $\partial\Omega$ и $cl\,\Omega$. Рассмотрим многозначное отображение $T: cl\,\Omega \to \mathcal{K}(E)$, где $\mathcal{K}(E)$ – множество всех непустых выпуклых компактных подмножеств пространства E. Допустим, что отображение T замкнуто (замкнут его график) и компактно (т.е. $T(cl\,\Omega)$ относительно компактно в E). Если T на границе $\partial\Omega$ не имеет неподвижных точек, т.е. $u \bar{\in} Tu$ для $u \in \partial\Omega$, то определено вращение $\gamma(I-T; \partial\Omega)$ векторного поля $u - Tu$ на границе $\partial\Omega$; за подробным определением и свойствами ссылаемся на [1,5]. Отметим здесь только, что вращение – это целочисленная характеристика отображения T, характеризирующая наличие у него неподвижных точек в Ω и инвариантная относительно гомотопий. Ниже будут использованы только линейные гомотопии. Напомним соответствующее определение: замкнутые компактные отображения $T, T_1 : cl\,\Omega \to \mathcal{K}(E)$ линейно гомотопны на $\partial\Omega$, если

$$u \bar{\in} \lambda T_1 u + (1-\lambda) Tu \qquad \forall u \in \partial\Omega, \quad \forall \lambda \in [0,1].$$

Теорема 1. Пусть отображения $T: cl\,\Omega \to \mathcal{K}(E)$ и $T_n: cl\,\Omega \to \mathcal{K}(E)$ $(n \in N = \{1,2,\ldots\})$ замкнуты и компактны, причем T_n компактно аппроксимируют T в том смысле, что

$$u_n \in cl\,\Omega,\ v_n \in T_n u_n\ (n \in N) \Rightarrow (v_n) \text{ относительно компактна}; \quad (1)$$

$$u_n \in cl\,\Omega,\ v_n \in T_n u_n,\ u_n \to u,\ v_n \to v\ (n \in N' \subseteq N) \Rightarrow v \in Tu. \quad (2)$$

Пусть T не имеет на $\partial\Omega$ неподвижных точек. Тогда при достаточно больших n таким же свойством обладают T_n, и

$$\gamma(I-T_n;\partial\Omega)=\gamma(I-T;\partial\Omega) \quad (n=n_0,n_0+1,\ldots). \quad (3)$$

Если при этом $\gamma(I-T;\partial\Omega)\neq 0$, то при достаточно больших n отображение T_n имеет в Ω хотя бы одну неподвижную точку u_n; любая последовательность $(u_n)_{n\geq n_0}$, составленная из неподвижных точек отображений T_n, относительно компактна, и ее предельные точки являются неподвижными для T.

Доказательство. Из условий теоремы следует, что отображения T и T_n при достаточно больших n линейно гомотопны на $\partial\Omega$. Действительно, рассуждая от противного, допустим, что для некоторых $\lambda_n\in[0,1]$ и $u_n\in\partial\Omega$ имеем

$$u_n\in\lambda_n T_n u_n+(1-\lambda_n)Tu_n \quad (n\in N'\subseteq N).$$

Выберем $v_n\in T_n u_n$ и $v'_n\in Tu_n$ таким образом, что

$$u_n=\lambda_n v_n+(1-\lambda_n)v'_n.$$

Последовательность (v_n) относительно компактна в силу условия (I); последовательность (v'_n) относительно компактна в силу компактности T. Итак, последовательность (u_n) относительно компактна. Пусть

$$\lambda_n\to\lambda,\ u_n\to u,\ v_n\to v,\ v'_n\to v' \quad (n\in N''\subseteq N').$$

Тогда $\lambda\in[0,1]$, $u\in\partial\Omega$ и

$$u=\lambda v+(1-\lambda)v'.$$

Из (2) следует, что $v\in Tu$; из замкнутости T следует, что $v'\in Tu$. Значит, $u\in Tu$, и мы пришли к противоречию с допущением, что T не имеет на $\partial\Omega$ неподвижных точек. Тем самым (3) доказано.

Пусть $\gamma(I-T;\partial\Omega)\neq 0$. Тогда в силу (3) при $n\geq n_0$ отображение T_n имеет в Ω неподвижную точку $u_n\in T_n u_n$. Последовательность $(u_n)_{n\geq n_0}$ относительно компактна в силу (I), а для ее предельных точек u в силу (2) имеем $u\in Tu$.

Теорема I доказана.

§ 2. Случай аппроксимации пространства

Пусть E и E_n $(n\in N)$ – банаховы пространства, а $\mathcal{P}=(p_n)$ – система отображений $p_n: E\to E_n$ $(n\in N)$, обладающих тем свойством, что для любых $u,u'\in E$ и любых скаляров a,a'

$$\|p_n u\|_{E_n}\to\|u\|_E,\quad \|p_n(au+a'u')-(ap_n u+a'p_n u')\|_{E_n}\to 0 \quad (n\in N).$$

Под $\mathcal{P}$-сходимостью $u_n\xrightarrow{\mathcal{P}}u$, $u_n\in E_n$, $u\in E$, подразумевается, что $\|u_n-p_n u\|_{E_n}\to 0$ при $n\to\infty$; соответствующий смысл придается и $\mathcal{P}$-компактности последовательности $(u_n)_{n\in N}$, $u_n\in E_n$, а именно, указанная последовательность $\mathcal{P}$-компактна, если для любого (бесконечного) $N'\subseteq N$ существует (бесконечное) $N''\subseteq N'$, такое что подпоследовательность $(u_n)_{n\in N''}$ $\mathcal{P}$-сходится. Подробнее об этих понятиях см., например, [4,6].

В дальнейшем через I будем обозначать единичный оператор как в E, так и в E_n. Пусть $\Omega\subset E$ и $\Omega_n\subset E_n$ $(n\in N)$ – непустые открытые множества. Следующая теорема является аналогом теоремы I в случае $\mathcal{P}$-сходимости и обобщает соответствующие результаты из [2,3,6] об однозначных отображениях.

Теорема 2. Пусть: I) подмножества $\Omega\subset E$ и $\Omega_n\subset E_n$ $(n\in N)$ согласованы в том смысле, что

$$\forall u\in cl\,\Omega \quad \exists (u_n)_{n\in N}: \ u_n\in cl\,\Omega_n,\ u_n\xrightarrow{\mathcal{P}}u\ (n\in N); \tag{4}$$

$$u_n\in cl\,\Omega_n,\ u_n\xrightarrow{\mathcal{P}}u\ (n\in N'\subseteq N)\Rightarrow u\in cl\,\Omega; \tag{5}$$

$$u_n\in\partial\Omega_n,\ u_n\xrightarrow{\mathcal{P}}u\ (n\in N'\subseteq N)\Rightarrow u\in\partial\Omega; \tag{6}$$

2) отображения $T: cl\,\Omega\to\mathcal{K}(E)$ и $T_n: cl\,\Omega_n\to\mathcal{K}(E_n)$ $(n\in N)$ замкнуты и компактны, причем T_n компактно аппроксимируют T в том смысле, что

$$u_n\in cl\,\Omega_n,\ v_n\in T_n u_n\ (n\in N)\Rightarrow (v_n)\ \ \mathcal{P}\text{-компактна}; \tag{7}$$

$$u_n\in cl\,\Omega_n,\ v_n\in T_n u_n,\ u_n\xrightarrow{\mathcal{P}}u,\ v_n\xrightarrow{\mathcal{P}}v\ (n\in N'\subseteq N)\Rightarrow v\in Tu; \tag{8}$$

3) отображение T не имеет на границе $\partial\Omega$ неподвижных точек.

Тогда при достаточно больших n отображение T_n не имеет на $\partial\Omega_n$ неподвижных точек, и

$$\gamma(I-T_n;\partial\Omega_n)=\gamma(I-T;\partial\Omega) \qquad (n\geq n_0). \tag{9}$$

Если при этом $\gamma(I-T;\partial\Omega)\neq 0$, то при достаточно больших n отображение T_n имеет в Ω_n хотя бы одну неподвижную точку u_n; любая последовательность $(u_n)_{n\geq n_0}$, составленная из неподвижных точек отображений T_n, $\mathcal{P}$-компактна, и ее $\mathcal{P}$-предельные точки являются неподвижными для T.

Доказательство. 0. Идея доказательства заключается в следующем: строим линейно-гомотопический переход от T и T_n к некоторым отображениям S и S_n, образы которых лежат в некоторых конечномерных подпространствах $E^\circ\subseteq E$ и $E_n^\circ\subseteq E_n$ одинаковой размерности; затем линейным преобразованием сводим эти отображения в отображения в одном общем пространстве (в пространстве E°) и устанавливаем линейную гомотопию между ними. Некоторые детали доказательства повторяют рассуждения [2,3,6],некоторые детали за счет допустимости многозначности даже упрощаются.

I. Из замкнутости и компактности отображения T и условия 3 теоремы следует, что

$$\alpha \equiv \inf_{u\in\partial\Omega} dist(u, Tu) > 0. \tag{I0}$$

Положим $\varepsilon = \alpha/2$ и выберем конечную ε-сеть $\{w^1,\dots,w^r\}\subset E$ относительно компактного множества $T(cl\,\Omega)$. Пусть $w^1,\dots,w^s$ $(s\le r)$ линейно независимы, а $w^k=\sum_{j=1}^s a_{jk}w^j$ для $k=s+1,\dots,r$. Положим $w_n^j=p_n w^j$ $(n\in N;\ j=1,\dots,s)$ и $w_n^k=\sum_{j=1}^s a_{jk}w_n^j$ $(n\in N)$ для $k=s+1,\dots,r$. Тогда $w_n^j \xrightarrow{\mathcal{P}} w^j$ $(n\in N)$ для $j=1,\dots,r$. Обозначим через $E^\circ\subseteq E$ и $E_n^\circ\subseteq E_n$ линейные оболочки элементов $w^1,\dots,w^s$ и $w_n^1,\dots,w_n^s$ соответственно.При достаточно больших n элементы $w_n^1,\dots,w_n^s$ линейно независимые, и $\dim E_n^\circ=\dim E^\circ=s$ $(n\ge n_1)$. Условиями $\varphi_n w^j=w_n^j$ $(j=1,\dots,s)$ определим линейные операторы $\varphi_n: E^\circ\to E_n^\circ$ $(n\in N)$. При $n\ge n_1$ оператор φ_n обратим, $\|\varphi_n\|\to 1$, $\|\varphi_n^{-1}\|\to 1$ при $n\to\infty$. Для $u_n\in E_n^\circ$ $(n\in N'\subseteq N)$ имеем $u_n\xrightarrow{\mathcal{P}} u$ тогда и только тогда, когда $\|\varphi_n^{-1}u_n-u\|_E\to 0$. Ясно также, что из условий $u_n\in E_n^\circ$, $\|u_n\|\le const$ $(n\in N)$ следует $\mathcal{P}$-компактность последовательности (u_n), и ее $\mathcal{P}$-предельные точки лежат в E°.

2. Введем многозначные отображение $S: cl\,\Omega\to\mathcal{K}(E)$ и $S_n: cl\,\Omega_n\to\mathcal{K}(E_n)$, положив

$$Su=(Tu)^{dist(Tu,E^\circ)}\cap E^\circ,\quad u\in cl\,\Omega,$$
$$S_n u_n=(T_n u_n)^{dist(T_n u_n,E_n^\circ)}\cap E_n^\circ,\quad u_n\in cl\,\Omega_n,$$

где $(Tu)^\rho$-замыкание ρ-окрестности множества $Tu\subset E$, а

$$dist(A,B)=\sup_{u\in A}\inf_{v\in B}\|u-v\|\quad (A,B\subset E).$$

Отображения S и S_n компактны ввиду конечномерности E° и E_n°. Менее очевидна их замкнутость. Проведем доказательство замкнутости S; для S_n доказательство такое же.

Пусть $u^n\in cl\,\Omega$, $v^n\in Su^n$, $u^n\to u$, $v^n\to v$ $(n\in N)$. Нужно показать, что $v\in Su$. Поскольку

$$v\in(Tu)^{dist(v,Tu)}$$

(такое включение верно для любого $v\in E$) и $v\in E^\circ$, то достаточно установить неравенство

$$dist(v,Tu)\le dist(Tu,E^\circ).$$

Имеем

$$dist(v, Tu) \le \lim \|v - v^n\| + \overline{\lim}\, dist(v^n, Tu^n) + \\ + \overline{\lim}\, dist(Tu^n, Tu) = \overline{\lim}\, dist(v^n, Tu^n),$$

так как $dist(Tu^n, Tu) \to 0$ при $n \to \infty$ (полунепрерывность отображения T сверху), что следует из компактности и замкнутости отображения T. В соответствии с условием $v^n \in Su^n$ и определением S представим v^n в виде $v^n = x^n + y^n$, $x_n \in Tu^n$, $\|y^n\| \le dist(Tu^n, E^\circ)$. Отсюда $dist(v^n, Tu^n) \le \|y^n\| \le dist(Tu^n, E^\circ)$ и

$$dist(v, Tu) \le \overline{\lim}\, dist(Tu^n, E^\circ) \le \\ \le \overline{\lim}\, dist(Tu^n, Tu) + dist(Tu, E^\circ) = dist(Tu, E^\circ),$$

что и требовалось доказать.

3. Из компактности и замкнутости оператора T следует импликация

$$u^n \in cl\,\Omega,\ u^n \to u \Rightarrow \overline{\lim}\, dist(Tu^n, E^\circ) \le dist(Tu, E^\circ). \qquad (11)$$

Установим импликацию

$$u_n \in cl\,\Omega_n,\ u_n \xrightarrow{\mathcal{P}} u \Rightarrow \overline{\lim}\, dist(T_n u_n, E_n^\circ) \le dist(Tu, E^\circ). \qquad (12)$$

Действительно, выберем $N' \subseteq N$ так что

$$\overline{\lim}\, dist(T_n u_n, E_n^\circ) = \lim_{n \in N'} dist(T_n u_n, E_n^\circ).$$

Ввиду компактности множества $T_n u_n$ и конечномерности E_n° найдутся такие $v_n \in T_n u_n$ и $v_n^\circ \in E_n^\circ$, что $dist(T_n u_n, E_n^\circ) = \|v_n - v_n^\circ\|$. Последовательность (v_n) $\mathcal{P}$-компактна в силу (7), а последовательность (v_n°) в силу построения подпространств $E_n^\circ \subseteq E_n$; если $v_n \xrightarrow{\mathcal{P}} v$, $v_n^\circ \xrightarrow{\mathcal{P}} v^\circ$ $(n \in N'' \subseteq N)$, то $v \in Tu$ в силу (8) и $v^\circ \in E^\circ$, причем $\|v - v^\circ\| \le dist(v, E^\circ)$ — в противном случае при больших n приходим к противоречию с определением v_n° как ближайшего к v_n элемента из E_n°. Отсюда следует (12):

$$\overline{\lim}\, dist(T_n u_n, E_n^\circ) = \lim_{n \in N''} \|v_n - v_n^\circ\| = \|v - v^\circ\| \le dist(v, E^\circ) \le dist(Tu, E^\circ).$$

4. Покажем, что отображения T и S линейно гомотопны на $\partial\Omega$ и, значит,

$$\gamma(I - T; \partial\Omega) = \gamma(I - S; \partial\Omega). \qquad (13)$$

Действительно, если для некоторых $u \in cl\,\Omega$ и $\lambda \in [0, 1]$ имеем $u \in \lambda Tu + (1-\lambda) Su$, то u представим в виде

$$u = \lambda v + (1-\lambda)(v' + z), \quad v, v' \in Tu, \quad \|z\| \le dist(Tu, E^\circ),$$

поэтому

$$\operatorname{dist}(u, Tu) \le \|z\| \le \operatorname{dist}(Tu, E^\circ) \le \varepsilon = \frac{\alpha}{2},$$

и $u \bar{\in} \partial\Omega$ (см. определение α в (10)).

5. Покажем, что при достаточно больших n отображения T_n и S_n линейно гомотопны на $\partial\Omega_n$ и, значит,

$$\gamma(I - T_n; \partial\Omega_n) = \gamma(I - S_n; \partial\Omega_n) \quad (n \ge n_2). \tag{14}$$

Рассуждая от противного, допустим что для некоторых $u_n \in \partial\Omega_n$, $\lambda_n \in [0,1]$ имеем

$$u_n \in \lambda_n T_n u_n + (1 - \lambda_n) S_n u_n \quad (n \in N' \subseteq N).$$

Выберем $v_n, v_n' \in T_n u_n$ таким образом, что

$$u_n = \lambda_n v_n + (1 - \lambda_n)(v_n' + z_n), \quad \|z_n\| \le \operatorname{dist}(T_n u_n, E_n^\circ).$$

Последовательности (v_n) и (v_n') $\mathcal{P}$-компактны по условию (7) последовательность $(v_n' + z_n)$ $\mathcal{P}$-компактна в силу того, что $v_n' + z_n \in S_n u_n \subset E_n^\circ$ $(n \in N')$. Значит, последовательности (u_n) и (z_n) тоже $\mathcal{P}$-компактны. Пусть

$$\lambda_n \to \lambda, \; u_n \xrightarrow{\mathcal{P}} u, \; v_n \xrightarrow{\mathcal{P}} v, \; v_n' \xrightarrow{\mathcal{P}} v', \; z_n \xrightarrow{\mathcal{P}} z \quad (n \in N'' \subseteq N').$$

Тогда $\lambda \in [0,1]$, $u \in \partial\Omega$, $v, v' \in Tu$ (см. (6) и (8)),

$$u = \lambda v + (1 - \lambda) v' + (1 - \lambda) z.$$

Отсюда заключаем, что (см. (12))

$$\operatorname{dist}(u, Tu) \le (1 - \lambda)\|z\| \le \underline{\lim} \|z_n\| \le \overline{\lim} \operatorname{dist}(T_n u_n, E_n^\circ) \le$$
$$\le \operatorname{dist}(Tu, E^\circ) \le \varepsilon = \frac{\alpha}{2},$$

вопреки (10). Противоречие доказывает утверждение.

6. Отображения $S: cl\,\Omega \to \mathcal{K}(E)$ и $S_n: cl\,\Omega_n \to \mathcal{K}(E_n)$ можно рассматривать и как отображения со значениями в $\mathcal{K}(E^\circ)$ и $\mathcal{K}(E_n^\circ)$ соответственно. По известным свойствам вращения

$$\gamma(I - S; \partial\Omega) = \gamma(I - S^\circ; \partial\Omega^\circ), \quad \gamma(I - S_n; \partial\Omega_n) = \gamma(I - S_n^\circ; \partial\Omega_n^\circ), \tag{15}$$

где $\Omega^\circ = \Omega \cap E^\circ$, $\partial\Omega^\circ$ - граница Ω° в E°, $\Omega_n^\circ = \Omega_n \cap E_n^\circ$, $\partial\Omega_n^\circ$ - граница Ω_n° в E_n°, а

$$S^\circ: cl\,\Omega^\circ \to \mathcal{K}(E^\circ), \quad S_n^\circ: cl\,\Omega_n^\circ \to \mathcal{K}(E_n^\circ)$$

являются сужениями отображений S и S_n.

По известным свойствам вращения имеем также

$$\gamma(I - S_n^\circ; \partial\Omega_n^\circ) = \gamma(I - \varphi_n^{-1} S_n^\circ \varphi_n; \partial(\varphi_n^{-1}\Omega_n^\circ)) \quad (n \ge n_1). \tag{16}$$

7. Дополнив цепочку равенств (13)-(16) равенством

$$\gamma(I - \varphi_n^{-1} S_n^\circ \varphi_n; \partial(\varphi_n^{-1}\Omega_n^\circ)) = \gamma(I - S^\circ; \partial\Omega^\circ) \quad (n \ge n_3), \tag{17}$$

приходим к утверждению (9) доказываемой теоремы.

Для доказательства равенств (I7) достаточно установить, что, во-первых, отображения $\varphi_n^{-1}S_n^\circ\varphi_n$ и S° линейно гомотопны на $\partial(\Omega^\circ\cap\varphi_n^{-1}\Omega_n^\circ)$, во-вторых, $\varphi_n^{-1}S_n^\circ\varphi_n$ не имеет неподвижных точек в $cl(\varphi_n^{-1}\Omega_n^\circ)\setminus\Omega^\circ$ и, в-третьих, S° не имеет неподвижных точек в $cl\,\Omega^\circ\setminus\varphi_n^{-1}\Omega_n^\circ$. Эти три утверждения доказываются по единой схеме, и мы ограничимся подробным доказательством наиболее сложного из них – первого. Рассуждая от противного, допустим, что для некоторых $u^n\in\partial(\Omega^\circ\cap\varphi_n^{-1}\Omega_n^\circ)$ и $\lambda_n\in[0,1]$ имеем

$$u^n\in\lambda_n S^\circ u^n+(1-\lambda_n)\varphi_n^{-1}S_n^\circ\varphi_n u^n \quad (n\in N'\subseteq N).$$

Тогда

$$u^n=\lambda_n(v^n+y^n)+(1-\lambda_n)\varphi_n^{-1}(v_n'+z_n) \quad (n\in N'),$$

где

$$v^n+y^n\in S^\circ u^n;\quad v^n\in Tu^n;\quad \|y^n\|\le dist(Tu^n,E^\circ),$$

$$v_n'+z_n\in S_n^\circ\varphi_n u^n,\quad v_n'\in T_n\varphi_n u^n;\quad \|z_n\|\le dist(T_n\varphi_n u^n,E_n^\circ).$$

Последовательности (v^n+y^n) и (v^n) компактны в силу компактности отображений T и S, последовательности (v_n') и $(v_n'+z_n)$ $\mathcal{P}$-компактны в силу (7) и свойств E_n° $(n\in N)$. Пусть

$$\lambda_n\to\lambda,\ u^n\to u,\ v^n\to v,\ y^n\to y,\ v_n'\xrightarrow{\mathcal{P}}v',\ z_n\xrightarrow{\mathcal{P}}z \quad (n\in N''\subseteq N').$$

Тогда $v\in Tu$ в силу замкнутости T, $v'\in Tu$ в силу (8),

$$u=\lambda v+(1-\lambda)v'+\lambda y+(1-\lambda)z,$$

откуда

$$dist(u,Tu)\le\lambda\|y\|+(1-\lambda)\|z\|=\lambda\lim\|y^n\|+(1-\lambda)\lim\|z_n\|\le$$
$$\le\lambda\,\overline{\lim}\,dist(Tu^n,E^\circ)+(1-\lambda)\overline{\lim}\,dist(T_n\varphi_n u_n,E_n^\circ)\le dist(Tu,E^\circ);$$

на последнем шаге мы воспользовались неравенствами (II) и (I2). Итак,

$$dist(u,Tu)\le dist(Tu,E^\circ)\le\varepsilon=\frac{\alpha}{2},$$

поэтому $u\,\overline{\in}\,\partial\Omega$ (см. (I0)). С другой стороны, из включений $u^n\in\partial(\Omega^\circ\cap\varphi_n^{-1}\Omega_n^\circ)$ на основании условия (6) следует, что $u\in\partial\Omega$. Противоречие доказывает наше утверждение о линейной гомотопности. При доказательстве двух оставшихся утверждений учтем, что

$$u^n\in cl(\varphi_n^{-1}\Omega_n^\circ)\setminus\Omega^\circ,\quad u^n\to u\Rightarrow u\in\partial\Omega;$$

$$u^n\in cl\,\Omega^\circ\setminus\varphi_n^{-1}\Omega_n^\circ,\quad u^n\to u\Rightarrow u\in\partial\Omega.$$

Эти соотношения вытекают из (4)-(6).

8. Итак, равенство (9) установлено. Остальные утверждения теоремы 2 элементарны и устанавливаются по той же схе-

ме, что и в доказательстве теоремы I.

Доказательство теоремы 2 завершено.

Литература

I. Борисович Ю.Г., Гельман Б.Д., Мышкис А.Д., Обуховский В.В., Топологические методы в теории неподвижных точек многозначных отображений. Успехи матем. наук, I980, № I,59-I26.

2. Вайникко Г.М., Компактная аппроксимация операторов и приближенное решение уравнений. Тарту, I970.

3. Вайникко Г., О приближении неподвижных точек вполне непрерывных операторов. Уч. зап. Тартуск. ун-та, I974, 342, 225-236.

4. Вайникко Г., Анализ дискретизационных методов. Тарту, I976.

5. Красносельский М.А., Забрейко П.П., Геометрические методы нелинейного анализа. Москва, I975.

6. Vainikko, G., Approximative methods for nonlinear equations (two approaches to the convergence problem). Nonlinear Analysis, 1978, 2, N 6, 647-687.

Поступило
9 II I982

INVARIANCE OF FIXED POINT INDEX BY APPROXIMATION OF MULTIVALUED MAPS

G.Vainikko

Summary

We establish the invariance of the fixed point index and the convergence of fixed points by the compact approximation of compact closed convex-valued maps. An elementary theorem 1 is based on the linear homotopy of maps. The main result is contained in theorem 2 and is concerned with a simultaneous approximation of the basic Banach space.

Уч.зап.Тартуск.ун-та,1982,633,II-I8

КОМПАКТНАЯ СХОДИМОСТЬ ПРИ АППРОКСИМАЦИИ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ В БАНАХОВОМ ПРОСТРАНСТВЕ

С. Пискарёв

Рассмотрим в банаховом пространстве E равномерно корректно поставленную задачу Коши

$$u''(t)=Au(t)\,,\quad u(0)=u^0\,,\quad u'(0)=u^1,\quad t\in R_+=[0,\infty) \qquad (\mathrm{I})$$

решение которой записывается как

$$u(t)=C(t)u^0+S(t)u^1$$

где $C(t)$ - косинус, а $S(t)$ - синус оператор-функции соответственно (сокращенно КОФ и СОФ) (см. [8,I3,I4]). В [I4] установлено, что $S(t)$ компактна при любом $t\in R=(-\infty,\infty)$ тогда и только тогда, когда резольвента $R(\lambda;A)=(\lambda-A)^{-1}$ компактна при любом $\lambda\in\rho(A)$. Известно (см. [9]), что компактная сходимость $S_n(t)\to S(t)$ имеет место в том и только в том случае, когда $R(\lambda;A_n)\to R(\lambda;A)$ компактно. В данной статье мы частично отвечаем на вопрос П.Оя: при каких условиях одновременно имеют место компактные сходимости $S_n(t)\to S(t)$ и $C_n(t)-J_n\to C(t)-J$, $t\in R$?

§ I. Подготовительные сведения

В данной работе будем рассматривать аппроксимацию задачи (I) задачами Коши

$$u_n''(t)=A_nu_n(t)\,,\quad u_n(0)=u_n^0\,,\quad u_n'(0)=u_n^1\,,\quad t\in R_+\,, \qquad (2)$$

заданным в банаховых пространствах E_n. Предполагается, что операторы A_n генерируют КОФ $C_n(t)$, а пространства E и E_n связаны непрерывными линейными отображениями $p_n: E\to E_n$ со свойством

$$\|p_nx\|\to\|x\| \quad \text{для любого} \quad x\in E\,.$$

Подробные сведения о сходимостях на указанной аппроксимационной схеме можно найти в [I,7].

Параллельно с исследованием (I) и (2) мы будем формулировать результаты о аппроксимации задачи Коши

$$u'(t)=Au(t), \quad u(0)=u^0\in E, \quad t\in R_+ \tag{3}$$

с оператором A, генерирующим C_0-полугруппу $exp(tA)$, задачами Коши

$$u_n'(t)=A_n u_n(t), \quad u_n(0)=u_n^0\in E_n, \quad t\in R_+, \tag{4}$$

где A_n также генерируют C_0-полугруппы $exp(tA_n)$. Дискретные полугруппы $\mathcal{N}_n(t)$ и $\mathcal{M}_n(t)$ определяют соответственно явный и неявный методы решения (4), а именно (см. [5,7]) ($k=0,1,\ldots$)

$$\mathcal{N}_n(k\tau+\tau)-\mathcal{N}_n(k\tau)=\tau A_n\mathcal{N}_n(k\tau), \quad \mathcal{N}_n(0)=u_n^0\in E_n \tag{5}$$

$$\mathcal{M}_n(k\tau+\tau)-\mathcal{M}_n(k\tau)=\tau A_n\mathcal{M}_n(k\tau+\tau), \quad \mathcal{M}_n(0)=u_n^0\in E_n \tag{6}$$

Если пространства $E=H$ и $E_n=H_n$ - гильбертовы, а операторы A и A_n - самосопряженные неположительные, то для дискретизации (2) будет изучаться схема ($k=0,1,\ldots$)

$$\mathcal{U}_n(k\tau+\tau)-2\mathcal{U}_n(k\tau)+\mathcal{U}_n(k\tau-\tau)=\tau^2 A_n\mathcal{U}_n(k\tau),$$

$$\mathcal{U}_n(0)=u_n^0, \quad \mathcal{U}_n(\tau)=\mathscr{C}_n(\tau)u_n^0+\mathfrak{S}_n(\tau)u_n^1,$$

где $\mathscr{C}_n(t)$ и $\mathfrak{S}_n(t)$ - дискретные косинус и синус оператор-функции (ДКОФ и ДСОФ) соответственно (см. [8]).

Для формулировок результатов нам понадобятся:

Определение I. Последовательность ограниченных оператор-функций $B_n(t)$ называется равностепенно сходящейся к ограниченной оператор-функции $B(t)$ если для любого t имеем $B_n(t)\to B(t)$ и, более того, для любого компакта $[0,T]$ и любого $\varepsilon>0$ можно указать такие числа N и $\delta>0$, что при $n\geq N$ и $|t_1-t_2|\leq\delta$, $t_1,t_2\in[0,T]$ выполняется

$$\|B_n(t_1)-B_n(t_2)\|\leq\varepsilon.$$

Определение 2. Последовательность ограниченных оператор-функций $B_n(t)$ называется дискретно равностепенно сходящейся к ограниченной оператор-функции $B(t)$ с шагом по времени τ_n, если $B_n(t)\to B(t)$ равномерно на любом компакте и для произвольного числа $T>0$ найдется такая константа $M>0$, что

$$\|B_n(t+\tau_n)-B_n(t)\|\leq M\tau_n \text{ при } t\in[0,T].$$

§ 2. Основные результаты

Теорема I. Пусть оператор A является инфинитезимальным генератором КОФ $C(t)$, а $S(t)$ - ассоциированная СОФ. Следующие условия эквивалентны:

(i) оператор $C(t) - J$ компактен при любом $t \geq 0$;
(ii) оператор $S(t) - tJ$ компактен при любом $t \geq 0$;
(iii) оператор A компактен.

Замечание I. Если операторы $S(t)$ и $C(t) - J$ одновременно компактны при любом $t \in [0, \infty)$ то пространство E с необходимостью конечномерно. Это утверждение следует из того факта, что компактными по предположению должны быть $R(\lambda; A)$ и $\lambda R(\lambda; A) - J$ и, значит, единичный оператор тоже компактен.

Замечание 2. Известно (см. [2], стр. II6), что задача (I) с ограниченным оператором A в гильбертовом пространстве имеет ограниченное на всей оси решение $\|u(t)\| \leq c$, $t \in (-\infty, \infty)$ тогда и только тогда, когда A подобен строго отрицательному оператору. Таким образом, при компактности $C(t) - J$, $t \geq 0$, решение задачи (I) не может быть равномерно ограниченным. Свойство компактности оператора $C(t) - J$ не сохраняется при сдвиге $A \to A + \xi J$. Однако, как недавно показал Гольдштейн, сдвигом $A_{\delta} = A + \delta J$ инфинитезимального генератора не всегда можно получить равномерную ограниченность

$$\|C(t)\| \leq const, \quad t \in [0, \infty).$$

Теорема 2. Пусть операторы A_n и A ограничены. Компактная сходимость $A_n \to A$ имеет место тогда и только тогда, когда $C_n(t) - J_n \to C(t) - J$ равностепенно компактно. Это же касается сходимости $S_n(t) - tJ_n \to S(t) - tJ$.

Теорема 3. Пусть операторы A_n и A ограничены. Тогда компактная сходимость инфинитезимальных генераторов $A_n \to A$ необходима и достаточна для того, чтобы $exp(tA_n) - J_n \to$ $\to exp(tA) - J$ равностепенно компактно.

Теорема 4. Пусть операторы A_n и A ограничены. Тогда компактная сходимость инфинитезимальных генераторов $A_n \to A$ необходима и достаточна для того, чтобы $\mathcal{N}_n(t) - J_n \to$ $exp(tA) - J$ дискретно равностепенно компактно. Это же касается сходимости $\mathfrak{M}_n(t) - J_n \to exp(tA) - J$.

Наконец, пусть пространства $E = H$ и $E_n = H_n$, а операторы A_n, A самосопряженные и неположительные.

Теорема 5. Компактная сходимость $A_n \to A$ имеет место тогда и только тогда, когда $C_n(t) - J_n \to C(t) - J$ компактно при $t \in R$.

Теорема 6. Компактная сходимость инфинитезимальных генераторов $A_n \to A$ имеет место тогда и только тогда, когда $\mathcal{C}_n(t) - J_n \to \mathcal{C}(t) - J$ дискретно равностепенно компактно. Это же касается сходимости $G_n(t) - tJ_n \to S(t) - tJ$.

§ 3. Доказательства и обсуждения

Доказательству теоремы I предпошлем две леммы:

Лемма I. Оператор A является вполне непреывным тогда и только тогда, когда оператор $\lambda R(\lambda, A) - J$ вполне непрерывен при достаточно больших $\lambda > \omega$.

Лемма 2. Пусть $\mathcal{C}(t) - J$ компактный оператор при любом $t \in R$. Тогда спектр $\delta(A)$ ограничен.

Доказательство. Как показано в [13] при $t > 0$ имеет место равенство

$$ch(t\sqrt{P\delta(A)}) = P\delta(\mathcal{C}(t)). \tag{7}$$

При этом множества $C\delta(A)$ и $R\delta(A)$ отображаются функцией $ch\, z = 1/2(exp(-z) + exp(z))$ соответственно в $C\delta(\mathcal{C}(t))$ и $R\delta(\mathcal{C}(t))$. Учитывая, что $\delta(A)$ лежит в параболе, направленной влево, и симметричной относительно оси абсцисс, а также, что $\mathcal{C}(t) - J$ компактен находим $\delta(A) = P\delta(A)$ и

$$0 \leq Re\,\lambda \leq \omega, \quad \lambda \in \sqrt{P\delta(A)}.$$

Поэтому для доказательства леммы достаточно установить ограниченность множества $\{Jm\, z : z \in \sqrt{P\delta(A)}\}$. Предположим от противного существование последовательности $\{a_n\} \subset \{Jm\,\lambda : \lambda \in \sqrt{P\delta(A)}\}$ такой, что

I^0 $\quad a_n \to \infty$ при $n \to \infty$.

Так как оператор $C(t) - J$ компактен, и, следовательно, $P\delta(C(t))$ имеет точкой сгущения возможно лишь 1, а функция $ch(ix)$ является 2π-периодической при вещественном x, то, учитывая (7), получаем

2^0 для любого $\varepsilon > 0$ и любого $t > 0$ только конечное число элементов последовательности $\{a_n t\}$ более чем на ε отличается от целых чисел, умноженных на 2π.

Покажем, что такой $\{a_n\}$, обладающей свойствами I^0 и 2^0, не существует. Положим

$$G_\varepsilon = \bigcup_{n=1}^{\infty} [2\pi n - \varepsilon, 2\pi n + \varepsilon], \quad \varepsilon \in (0, 1/2).$$

Для любого $t > 0$ в силу 2^0 должен найтись номер N, такой что

$t \in \bigcap_{k=N}^{\infty} G_\varepsilon a_k^{-1}$, поэтому $(0,\infty) \subseteq \bigcup_{N=1}^{\infty} \bigcap_{k=N}^{\infty} G_\varepsilon a_k^{-1}$.

По теореме Бэра о категориях (см. [4], стр. 24) существует число N_0, для которого $\bigcap_{k=N_0}^{\infty} G_\varepsilon a_k^{-1}$ содержит некоторый интервал $[a, b]$ положительной меры. Однако, последнее не возможно ввиду $2\varepsilon a_n^{-1} < (b-a)$ при достаточно больших n. Полученное противоречие доказывает лемму. Лемма доказана.

Доказательство теоремы I проведем по схеме $(i) \Rightarrow (ii) \Rightarrow$ $\Rightarrow (iii) \Rightarrow (i)$. Установим $(i) \Rightarrow (ii)$. Для вещественных $\alpha > 0$ и $\rho > 0$ обозначим через $S(\alpha, \rho)$ множество комплексных чисел $\lambda \in \mathbb{C}$ со свойствами $Re\,\lambda \le \alpha$, $|\alpha - \lambda| = \rho$. В силу леммы 2 спектр $P\sigma(A)$ ограничен: $|\sqrt{P\sigma(A)}| \le c_1$. Определим ограниченный оператор $\tilde{C}(t)$ по формуле

$$\tilde{C}(t)x = 1/(2\pi i)\int_{\Gamma_1} exp(\lambda t)\lambda R(\lambda^2; A)x\,d\lambda\,, \qquad x \in E, \tag{8}$$

где $\Gamma_1 = \Gamma_2 \cup \Gamma_3$, $\Gamma_2 = \{\lambda = \alpha \pm i\beta : \alpha = \omega,\ 0 \le \beta \le c_1^2 + 2\omega\}$, $\Gamma_3 = \{\lambda : \lambda \in S(\omega, c_1^2 + 2\omega)\}$, ω – число из доказательства леммы 2. В то же время для любого $x \in \mathcal{D}(A)$ имеем (см. [3], стр. I9)

$$C(t)x = 1/(2\pi i)\cdot\int_{\Gamma_4} exp(\lambda t)\lambda R(\lambda^2; A)x\,d\lambda\,,$$

где $\Gamma_4 = \{\lambda = \alpha + i\beta :\ \alpha = \omega,\ -\infty < \beta < \infty\}$. Так как все полюсы резольвенты $R(\lambda^2; A)$ лежат внутри Γ_1, то по теореме о вычислении интеграла Меллина (см. [I0], стр. 364) с помощью теории вычетов, находим

$$C(t)x = \tilde{C}(t)x \text{ для любого } x \in \mathcal{D}(A).$$

Множество $\mathcal{D}(A)$ плотно в E, поэтому $C(t) = \tilde{C}(t)$. Таким образом СОФ $C(t)$ оказалась непрерывной в равномерной операторной топология, а, следовательно, равномерно непрерывной на любом компакте $[\delta, T]$ вещественной прямой. Мы завершим доказательство первого шага $(i) \Rightarrow (ii)$, заметив, что $S(t) - tJ = \int_0^t (C(s) - J)\,ds$, а интеграл справа является пределом в равномерной операторной топологии риманоских сумм компактных операторов. Импликация $(ii) \Rightarrow (iii)$ устанавливается с использованием выражения

$$R(\lambda^2; A) - J/\lambda^2 \cdot J = \int_0^\infty exp(-\lambda t)(S(t) - tJ)\,dt, \quad \lambda > \omega\,,$$

и леммы I. Напомним, что $S(t)$ равномерно непрерывна на компакте и $\|S(t)\| \le M exp(\omega t)$.

Наконец, пусть A – вполне непрерывный. Принимая во внимание равенство

$$C(t)=\sum_{k=0}^{\infty} t^{2k}A^k/(2k)!\,,\quad t\in R, \tag{9}$$

получаем компактность оператора $C(t)-J$ как предел по норме сумм компактных операторов. Теорема доказана.

Предложение I. Пусть оператор $C(t)-J$ компактен для любого $t\in R$. Тогда найдется такое $\delta>0$, что при $t\in[0,\delta]$ существуют ограниченные обратные $C^{-1}(t)$ и $S^{-1}(t)$.

Доказательство. Обратимость $C(t)$ сразу вытекает из (9) т.к. $\|C(t)-J\|=O(t^2)$ при $t\to 0$. Для $S(t)$ утверждение следует из $\|S(t)-tJ\|=O(t^3)$, т.к. оператор $S(t)-tJ$ является близким к обратиму tJ. Предложение доказано.

Если для некоторого $T>0$ оператор $S^{-1}(T)$ существует и ограничен, то в силу $C(2T)-C(0)=2AS^2(T)$ вытекает ограниченность оператора A. В [I2] изучается задача для уравнения

$$u''(t)=Au'(t)+Bu(t)+f(t) \tag{I0}$$

с условиями $u(0)=0,\ u(T)=0$. При этом предполагается, что $1/4A^2+B$ генерирует КОФ и существует ограниченный обратный $S^{-1}(t)$. Из равенства $C(t+h)-C(t-h)=2AS(t)S(h)$ видно, что предположение непрерывности $S^{-1}(T)$ весьма ограничительно (в случае компактности резольвенты $R(\lambda;A)$ оно вообще не выполняется). Естественный подход в изучении уравнения (I0) указан, например, в [3,6].

Доказательство теоремы 2. Для того, чтобы установить компактную сходимость $C_n(t)-J_n\to C(t)-J$, достаточно показать, что для любой последовательности $\{x_n\}$, $\|x_n\|=1$ имеем $\mu(((C_n(t)-J_n)x_n))=0$. Последнее утверждение вытекает из неравенства (μ -мера некомпактности)

$$\mu(((C_n(t)-J_n)x_n))\le\sum_{k=1}^{N}\mu((A_n^k x_n))t^{2k}/(2k)!+\mu\Big(\Big(\sum_{k=N+1}^{\infty}t^{2k}A_n^k x_n/(2k)!\Big)\Big),$$

где за счет выбора числа N сколь угодно большим, но конечным, правая часть становится сколь угодно малой. Равностепенность семейства $\{C_n(t)-J_n\}$ следует из ограниченности $\{A_n\}$, $\|A_n\|=O(1)$. Обратно. Пусть $C_n(t)-J_n\to C(t)-J$ равностепенно компактно. Тогда в силу равенства $\lambda R(\lambda^2;A)x=\int_0^{\infty}exp(-\lambda t)C(t)x\,dt$, $x\in E$, имеем

$$\lambda R(\lambda;A_n)-J_n \to \lambda R(\lambda;A)-J \text{ компактно.} \qquad (II)$$

Поэтому $\lambda R(\lambda;A_n) \to \lambda R(\lambda;A)$ собственно и, учитывая, что при $\lambda \neq 0$, $\lambda \notin \delta(A)$ имеем $N(\lambda R(\lambda;A)) = \{0\}$, получим (см. [I], стр. 35) равномерную ограниченность $\{A_n\}$, т.е. $\|A_n\| = \mathcal{O}(1)$. Для доказательства компактной сходимости $A_n \to A$ остается воспользоваться равенством $\lambda R(\lambda;A_n)-J_n = A_n R(\lambda;A_n)$ и утверждением (II). Теорема доказана.

<u>Доказательство теоремы 5.</u> Установим лишь, что из компактной сходимости $C_n(t)-J_n \to C(t)-J$ следует равномерная ограниченность $\|A_n\| = \mathcal{O}(1)$. Спектр оператора $C(0)=J$ состоит из единственной точки I. В силу непрерывности $C(t)$ по норме найдется $\delta>0$ такое, что $\delta(C(t))$ при $t\in[0,\delta]$ лежит в правой полуплоскости (см. [5], стр. 264). Поэтому, учитывая компактную сходимость $C_n(t)-J_n \to C(t)-J$ заключаем, что спектр $\delta(C_n(t))$, $t\in[0,\delta]$, не может иметь точки сгущения в левой полуплоскости. Предположим неограниченность последовательности $\|A_n\| \to \infty$. Мы придем к противоречию, показав, что спектр $\delta(C_n(t_n))$ аппроксимирует точку -1 при $t_n \in [0,\delta]$. Обозначим $\beta_n = \sqrt{\lambda_n}$; $\lambda_n = \|A_n\|$. Справедливо представление

$$\beta_n = 2\pi k(n) + \alpha_n, \quad k(n) \text{ - целое}, \quad 0 \le \alpha_n < 2\pi.$$

Выбрав $t_{k(n)}$ из соотношения $k(n)t_{k(n)} = [k(n)\delta/10] - \alpha_n t_{k(n)}/2\pi + 1/2$, где $[p]$ - целая часть числа p, получим $ch(it_{k(n)}\beta_n) \to -1$, $t_{k(n)} \to \delta/10$. Теорема доказана.

Литература

I. В а й н и к к о Г., Анализ дискретизационных методов. Тарту, 1976.

2. Д а л е ц к и й Ю.Л., К р е й н М.Г., Устойчивость решений дифференциальных уравнений в банаховом пространстве. Москва, 1970.

3. К р е й н С.Г., Линейные дифференциальные уравнения в банаховом пространстве. Москва, 1967.

4. И о с и д а К., Функциональный анализ. Москва, 1967.

5. К а т о Т., Теория возмущений линейных операторов. Москва, 1972.

6. М е л ь н и к о в а И.В., М о р н о в а В.М., О задаче Коши и методе квазиобращения для уравнений второго

порядка в банаховом пространстве. Дифф. уравнения, 1979, 15, 4, 614-618.

7. П и с к а р ё в С., Об аппроксимации голоморфных полугрупп. Уч. зап. Тартуск. ун-та, 1979, 492, 3-14.

8. П и с к а р ё в С., Дискретизация абстрактного гиперболического уравнения. Уч. зап. Тартуск. ун-та, 1979, 500, 3-23.

9. П и с к а р ё в С., Решение неоднородного абстрактного линейного гиперболического уравнения. Уч. зап. Тартуск. ун-та, 1979, 500, 24-32.

10. П ч е л и н Б.К., Специальные разделы высшей математики, Москва, 1973.

11. C u t h b e r t, J.R., On semigroups such that $T_t - J$ is compact for some $t>0$. Z. Wahrs., 1971, 18, 1-2, 9-16.

12. H e i m e s, K.A., Green's functions for linear second order systems. SIAM J.Math. Anal., 1978, 9, 1, 207-214.

13. N a g y, B., On cosine operator functions in Banach spaces. Acta Sci. Math., 1974, 36, 3-4, 281-289.

14. T r a v i s, C.C., W e b b, G.F., Compactness, regularity, and uniform continuity properties of strongly continuous cosine families. Houston J. Math., 1977, 3, 4, 555-567.

Поступило
21 I 1981

COMPACT APPROXIMATION OF DIFFERENTIAL EQUATIONS IN BANACH SPACE

S.Piskarjov

Summary

Properties of the infinitesimal generator of the cosine operator function $C(t)$ where $C(t)-J$ is a compact operator are investigated. In addition, compact approximation of the generator which gives rise to such cosine families is studied.

Уч.зап.Тартуск.ун-та,1982,633,19-28

АППРОКСИМАЦИЯ В ПРОБЛЕМЕ СОБСТВЕННЫХ ЗНАЧЕНИЙ С ГОЛОМОРФНОЙ ЗАВИСИМОСТЬЮ ОПЕРАТОРА ОТ ПАРАМЕТРА (II)

О. Карма

Рассматривается аппроксимация проблемы собственных значений $A(\lambda)u=0$ для голоморфной фредгольмовой оператор-функции A. В предположении регулярной аппроксимации исследуется сходимость жордановых цепочек. Основной результат статьи – теорема 6.

§ 1. Предположение регулярной аппроксимации

Пусть $\mathcal{J}$ – основное множество индексов – некоторое фиксированное направленное множество, а $\mathcal{J}', \mathcal{J}'', \mathcal{J}_1, \ldots$ – его конфинальные подмножества ($\forall i\in \mathcal{J}\ \ \exists i'\in\mathcal{J}': i'\geq i$). Семейства вида $(\alpha_i,\ i\in\mathcal{J}')$ будем называть (под)последовательностями. Для числовой последовательности $(\alpha_i,\ i\in\mathcal{J}')$ запись $\alpha_i\to\alpha$ $(i\in\mathcal{J}')$ означает сходимость в следующем смысле:

$$\forall\varepsilon>0\quad \exists i(\varepsilon)\in\mathcal{J}':\quad i\geq i(\varepsilon),\, i\in\mathcal{J}'\Rightarrow|\alpha_i-\alpha|<\varepsilon.$$

Пусть, далее, $\mathcal{U},\mathcal{V},\mathcal{X}_i,\mathcal{Y}_i$ ($i\in\mathcal{J}$) – комплексные банаховы пространства; соответствующие строчные буквы используются ниже только для обозначения элементов этих пространств $(u,u',\ldots\in\mathcal{U},\ x_i,x_i',\ldots\in\mathcal{X}_i,\ \ldots)$. Во всей этой статье мы будем предполагать, что на некоторой области (открытом звязном подмножестве) $\Lambda\subseteq\mathbb{C}$ заданы оператор-функций[1] $A:\Lambda\to\mathcal{B}(\mathcal{U},\mathcal{V})$ и $B_i:\Lambda\to\mathcal{B}(\mathcal{X}_i,\mathcal{Y}_i)$, и фиксированы некоторые семейства "связывающих" отображений ($p_i:\mathcal{U}\to\mathcal{X}_i,\ i\in\mathcal{J}$) и ($q_i:\mathcal{V}\to\mathcal{Y}_i,\ i\in\mathcal{J}$), причем выполнены следующие требования 1°–9°:

1° p_i ($i\in\mathcal{J}$) определены на всем $\mathcal{U}$ и q_i ($i\in\mathcal{J}$) – на всем $\mathcal{V}$;

2° $\|p_iu\|\to\|u\|\ (i\in\mathcal{J})\ \forall u\in\mathcal{U},\ \|q_iv\|\to\|v\|\ (i\in\mathcal{J})\ \forall v\in\mathcal{V}$;

3° p_i и q_i $(i\in\mathcal{J})$ – "асимптотически линейны" [14$_{\mathrm{I}}$];

4° A и B_i $(i\in\mathcal{J})$ – голоморфны на Λ [5];

[1] Через $\mathcal{B}(\cdot,\cdot)$ обозначено банахово пространство линейных ограниченных операторов.

5^0 $A(\lambda)$ и $B_i(\lambda)$ $(i\in J, \lambda\in\Lambda)$ – фредгольмовы с индексом 0;

6^0 множество $\rho(A)=\{\lambda\in\Lambda : \exists A^{-1}(\lambda)\in\mathcal{B}(\mathcal{V},\mathcal{U})\}$ не пусто;

7^0 $B_i(\lambda)$ равномерно ограничены по норме на каждом компакте $\Lambda_0\subset\Lambda$: $\sup\{\|B_i(\lambda)\| : \lambda\in\Lambda_0, i\in J\}\le C(\Lambda_0)<\infty$;

8^0 для всех $\lambda\in\Lambda$ последовательность операторов ($B_i(\lambda)$, $i\in J$) сходится к $A(\lambda)$:

$$\|x_i-p_i u\|\to 0 \ (i\in J) \Rightarrow \|B_i(\lambda)x_i-q_i A(\lambda)u\|\to 0 \ (i\in J);$$

9^0 для всех $\lambda\in\Lambda$ последовательность операторов ($B_i(\lambda)$, $i\in J$) регулярна: $\|x_i\|\le 1$, $\|B_i(\lambda)x_i-q_i v\|\to 0 \ (i\in J')\Rightarrow$

$$\Rightarrow \exists u\in\mathcal{U}, J''\subseteq J' : \|x_i-p_i u\|\to 0 \ (i\in J'').$$

Требования 7^0–9^0 мы назовем требованиями регулярной аппроксимации A последовательностью (B_i , $i\in J$) на $\Lambda\subseteq\mathbb{C}$. (Дискретную) сходимость и (дискретную) компактность последовательностей (x_i, $i\in J$) с $x_i\in\mathcal{X}_i$ и (y_i, $i\in J$) с $y_i\in\mathcal{Y}_i$ будем понимать, как в [14_I] :

$$x_i\to u \ (i\in J') \Leftrightarrow \|x_i-p_i u\|\to 0 \ (i\in J'),$$

$$(x_i, i\in J')\text{-комп.} \Leftrightarrow \forall J''\subseteq J' \ \exists J'''\subseteq J'', u\in\mathcal{U}: x_i\to u \ (i\in J''').$$

Замечание I. В [6] схема регулярной аппроксимации (там: y_0– аппроксимации) фактически применяется в случае, когда все $\mathcal{X}_i$ и $\mathcal{Y}_i$ являются изоморфными или подпространствам или фактор-пространствам пространств $\mathcal{U}$ и $\mathcal{V}$, соответственно.

Отметим некоторые нужные нам следствия из предположений 1^0–9^0, вытекающие несложно из следующих предложений I и 2.

Предложение I [5]. Для голоморфной функции со значениями в банаховом пространстве, и для ее производных, имеет место интегральная формула Коши :

$$f^{(j)}(\lambda)=j!\,(2\pi i)^{-1}\int_\Gamma (\lambda-\lambda')^{-j-1}f(\lambda')\,d\lambda'.$$

Предложение 2 [14_{II}]. Пусть $\Gamma\subset\mathbb{C}$–спрямляемая жорданова кривая, а ($w_i : \Gamma\to F_i$, $i\in J'$) – семейство равностепенно непрерывных на Γ функций таких, что $w_i(\lambda)\to w(\lambda)\in\mathcal{V}$ ($i\in J'$) при каждом $\lambda\in\Gamma$. Тогда $\int_\Gamma w_i(\lambda)d\lambda\to\int_\Gamma w(\lambda)d\lambda$ $(i\in J')$, если только все эти интегралы существуют.

Следствие I. При любом $j=1,2,\ldots$ производные $B^{(j)}(\lambda)$ равномерно ограничены по норме на каждом компакте из Λ .

Следствие 2. На каждом компакте $\Lambda_0\subset\Lambda$ семейство (B_i, $i\in J$) равностепенно непрерывно по норме:

$$\forall\varepsilon>0 \ \exists\delta>0 : \lambda,\lambda'\in\Lambda_0, |\lambda-\lambda'|<\delta\Rightarrow\|B_i(\lambda)-B_i(\lambda')\|<\varepsilon \quad \forall i\in J.$$

Следствие 3. При любом $j=0,1,\ldots$ имеем :

$$\lambda_i\in\Lambda,\ \lambda_i\to\lambda\in\Lambda,\ x_i\to u\ (i\in J')\Rightarrow B_i^{(j)}(\lambda_i)x_i\to A^{(j)}(\lambda)u\ \ (i\in J').$$

Следствие 4. Для любой последовательности $\lambda_i\to\lambda\in\Lambda$ $(i\in J)$ с $\lambda_i\in\Lambda$ последовательность операторов $(B_i(\lambda_i),\ i\in J)$ регулярна: $\|x_i\|\le C,\ (B_i(\lambda_i)x_i,\ i\in J)$-*комп.* $\Rightarrow (x_i,\ i\in J)$-*комп.*

§ 2. Понятие жордановой цепочки

Точка $\lambda_0\in\Lambda$ называется собственным значением для A и элемент u^0 соответствующим ему собственным элементом, если $A(\lambda^0)u^0=0,\ u^0\neq 0$. В наших предположениях спектр $\sigma(A)=$ $=\{\lambda\in\Lambda:\ \neg\exists A^{-1}(\lambda)\in\mathcal{B}(V,U)\}$ оператор-функции A состоит только из собственных значений, которых в каждом компакте $\Lambda_0\subset\Lambda$ не более чем конечное число [4]. При этом для всех $\lambda_0\in\sigma(A)$ размерность ядра $N(A,\lambda_0)=N(A(\lambda_0))=\{u: A(\lambda_0)u=0\}$ конечна.

Жордановыми (корневыми) цепочками длины $n+1$ (для A в точке λ_0) называются цепочки элементов $(u^0,u^1,\ldots,u^n)$ с $u^0\neq 0$, удовлетворяющие соотношениям

$$A(\lambda_0)u^0=0,\quad A'(\lambda_0)u^0+A(\lambda_0)u^1=0,\ \ldots,$$
$$\ldots,\ \frac{1}{n!}A^{(n)}(\lambda_0)u^0+\frac{1}{(n-1)!}A^{(n-1)}(\lambda_0)u^1+\ldots+A(\lambda_0)u^n=0. \tag{2.1}$$

Линейную оболочку всех элементов из $\mathcal{U}$, которые могут встретиться в некоторой из жордановых цепочек для A в точке λ_0, будем называть подпространством обобщенных собственных элементов и обозначать через $G(A,\lambda_0)$.

Алгебраическая кратность $\nu(u^0)$ собственного элемента u^0 есть число, равное максимальной длине жордановых цепочек, начинающихся с u^0. В наших предположениях $\max\{\nu(u^0):$ $: A(\lambda_0)u^0=0,\ u^0\neq 0\}=\varkappa<\infty$, где $\varkappa=\varkappa(A,\lambda_0)$ – порядок полюса оператор-функции A^{-1} в точке λ_0 [11]. Алгебраическая кратность $\nu(\lambda_0,A)$ точки λ_0 для A может быть определена [13], как размерность в $\mathcal{U}^{\varkappa}=\mathcal{U}\times\ldots\times\mathcal{U}$ линейной оболочки всех жордановых цепочек для A в точке λ_0, дополненных спереди нулями (т.е. каждая жорданова цепочка $(u^0,\ldots,u^n)$ с $n+1<\varkappa$ рассматривается как элемент $(0,\ldots,0,u^0,\ldots,u^n)\in\mathcal{U}^{\varkappa}$).

Аналогично определяют соответствующие понятия для B_i.

Теорема 1 [6,7,15,10]. Для любого компакта $\Lambda_0\subset\Lambda$ с границей $\partial\Lambda_0\subset\rho(A)$ существует индекс $i(\Lambda_0)$ такой, что при всех $i\ge i(\Lambda_0)$ суммарные алгебраические кратности всех собственных значений в Λ_0 для A и B_i равны: $\sum\nu(\lambda,A)=$ $=\sum\nu(\mu,B_i)$, где суммируется по всем $\lambda\in\sigma(A)\cap\Lambda_0$ и

$\mu \in \sigma(B_i) \cap \Lambda_0$.

Замечание 2. В [I5] эта теорема доказана в случае сепарабельности $\mathcal{U}$ (достаточно выполнения требования (K) из § 4 ниже). В [I0] для доказательства этой теоремы (там — теоремы 2.3) используется та же конструкция, что и при доказательстве оценок скорости сходимости собственных значений. С этим связано добавочное требование на операторы q_i $(i\in \mathcal{J})$ (то же, что в последнем утверждении теоремы 6 ниже). Для доказательства только нашей теоремы I эту конструкцию легко упростить так, что достаточно предполагать лишь линейность операторов q_i на некотором конечномерном подпространстве. А это мы в теореме I без ограничения общности можем предпологать (см., например, замечание в [I0] п. 2.3).

§ 3. Сходимость жордановых цепочек

Из следствий 4 и 3 и свойств сходимости [14_{I}] непосредственно вытекает следующая теорема 2.

Теорема 2 [6,3]. Пусть $\lambda_0 \in \sigma(A)$, $\lambda_i \in \sigma(B_i)$ и $\lambda_i \to \lambda_0$ $(i\in \mathcal{J})$, а $(x_i,\ i\in\mathcal{J})$ – нормированная последовательность собственных элементов для B_i, соответствующих λ_i:

$$\|x_i\| = 1, \quad B_i(\lambda_i)x_i = 0 \quad \forall i\in\mathcal{J}.$$

Тогда $(x_i,\ i\in\mathcal{J})$ компактна, а из сходимости $x_i \to u^0$ $(i\in\mathcal{J}')$ следует, что u^0 – нормированный собственный элемент для A: $A(\lambda_0)u^0 = 1$, $\|u^0\| = 1$.

Замечание 3. В условиях 1^0–9^0 для произвольного $u^0 \in N(A,\lambda_0)$ не обязательно найдется сходящаяся к нему последовательность элементов из линейной оболочки подпространств $N(B_i,\lambda_i)$ с $\lambda_i \in \sigma(B_i)$, $\lambda_i \to \lambda_0$ $(i\in\mathcal{J})$.

Для $x_i \in \mathcal{X}_i$, $M_i \subseteq \mathcal{X}_i$, $u \in \mathcal{U}$, $M \subseteq \mathcal{U}$ введем обозначения

$$d_i(x_i, M) = \inf\{\|x_i - p_i u\| : u\in M\}, \quad d_i(u, M_i) = \inf\{\|p_i u - x_i\| : x_i \in M_i\}.$$

Теорема 3 [6,9]. Пусть λ_0 – единственная точка спектра $\sigma(A)$ в компакте $\Lambda_0 \subset \Lambda$ с границей $\Gamma \subset \rho(A)$. Тогда

$$\max\{d_i(u, G(B_i,\Lambda_0)) : u \in G(A,\lambda_0),\ \|u\| = 1\} \to 0 \quad (i\in\mathcal{J}),$$

где $G(B_i,\Lambda_0)$ – линейная оболочка всех подпространств $G(B_i,\lambda_i)$ с $\lambda_i \in \sigma(B_i)\cap\Lambda_0$.

Доказательство. Воспользуемся фактом [II], что $G(A,\lambda_0)$ совпадает с линейной оболочкой областей значений операторов $S_j = (2\pi i)^{-1}\int_\Gamma (\lambda-\lambda_0)^{j-1}A^{-1}(\lambda)d\lambda$, $j = 1,\ldots,\varkappa(A,\lambda_0)$. Напомним, что каждая оператор-функция B_i^{-1} голоморфна на $\rho(B_i)$ [5], а для $\Gamma\subset\rho(A)$ существует [6,2,10] индекс $i(\Gamma)$ такой, что $\Gamma\subset\subset\rho(B_i)$ при всех $i\geqslant i(\Gamma)$ и

1) $\sup\{\|B_i^{-1}(\lambda)\| : \lambda\in\Gamma, i\geqslant i(\Gamma)\}\leq c<\infty$,

2) $y_i\to v\ (i\in J)\Rightarrow B_i^{-1}(\lambda)y_i\to A^{-1}(\lambda)v\quad (i\in J, i\geqslant i(\Gamma))\ \forall\lambda\in\Gamma$.

Значит, по предложению 2, при $i\geqslant i(\Gamma)$ операторы $S_{ij}= =(2\pi i)^{-1}\int_\Gamma(\lambda-\lambda_0)^{j-1}B_i^{-1}(\lambda)\,d\lambda$ будут сходиться к S_j:

$$y_i\to v\ (i\in J)\ \Rightarrow\ S_{ij}y_i\to S_jv\ (i\in J)\quad \forall j=1,2,\dots\varkappa.$$

Но при всех $i\geqslant i(\Gamma)$ для достаточно малых $\delta_i>0$ имеем

$$S_{ij}=\sum_{\lambda_i\in\sigma(B_i)\cap\Lambda_0}(2\pi i)^{-1}\int_{|\lambda-\lambda_0|=\delta_i}[(\lambda_i-\lambda_0)+(\lambda-\lambda_i)]^{j-1}B_i^{-1}(\lambda)\,d\lambda=$$

$$=\sum_{\lambda_i\in\sigma(B_i)\cap\Lambda_0}\ \sum_{\ell=0}^{j-1}\binom{j-1}{\ell}(\lambda_i-\lambda_0)^{j-1-\ell}(2\pi i)^{-1}\int_{|\lambda-\lambda_i|=\delta_i}(\lambda-\lambda_i)^{\ell}B_i^{-1}(\lambda)\,d\lambda,$$

откуда видно, что значения операторов S_{ij} принадлежат $G(B_i,\Lambda_0)$.

Ввиду конечномерности $G(A,\lambda_0)$ для доказательства теоремы 3 остается заметить, что для каждого u из базиса $G(A,\lambda_0)$ можно подобрать элементы $v_1,\dots,v_\varkappa$ такие, что $u=S_1v_1+\dots+S_\varkappa v_\varkappa$, а тогда для $(x_i,\ i\in J,\ i\geqslant i(\Gamma))$ с $x_i=S_{i1}q_iv_1+\dots+S_{i\varkappa}q_iv_\varkappa$ имеем $x_i\in G(B_i,\Lambda_0)$, $x_i\to u\ (i\in J)$.

Замечание 4. В условиях теоремы 3 не обязательно $\max\{d_i(x_i,G(A,\lambda_0)) : x_i\in G(B_i,\Lambda_0),\ \|x_i\|=1\}\to 0\ (i\in J)$, но в случае линейной зависимости A и B_i от параметра такая сходимость имеет место [1].

Теорема 4 [6,9]. Пусть $\lambda_0\in\sigma(A)$, $\lambda_i\in\sigma(B_i)$ и $\lambda_i\to\lambda_0$ $(i\in J)$, а $((x_i^0,x_i^1,\dots,x_i^{r-1}),\ i\in J)$ – последовательность жордановых цепочек длины r для B_i в точках λ_i, причём

$$\max\{\|x_i^0\|,\|x_i^1\|,\dots,\|x_i^{r-1}\|\}=1\quad\forall i\in J.$$

Тогда все $(x_i^j,\ i\in J)$, $j=0,1,\dots,r-1$ компактны, а из сходимости $x_i^0\to u^0,\dots,x_i^{r-1}\to u^{r-1}\ (i\in J')$ следует, что $\max\{\|u^0\|,\dots,\|u^{r-1}\|\}=1$ и $(u^0,u^1,\dots,u^{r-1})$ – жорданова цепочка для A в точке λ_0, дополненная спереди нулями: если $u^0=\dots=u^{\ell-1}=0\neq u^\ell$, то $(u^\ell,\dots,u^{r-1})$ жорданова цепочка.

Доказательство. Так как $(x_i^0,\ i\in J)$ ограничена по норме и $(B_i(\lambda_i)x_i^0,\ i\in J)$ компактна ($B_i(\lambda_i)x_i^0=0,\ i\in J$), то, по следствию 4, последовательность $(x_i^0, i \in J)$ должна быть компактной. Но тогда, по следствию 3, будут компактными также последовательности $(B_i'(\lambda_i)x_i^0, i\in J)$ и $(B_i(\lambda_i)x_i^1, i\in J)$ (так как $B_i(\lambda_i)x_i^1 + B_i'(\lambda_i)x_i^0 = 0$, $i \in J$). Из компактности $(B_i(\lambda_i)x_i^1, i\in J)$ ввиду ограниченности $(x_i^1, i\in J)$ следует, в свою очередь, компактность $(x_i^1, i \in J)$. Продолжая рассуждать таким же образом, мы получим компактность всех последовательностей $(x_i^j, i\in J)$, $j = 1,\dots,r-1$.

Остальные утверждения теоремы 4 вытекают непосредственно из следствия 3, определения жордановых цепочек и свойств сходимости.

Замечание 5. В [13] в пространстве $\mathcal{U}^{\varkappa} = \mathcal{U}\times\dots\times\mathcal{U}$ построен проектор E на подпространство жордановых цепочек для A в точке λ_0, дополненных спереди нулями. Указаны также операторы $E_i \in \mathcal{B}(\mathcal{X}_i^{\varkappa}, \mathcal{X}_i^{\varkappa})$, сходящиеся дискретно к E и показано, как E (E_i) может быть построен, если известны элементы жордановых цепочек оператора и его сопряженного.

§ 4. Скорость сходимости жордановых цепочек

В [3] доказана следующая теорема 5.

Теорема 5. Пусть $\lambda_0 \in \sigma(A)$, $\lambda_i \in \sigma(B_i)$ и $\lambda_i \to \lambda_0$ $(i \in J)$. Пусть операторы p_i $(i\in J)$ линейны на $N(A,\lambda_0)$ и $q_i 0 = 0$ $(i \in J)$. Тогда существует индекс i_0 такой, что при всех $i \geq i_0$ имеет место оценка:

$$\max\{d_i(x_i, N(A,\lambda_0)) : x_i \in N(B_i,\lambda_i),\ \|x_i\| = 1\} \leq$$
$$\leq c_1|\lambda_i - \lambda_0| + c_2 \max\{\|[B_i(\lambda_0)p_i - q_i A(\lambda_0)]u\| : u \in N(A,\lambda_0), \|u\|=1\}. \tag{4.1}$$

Замечание 6. В [6] получена более грубая оценка:

$$\max\{d_i(x_i, N(A,\lambda_0)) : x_i\in N(B_i,\lambda_i), \|x_i\|=1\} \leq c(|\lambda_i - \lambda_0| + \bar{\varepsilon}_i),$$
$$\bar{\varepsilon}_i = \sup\{\|[B_i(\lambda)p_i - q_i A(\lambda)]u\| : |\lambda-\lambda_0| = \delta,\ u\in G(A,\lambda_0),\ \|u\|=1\}, \tag{4.2}$$

с любым $\delta > 0$ таким, что $\{\lambda : |\lambda - \lambda_0| \leq \delta\} \subset \Lambda$.

Пользуясь идеей "линеаризации" проблем $A(\lambda)u = 0$ и $B_i(\lambda)x_i = 0$ [13], мы на основе теоремы 5 получим, далее, оценку скорости сходимости для жордановых цепочек.

Введем банаховы пространства $\mathfrak{U} = \mathcal{U}^r = \mathcal{U}\times\dots\times\mathcal{U}$ (r раз) и $\mathfrak{V} = \mathcal{V}^r = \mathcal{V}\times\dots\times\mathcal{V}$ с нормами $\|u\| = \|(u_1,\dots,u_r)\| = \max\{\|u_1\|,\dots,\|u_r\|\}$ и $\|v\| = \|(v_1,\dots,v_r)\| = \max\{\|v_1\|,\dots,\|v_r\|\}$.

Пусть (1) $\lambda_0 \in \sigma(A)$, (2) $\dim N(A,\lambda_0) = m$,

(3) $(u_1^0, \ldots, u_m^0)$ – некоторый базис в $N(A, \lambda_0)$,
(4) $(v_1, \ldots, v_m)$ – базис в некотором прямом дополнении V' подпространства $A(\lambda_0)\mathcal{U}$ в V, (5) $(g_1, \ldots, g_m)$ – система функционалов из $\mathcal{B}(\mathcal{U}, \mathbb{C})$, биортогональная к системе[2] $(u_1^0, \ldots, u_m^0)$: $\langle u_k^0, g_j\rangle = \delta_{kj}$, $k, j = 1, \ldots, m$.

Рассмотрим проблему собственных значений

$$\mathbb{A}(\mu)\mathbf{u} \equiv \mathbb{A}_0\mathbf{u} + \mu\mathbb{K}\mathbf{u} = 0 \quad (\mu \in \mathbb{C}) \qquad (4.3)$$

с $\mathbb{A}_0, \mathbb{K} \in \mathcal{B}(\mathbb{U}, \mathbb{V})$, определенными соотношениями

$$\mathbb{A}_0\mathbf{u} = \begin{pmatrix} A_0 & 0 & \ldots & 0 \\ A_1 & A_0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ A_{r-1} & A_{r-2} & \ldots & A_0 \end{pmatrix}\mathbf{u}, \quad \mathbb{K}\mathbf{u} = \begin{pmatrix} K & 0 & \ldots & 0 \\ 0 & K & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & K \end{pmatrix}\mathbf{u}, \qquad (4.4)$$

где $A_j = \frac{1}{j!}A^{(j)}(\lambda_0)$ и $Ku = \sum_{k=1}^{m} \langle u, g_k\rangle v_k$.

Нетрудно убедиться, что все операторы $\mathbb{A}(\mu)$ фредгольмовы с индексом 0, $\sigma(\mathbb{A}) = \{0\}$ и $\max\{\nu(\mathbf{u}^0) : \mathbf{u}^0 \in N(\mathbb{A}, 0)\} = 1$. При этом $\mathbf{u}^0 = (0, \ldots, 0, u^0, \ldots, u^{n-1})$ с $u^0 \neq 0$ будет собственным элементом для $\mathbb{A}$ в точке $\mu = 0$ тогда и только тогда, когда $(u^0, \ldots, u^{n-1})$ – жорданова цепочка для A в точке λ_0.

Введем, далее, банаховы пространства $\mathfrak{X}_i = \mathcal{X}_i^r$ и $\mathbb{Y}_i = \mathcal{Y}_i^r$ аналогично $\mathbb{U}$ и $\mathbb{V}$. Определим "связывающие" отображения $\mathbb{p}_i : \mathbb{U} \to \mathfrak{X}_i$ и $\mathbb{q}_i : \mathbb{V} \to \mathbb{Y}_i$ соотношениями $\mathbb{p}_i(u_1, \ldots, u_r) = (p_i u_1, \ldots, p_i u_r)$, $\mathbb{q}_i(v_1, \ldots, v_r) = (q_i v_1, \ldots, q_i v_r)$. Отметим, что для $\mathbb{p}_i$ и $\mathbb{q}_i$, $i \in J$ выполнены требования 1^0–3^0.

Введем следующее дополнительное требование (K), которое выполнено, например, в случае сепарабельности $\mathcal{U}$ [12$_1$].

<u>Требование (K)</u>. Для каждого g_k, $k = 1, \ldots, m$ в определении оператора K существует последовательность функционалов $(f_i^k, i \in J)$ с $f_i^k \in \mathcal{B}(\mathcal{X}_i, \mathbb{C})$, слабо сходящаяся к g_k: $x_i \to u \ (i \in J) \Rightarrow \langle x_i, f_i^k\rangle \to \langle u, g^k\rangle \ (i \in J)$.

Пусть, далее, (6) $\lambda_i \in \sigma(B_i)$, (7) $\lambda_i \to \lambda_0 \ (i \in J)$, (8) $(f_i^k, i \in J)$, $k = 1, \ldots, m$ – последовательности, оговоренные требованием (K). Рассмотрим при каждом $i \in J$ аналогичную проблеме (4.3) проблему собственных значений

$$\mathbb{B}_i(\mu)\mathbf{x}_i \equiv \mathbb{B}_i^0\mathbf{x}_i + \mu\mathbb{L}_i\mathbf{x}_i = 0 \qquad (\mu \in \mathbb{C}) \qquad (4.5)$$

[2] Через $\langle u, g\rangle$ обозначим значение функционала g на u.

с $\mathbb{B}_i^0, \mathbb{L}_i \in \mathcal{B}(\mathcal{X}_i, \mathcal{Y}_i)$, определенными соотношениями

$$\mathbb{B}_i^0 \mathbb{x}_i = \begin{pmatrix} B_i^0 & 0 & \dots & 0 \\ B_i^1 & B_i^0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ B_i^{r-1} & B_i^{r-2} & \dots & B_i^0 \end{pmatrix} \mathbb{x}_i, \quad \mathbb{L}_i \mathbb{x}_i = \begin{pmatrix} L_i & 0 & \dots & 0 \\ 0 & L_i & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & L_i \end{pmatrix} \mathbb{x}_i, \tag{4.6}$$

где $B_i^j = \frac{1}{j!} B_i^{(j)}(\lambda_i)$ и $L_i x_i = \sum_{k=1}^{m} \langle x_i, f_i^k \rangle q_i v_k$.

Нетрудно убедиться, что все операторы $\mathbb{B}_i(\mu)$ фредгольмовы с индексом 0 и $0 \in \sigma(\mathbb{B}_i)$ $\forall i \in \mathcal{J}$. При этом $(0, \dots, 0, x_i^0, \dots, x_i^{n-1})$ с $x_i^0 \neq 0$ – собственный элемент для $\mathbb{B}_i$ в точке 0 тогда и только тогда, когда $(x_i^0, \dots, x_i^{n-1})$ – жорданова цепочка для B_i в точке λ_i.

Для применения теоремы 5 к $\mathbb{A}$ и $\mathbb{B}_i$ остается еще проверить, что выполнены требования 7^0–9^0. Это делается несложно, если учесть при проверке 7^0, что сходящиеся слабо последовательности ограничены по норме [14_I], а при проверке 9^0, что для каждой ограниченной по норме последовательности $(x_i, i \in \mathcal{J})$ последовательность $(L_i x_i, i \in \mathcal{J})$ компактна.

Для $(x_i^1, \dots, x_i^r)$ и $\mathbb{M} = M^1 \times \dots \times M^r \subseteq \mathcal{U}^r$ введем обозначение

$$d_i((x_i^1, \dots, x_i^r), \mathbb{M}) = \max_{j=1,\dots,r} \inf \{ \| x_i^j - p_i u \| : u \in M^j \}.$$

<u>Теорема 6</u>. Пусть $\lambda_0 \in \sigma(A)$, $\lambda_i \in \sigma(B_i)$ и $\lambda_i \to \lambda_0$ $(i \in \mathcal{J})$. Пусть операторы p_i линейны на $G(A, \lambda_0)$, $q_i 0 = 0$ $(i \in \mathcal{J})$ и выполнено требование (K). Тогда существует индекс i_0 такой, что при всех $i \geq i_0$ имеет место оценка

$\max\{ d_i((x_i^0, \dots, x_i^{r-1}), N(\mathbb{A}_0)) : (x_i^0, \dots, x_i^{r-1})$ — жорданова цепочка для B_i в точке λ_i, $\max(\|x_i^0\|, \dots, \|x_i^{r-1}\|) = 1 \} \leq$

$\leq C \varepsilon_i = C \max\{ \| [\sum_{j=0}^{r-1} \frac{1}{j!} B_i^{(j)}(\lambda_i) p_i - q_i \sum_{j=0}^{r-1} \frac{1}{j!} A^{(j)}(\lambda_0)] u^{r-j-1} \| :$
$(u^0, \dots, u^{r-1}) \in N(\mathbb{A}_0)$, $\max(\|u^0\|, \dots, \|u^{r-1}\|) = 1 \}$, где $N(\mathbb{A}_0)$ – линейная оболочка в $\mathcal{U}^r = \mathcal{U} \times \dots \times \mathcal{U}$ всех жордановых цепочек длины $\leq r$ для A в точке λ_0, дополненных спереди нулями (т.е. каждая жорданова цепочка $(u^0, \dots, u^{n-1})$ с $n < r$ рассматривается как элемент $(0, \dots, 0, u^0, \dots, u^{n-1}) \in \mathcal{U}^r$).

При этом $\varepsilon_i \leq C |\lambda_i - \lambda_0| + \varepsilon_i^0$, где

$$\varepsilon_i^0 = \max\{ \| \sum_{j=0}^{r-1} \frac{1}{j!} B_i^{(j)}(\lambda_0) p_i u^{r-j-1} - q_i \sum_{j=0}^{r-1} \frac{1}{j!} A^{(j)}(\lambda_0) u^{r-j-1} \| : (u^0, \dots, u^{r-1}) \in N(\mathbb{A}_0), \max(\|u^0\|, \dots, \|u^{r-1}\|) = 1 \}. \tag{4.7}$$

Если, кроме того, операторы q_i $(i\in\mathcal{J})$ линейны и непрерывны на некотором подпространстве $\mathcal{V}_0\subseteq\mathcal{V}$ таком, что $\mathcal{V}_0\supseteq A(\lambda)G(A,\lambda_0)$ для всех λ на некоторой окружности $|\lambda-\lambda_0|=\delta>0$ с $\{\lambda:|\lambda-\lambda_0|\leq\delta\}\subset\Lambda$, то $\varepsilon_i^0\leq c\bar{\varepsilon}_i$ с $\bar{\varepsilon}_i$, определенными в (4.2).

<u>Доказательство</u>. Для доказательства оценки (4.7) заметим, что если $(u^k,\dots,u^r)$ — жорданова цепочка (для A в точке λ_0), то жордановой цепочкой будет также $(u^k,\dots,u^{r-1})$. Это позволяет в правой стороне оценки (4.7) ограничиться только суммами от 0 до $r-1$.

Оценка $\varepsilon_i\leq c|\lambda_i-\lambda_0|+\varepsilon_i^0$ следует из голоморфности и равномерной ограниченности ($B_i^{(j)}$, $i\in\mathcal{J}$) по норме на любом компакте $\Lambda_0\subset\Lambda$. Для доказательства неравенства $\varepsilon_i^0\leq c\varepsilon_i$ надо в выражении (4.7) производные заменить их представлением через интегральную формулу Коши.

Литература

1. В а й н и к к о Г., Анализ дискретизационных методов. Изд. Тартуск. ун-та, Тарту, 1976.
2. В а й н и к к о Г.М., К а р м а О.О., О сходимости приближенных методов решения линейных и нелинейных операторных уравнений. Ж. вычисл. матем. и матем. физ., 1974, <u>14</u>, № 4, 828-837.
3. В а й н и к к о Г.М., К а р м а О.О., О быстроте сходимости приближенных методов в проблеме собственных значений с нелинейным вхождением параметра. Ж. вычисл. матем. и матем. физ., 1974, <u>14</u>, № 6, 1393-1408.
4. Г о х б е р г И.Ц., О линейных операторах, аналитически зависящих от параметра. Докл. АН СССР, 1951, <u>78</u>, № 4, 629-632.
5. Д а н ф о р д Н., Ш в а р ц Дж., Линейные операторы. Общая теория. ИИЛ, Москва, 1962.
6. К а р м а О.О., Об аппроксимации оператор-функций и сходимости приближенных собственных значений. Канд. диссертация, Тарту, 1971.
7. К а р м а О.О., Асимптотические оценки скорости сходимости собственных значений при регулярной аппроксимации. В сб. Республиканский симпозиум по методам решения нелинейных уравнений и задач оптимизации. Пярну, 5-10 июня 1978 г. (Тезисы докладов). Таллин, 1978.

8. К а р м а О., Об аппроксимации в задачах на собственные значения для оператор-функций, голоморфных типа (A). Уч. зап. Тартуск. ун-та, 1981, 580, 9-23.

9. К а р м а О.О., О сходимости собственных элементов при регулярной аппроксимации оператор-функций. В. сб. "II Республиканский симпозиум по методам решения нелинейных уравнений и задач оптимизации. Хаапсалу, 4-7 июня 1981 г. Доклады и сообщения." Таллин, 1981.

10. К а р м а О., Аппроксимация в проблеме собственных значений с голоморфной зависимостью оператора от параметра (I). 1-47. Рукопись депонирована в ВИНИТИ 11. 01. 82 г. № 130-82 Деп.

11. Т р о ф и м о в В.П., О корневых подпространствах операторов, аналитически зависящих от параметра. Матем. исследования, 1968, 3, вып. 3, 117-125.

12. G r i g o r i e f f, R.D., Zur Theorie linearer approximationsregulären Operatoren I. Math. Nachr., 1973, 55, 233-249. II Math. Nachr., 1973, 55, 251-263.

13. J e g g l e, H., W e n d l a n d, W., On the Discrete Approximation of Eigenvalue Problems with Holomorphic Parameter Dependence. Proc. Roy. Soc. Edinburgh, 1978, 78A, 1-29.

14. S t u m m e l, F., Discrete Konvergenz linearer Operatoren I. Math. Annal. 1970, 190, 45-92; II. Math. Z. 1971, 120, 231-264; III. Int. Series of Numerical Mathematics, 20, Birkhäuser Verlag, Basel, 1972.

15. W o l f, R., Zur Stabilität der algebraischen Vielfachheit von Eigenwerten von holomorphen Fredholm-Operatorfunctionen. Appl. Analysis, 1979, 9, 165-177.

Поступило
8 III 1982

APPROXIMATION IN EIGENVALUE PROBLEMS
WITH HOLOMORPHIC DEPENDENCE ON PARAMETER (II)

O.Karma

Summary

Approximation of the eigenvalue problem $A(\lambda)u=0$ for the holomorphic Fredholm operator-function A is studied. On the assumptions of a regular convergence (assumptions 1°-9° in § 1) the convergence rate of Jordan chains is given.

Уч.зап.Тартуск.ун-та,1982,633,29-34

ОЦЕНКИ СКОРОСТИ СХОДИМОСТИ АППРОКСИМАЦИОННО-ИТЕРАЦИОННОГО МЕТОДА

И.Саарнийт

В статье уточняются результаты о скорости сходимости аппроксимационно-итерационного метода, полученные в [4,5]. Результаты иллюстрируются на примере решения краевой задачи для обыкновенного дифференциального уравнения.

1. <u>О скорости сходимости аппроксимационно-итерационного метода.</u> Пусть даны метрические пространства (E, ρ) и (E_n, ρ_n), $n = 0, 1, \ldots,$ а также система $\mathcal{P} = \{p_n\}$, $n = 0, 1, \ldots,$ связывающих отображений $p_n : E \to E_n$. Будем говорить, что последовательность $\{x_n\}$, $n = 0, 1, \ldots,$ с $x_n \in E_n$ $\mathcal{P}$-сходится к $x \in E$, если $\lim \rho_n(x_n, p_n x) = 0$, обозначаем это $\mathcal{P}\text{-}\lim x_n = x$ (подробнее см. [1]).

Пусть $T_n : E_n \to E_n$, $n = 0, 1, \ldots,$ – некоторые отображения метрических пространств E_n в себя, а $\vartheta_n : E_{n-1} \to E_n$, $n = 1, 2, \ldots,$ – отображения перехода из пространств E_{n-1} в пространства E_n. Пусть в пространстве E_0 задано начальное приближение x_0^0. Итерационный метод

$$z_n = (T_n)^{\alpha_n} \vartheta_n z_{n-1}, \quad n = 0, 1, \ldots, \tag{1}$$

где $\alpha_n \geq 0$ – целые числа, а $\vartheta_0 z_{-1} = x_0^0$, будем называть аппроксимационно-итерационным методом (АИМ).

Опишем условия, налагаемые на T_n и ϑ_n. Предположим, что

$$\exists x_n^* \in E_n : x_n^* = T_n x_n^*, \quad n = 0, 1, \ldots, \quad \mathcal{P}\text{-}\lim x_n^* = x^* \in E, \tag{2}$$

причем

$$\rho_0(x_0^0, x_0^*) \leq d_0 < \tau, \quad \rho_n(\vartheta_n x_{n-1}^*, x_n) \leq d_n < \tau, \; n \geq 1. \tag{3}$$

Пусть

$$\rho_n(T_n x_n, T_n x_n^*) \leq \omega(\rho_n(x_n, x_n^*)), \text{ для } \forall x_n : \rho_n(x_n, x_n^*) < \tau, \tag{4}$$

где $\omega(s)$ – определенная на интервале $[0, \tau)$, $0 < \tau \leq \infty$, неубывающая, непрерывная справа функция.

Об отображениях ϑ_n предположим, что

$$\rho_n(\vartheta_n x_{n-1}, \vartheta_n y_{n-1}) \leq K \rho_{n-1}(x_{n-1}, y_{n-1}), \tag{5}$$

$x_{n-1}, y_{n-1} \in E_{n-1}$, $K = \text{const}$. О сходимости АИМ (1) в [4] доказана следующая

Теорема 1. Пусть выполняются условия (2)-(5), а также

$$\lim d_n = 0 \tag{6}$$

и для β-й итерации $\omega^\beta(s)$ функции $\omega(s)$ имеет место неравенство

$$\omega^\beta(s) < s/K \quad \text{при} \quad 0 < s < \tau_0 = \tau_0(\beta) \le \tau. \tag{7}$$

Тогда, начиная от достаточно больших n, АИМ (1) будет сходится при $\alpha_n \ge \beta$.

Пусть нам известна оценка

$$\rho_n(x_n^*, p_n x^*) \le \varepsilon_n, \quad \lim \varepsilon_n = 0, \ \varepsilon_{n+1} \ge \eta \varepsilon_n, \ 0 < \eta < 1. \tag{8}$$

Изучим, когда погрешность АИМ имеет такой же порядок.

Теорема 2. Пусть выполняются условия (2)-(5), (8) а также условие

$$d_{n+1} \le \zeta d_n, \ n \ge 1, \ 0 < \zeta < 1. \tag{9}$$

Пусть существует число $q < \zeta/K$ такое, что

$$\omega^\beta(s) \le qs \quad \text{при} \quad 0 \le s \le \tau_1 = \tau_1(q, \beta) \le \tau, \tag{10}$$

и пусть выполняется неравенство

$$\omega^\beta(\zeta^t) \le \lambda \eta^t \quad \text{при} \quad \zeta^t < \tau, \tag{11}$$

где $\lambda = const < \infty$. Тогда, начиная от достаточно больших n, при $\alpha_n \ge \beta$ будет иметь место оценка

$$\rho_n(z_n, p_n x^*) \le const \cdot \varepsilon_n. \tag{12}$$

Доказательство. Определим рекуррентную последовательность $\{\delta_n\}$:

$$\delta_0 = \omega^{\alpha_0}(d_0), \ \delta_n = \omega^{\alpha_n}(K\delta_{n-1} + d_n), \ n \ge 1.$$

При этом α_n выбираем так, чтобы $K\delta_n + d_{n+1} < \tau$, $n \ge 0$. Так как

$$\rho_n(z_n, p_n x^*) \le \rho_n(z_n, x_n^*) + \rho_n(x_n^*, p_n x^*), \tag{13}$$

то ввиду

$$\rho_0(z_0, x_0^*) \le \rho_0(T_0^{\alpha_0} x_0^0, T^{\alpha_0} x_0^*) \le \omega^{\alpha_0}(\rho_0(x_0^0, x_0^*)) \le \delta_0,$$

$$\rho_n(z_n, x_n^*) \le \rho_n(T_n^{\alpha_n} \vartheta_n z_{n-1}, T_n^{\alpha_n} x_n^*) \le \omega^{\alpha_n}(\rho_n(\vartheta_n z_{n-1}, x_n^*)) \le$$

$$\le \omega^{\alpha_n}(K\rho_{n-1}(z_{n-1}, x_{n-1}^*) + \rho_n(\vartheta_n x_{n-1}^*, x_n^*)) \le \delta_n, \ n \ge 1,$$

достаточно доказать, существование индекса m такого, что

$$\delta_n \le const \cdot \eta^n \quad \text{при} \quad \alpha_n \ge \beta \quad \text{для всех} \quad n > m. \tag{14}$$

Выберем m такой, что

$$d_{m+1} \leq \frac{\tau_1}{\zeta}(\zeta - Kq). \tag{15}$$

Заметим, что в этом случае

$$q \leq \frac{\zeta}{\tau_1 K}(\tau_1 - d_{m+1}). \tag{16}$$

Выберем α_m при котором

$$K\delta_m + d_{m+1} \leq \tau_1. \tag{17}$$

Пусть теперь $\alpha_n \geq \beta$, $n \geq m+1$. Докажем, что

$$\delta_n \leq \omega^\beta(\tau_1 \zeta^{n-m-1}), \quad n \geq m+1. \tag{18}$$

Действительно, согласно неравенству (17)

$$\delta_{m+1} \leq \omega^\beta(K\delta_m + d_{m+1}) \leq \omega^\beta(\tau_1).$$

Пусть теперь $\delta_{n-1} \leq \omega^\beta(\tau_1 \zeta^{n-m-2})$. Тогда, ввиду (9) и (10),

$$K\delta_{n-1} + d_n \leq \zeta^{n-m-1}\left(\frac{Kq\tau_1}{\zeta} + d_{m+1}\right).$$

Используя (16), отсюда получаем, что

$$K\delta_{n-1} + d_n \leq \tau_1 \zeta^{n-m-1},$$

откуда, в свою очередь, следует (18).

Для завершения доказательства теоремы достаточно использовать в (18) неравенство (11), согласно которому

$$\delta_n \leq \lambda \cdot \eta^{\ln\tau_1/\ln\zeta - m-1} \cdot \eta^n.$$

<u>Замечания</u>. 1. Из неравенства

$$\rho_n(x_n^*, \vartheta_n x_{n-1}^*) \leq \rho_n(x_n^*, p_n x^*) + \rho_n(p_n x^*, \vartheta_n p_{n-1} x^*) + K\rho_{n-1}(x_{n-1}^*, p_{n-1}x^*)$$

видно, что практический интерес представляет лишь случай $\zeta \geq \eta$. Строгое неравенство $\zeta > \eta$ имеет место, если $\{\rho_n(p_n x^*, \vartheta_n p_{n-1} x^*)\}$ стремится к нулю медленнее,чем $\{\rho_n(x_n^*, p_n x^*)\}$.

2. Если $\zeta = \eta$, то (11) выполняется ввиду предположения $\omega(s) < s$. Если же $\zeta > \eta$, то из (11) следует, что для любого $q > 0$ существует $\tau_1 = \tau_1(q) \in (0, \tau]$ такой, что выполняется (10).

3. Из доказательства теоремы видно, что выполнение условия (9) требуется лишь начиная от достаточно больших n.

2. <u>Построение АИМ в случае сжимающих отображений</u>. Пусть

$$\omega(s) = \varkappa s, \quad 0 \leqslant \varkappa < 1, \quad \tau = +\infty .$$

Тогда $\omega^{\beta}(s) = \varkappa^{\beta} s$ и число итераций β надо выбирать так, чтобы $q = \varkappa^{\beta} < \zeta / K$. При этом τ_1 – любое положительное число. Условие (II) выполнимо лишь при $\zeta = \eta$, тогда $\lambda = \varkappa^{\beta}$. Поэтому операторы перехода ϑ_n должны удовлетворять условию

$$\rho_n(p_n x^*, \vartheta_n p_{n-1} x^*) \leqslant \mathrm{const} \cdot \eta^n .$$

Если

$$\tau_1 = \frac{\eta d_1}{\eta - K\varkappa^{\beta}},$$

то (I5) выполняется при $m = 0$. Пусть $\delta_0 \leqslant \dfrac{d_1 \varkappa^{\beta}}{\eta - K\varkappa^{\beta}}$, т.е. начальное приближение достаточно хорошо. Тогда при $m = 0$ выполняется (I7). Выбрав число итерации $\alpha_n = \beta$, $n \geqslant 1$, мы получаем из (I8) следующую оценку:

$$\delta_n \leqslant \frac{\varkappa^{\beta} d_1}{\eta - K\varkappa^{\beta}} \eta^n, \quad n \geqslant 1 .$$

3. <u>Построение АИМ на базе итерационного метода с квадратичной скоростью сходимости</u>. Пусть

$$\omega(s) = M s^2; \quad \tau = 1/M ,$$

где $M = \mathrm{const}$. Исследуем скорость сходимости АИМ при $\beta = 1$. Условие (I0) выполняется при любом $q < \zeta / K$ с $\tau_1 = q / M$, а (II) имеет место, если $\zeta^2 \leqslant \eta$. Неравенство (I5) принимает вид

$$d_{m+1} \leqslant \frac{q}{M\zeta}(\zeta - Kq),$$

такое q найдется, если $d_{m+1} \leqslant \zeta / 4KM$. Для выполнения условия (I7) надо подобрать α_m так, чтобы

$$K\delta_m \leqslant \frac{q}{M} - d_{m+1} .$$

Рассмотрим следующие два случая.

А: $\eta = \zeta^2$, т.е. операторы ϑ_n удовлетворяют условию

$$\rho_n(p_n x^*, \vartheta_n p_{n-1} x^*) \leqslant \mathrm{const}\ \eta^{n/2} .$$

Так как $\tau_1 = q / M$, то из оценки (I8) следует, что

$$\delta_n \leqslant \frac{q^2}{M} \eta^{-m-1} \cdot \eta^n .$$

Б: $\eta = \zeta$, т.е. операторы ϑ_n удовлетворяют условию

$$\rho_n(p_n x^*, \vartheta_n p_{n-1} x^*) \leqslant \mathrm{const}\ \eta^n .$$

Из неравенства (I8) получаем, что

$$\delta_n \leq \frac{q^2}{M}\eta^{-2(m+1)}\cdot\eta^{2n},$$

однако, ввиду неравенства (I3), оценка (I2) не улучшается.

4. <u>Приложение: приближенное решение краевой задачи для обыкновенного дифференциального уравнения.</u> Рассмотрим задачу

$$x'' = f(t,x,x'),\ 0<t<1;\ x(0)=x(1)=0, \tag{I9}$$

где f – заданная на $(0,1)\times\mathbb{R}^1\times\mathbb{R}^1$ функция. Пусть задача (I9) имеет решение $x^*(t)$. В качестве приближенной задачи положим систему конечноразностных уравнений

$$\left.\begin{aligned} &k^2(x^{i+1}-2x^i+x^{i-1}) = f\left(t^i, x^i, \tfrac{k}{2}(x^{i+1}-x^{i-1})\right),\ i=1,\dots,k-1;\\ &x^0 = x^k = 0, \end{aligned}\right\} \tag{20}$$

где $t^i = i/k$. Пусть задача (20) имеет при $k\geq k_0$ решение $(x^{*i})_{0\leq i\leq k}$, причем

$$\max_{1\leq i\leq k-1}\max\left(|x^*(t^i)-x^{*i}|,\ |x^{*\prime}(t^i)-\tfrac{k}{2}(x^{*,i+1}-x^{*,i-1})|\right)\leq \mathrm{const}\cdot\frac{1}{k^2}. \tag{2I}$$

Достаточные для этого условия даны, например, в [2].

Построим АИМ следующим образом. Рассмотрим последовательность задач (20) с $k=k_n=2^n\cdot k_0$. Пусть E_n – пространства векторов

$$E_n = \left\{x_n = (x_n^i)_{0\leq i\leq k_n} : x_n^0 = x_n^{k_n} = 0\right\},$$

снабженные нормой

$$\|x_n\| = \max_{1\leq i\leq k_n-1}\max\left(|x_n^i|,\ \tfrac{k_n}{2}|x_n^{i+1}-x_n^{i-1}|\right).$$

В качестве E положим пространство непрерывно дифференцируемых функций $C^1[0,1]$, а

$$p_n x = \left(x(t_n^i)\right)_{0\leq i\leq k_n}\in E_n,\ t_n^i = i/k_n.$$

Из неравенства (2I) следует оценка (8) с $\eta = 1/4$.

Для нахождения решения x_n^* системы (20) используем метод Ньютона-Канторовича. Непосредственно проверяется, что при достаточно гладкой функции f данный метод будет иметь квадратичную скорость сходимости. Поэтому достаточно построить отображения перехода ϑ_n, удовлетворяющие неравенству (9) с $\zeta = 1/2$. Можно построить ϑ_n, например, следующим

9

образом: $\vartheta_n x_{n-1} = y_n \in E_n$, где

$$y_n^i = \begin{cases} x_{n-1}^j, & \text{если } t_n^i = t_{n-1}^j, \\ \frac{1}{2}(x_{n-1}^{j+1} + x_{n-1}^j), & \text{если } t_n^i = \frac{1}{2}(t_{n-1}^{j+1} + t_{n-1}^j). \end{cases}$$

Литература

1. Вайникко Г., Анализ дискретизационных методов. Тарту, 1976.
2. Вайникко Г., Компактная аппроксимация операторов и приближенное решение уравнений. Тарту, 1970.
3. Красносельский М.А. и др., Приближенное решение операторных уравнений. Москва, 1969.
4. Саарнийт И., Об одной возможности построения итерационных методов. Уч. зап. Тартуск. ун-та, 1979, 500, 43-52.
5. Саарнийт И., О скорости сходимости аппроксимационно-итерационного метода. Тезисы конференции "Теоретические и прикладные вопросы математики". Тарту, 1980, 205-206.

Поступило
10 V 1982

SCHÄTZUNGEN DER KONVERGENZGESCHWINDIGKEIT DES APPROXIMATIONS-ITERATIONSVERFAHRENS

I. Saarniit

Zusammenfassung

Im Artikel werden die früheren Ergebnisse über die Konvergenzgeschwindigkeit des Approximations-Iterationsverfahrens präzisiert. Die Ergebnisse werden auf dem Beispiel der Lösung von der Randwertaufgabe für die gewöhnliche Differentialgleichung illustriert.

Уч.зап.Тартуск.ун-та,1982,633,35–40

РЕШЕНИЕ ОБРАТНОЙ ЗАДАЧИ ТЕПЛОПРОВОДНОСТИ МЕТОДОМ ПОСЛЕДОВАТЕЛЬНЫХ ПРИБЛИЖЕНИЙ

Э. Тамме

Обратная задача теплопроводности является некорректно поставленной задачей математической физики. Для численного решения этой задачи выработан ряд методов (см., например, [2 , 6]). В данной работе предлагается алгоритм последовательного приближения для решения этой задачи.

I. Обратную задачу теплопроводности можно сформулировать следующим образом (см. [2]): найти на отрезке $0 \le t \le T$ решение $u = u(t)$ задачи

$$u' + Au = f(t), \quad u(T) = h$$

или

$$Bu = g, \tag{I}$$

где $Bu = (u' + Au,\ u(T))$, $g = (f, h)$, а A – линейный непрерывный оператор, действующий из действительного гильбертова пространства V в сопряженное пространство V'. Пусть V всюду плотно и непрерывно вложено в гильбертово пространство H и $(u, v)_V$, $(u, v) = (u, v)_H$, $(u, v)_{V'}$, $\|u\|_V$, $\|u\|_H$, $\|u\|_{V'}$ – скалярные произведения и нормы в пространствах V, H и V'. Отождествляя H с его сопряженным пространством, получаем $V \subset H \subset V'$. Если $u \in V'$ и $v \in V$, то (u, v) обозначает значение функционала u на v и совпадает со скалярным произведением в H для $u \in H$. Обозначим через J канонический изоморфизм из V в V' (см., например, [4]), определенный соотношением

$$(u, v)_V = (Ju, v) \qquad \forall v \in V.$$

Тогда $(u, v)_{V'} = (u, J^{-1}v)$.

Пусть оператор A удовлетворяет условию коэрцитивности

$$(Au, u) \ge m_1 \|u\|_V^2 - m_2 \|u\|_H^2 \qquad \forall u \in V, \tag{2}$$

где $m_1 > 0$. Тогда B является изморфизмом между пространствами $\mathcal{H}_1$ и $\mathcal{H}_2$ (см. [2]), метрики которых определены скалярными произведениями

$$(u, v)_{\mathcal{H}_1} = \int_0^T \left[(u, v)_V + (u', v')_{V'} \right] dt,$$

$$(g,\bar g)_{\mathcal{H}_2} = \int_0^T (f,\bar f)_{V'}\,dt + (h,\bar h)_H,$$

где $\bar g = (\bar f, \bar h)$.

Для приближенного решения задачи (I) применяем метод последовательных приближений

$$u_{n+1} = (I - \alpha^2 B^* B)\, u_n + \alpha^2 B^* g \qquad (n = 0, 1, \dots), \tag{3}$$

где I – единичный оператор, u_0 – начальное приближение и $0 < \alpha^2 < 2/\|B\|^2$. Чтобы получить хорошее приближение к решению некорректной задачи (I), надо остановить итерации при таком n, при котором будет выполнено одно из правил останова, предложенных в [I] и учитывающих неточность задания f и h.

При выборе параметра α^2 метода (3) понадобится оценка нормы оператора B. Эту оценку можно найти следующим образом. Пусть $u \in \mathcal{H}_1$. По определению оператора B

$$\|Bu\|_{\mathcal{H}_2} = \Big(\int_0^T \|u' + Au\|^2_{V'}\,dt + \|u(T)\|^2_H \Big)^{1/2}. \tag{4}$$

Оценим сначала норму $\|u(T)\|_H$. Так как $\|u(t)\|_H$ непрерывна на отрезке $t \in [0, T]$, то существует $t_0 \in [0,T]$ такое, что $\|u(t_0)\|_H = \min\limits_{0\le t\le T} \|u(t)\|_H$. Тогда

$$\|u(T)\|^2_H = \|u(t_0)\|^2_H + \int_{t_0}^T \frac{d}{dt}\|u(t)\|^2_H\,dt = \|u(t_0)\|^2_H +$$

$$+ 2\int_0^T (u'(t), u(t))\,dt \le \frac{1}{T}\int_0^T \|u(t)\|^2_H\,dt + \int_0^T \|u'(t)\|^2_{V'}\,dt + \int_0^T \|u(t)\|^2_V\,dt.$$

Если $\|u\|_H \le c\,\|u\|_V$, то

$$\|u(T)\|^2_H \le \Big(\frac{c^2}{T} + 1\Big)\|u\|^2_{\mathcal{H}_1}.$$

Так как

$$\|u' + Au\|^2_{V'} \le (\|u'\|_{V'} + \|A\|\,\|u\|_V)^2 \le (1 + \|A\|^2)(\|u'\|^2_{V'} + \|u\|^2_V),$$

то из (4) получаем оценку

$$\|Bu\|_{\mathcal{H}_2} \le \Big(2 + \|A\|^2 + \frac{c^2}{T}\Big)^{1/2} \|u\|_{\mathcal{H}_1}.$$

Таким образом

$$\|B\| \le \Big(2 + \|A\|^2 + \frac{c^2}{T}\Big)^{1/2}. \tag{5}$$

При реализации метода (3) надо применить сопряженный оператор B^*. Элемент $w = B^* g$ найдем из соотношения

$$(Bu, g)_{\mathcal{H}_2} = (u, w)_{\mathcal{H}_1} \qquad \forall u \in \mathcal{H}_1.$$

Учитывая, что

$$(Bu, g)_{\mathcal{H}_2} = \int_0^T (u' + Au, f)_{V'} dt + (u(T), h)_H = -\int_0^T (J^{-1}f', u) dt +$$
$$+ \int_0^T (A^* J^{-1} f, u) dt + (J^{-1} f(T), u(T)) - (J^{-1} f(0), u(0)) + (h, u(T))$$

и

$$(u, w)_{\mathcal{H}_1} = \int_0^T [(u, w)_V + (u', w')_{V'}] dt = \int_0^T (Jw, u) dt -$$
$$- \int_0^T (J^{-1} w'', u) dt + (J^{-1} w'(T), u(T)) - (J^{-1} w'(0), u(0)),$$

получим для вычисления $w = B^* g$ задачу

$$-J^{-1} w'' + Jw = -J^{-1} f' + A^* J^{-1} f, \tag{6}$$
$$J^{-1} w'(0) = J^{-1} f(0), \quad J^{-1} w'(T) = J^{-1} f(T) + h.$$

2. Рассмотрим более подробно реализацию метода последовательных приближений (3) в случае, если $A = a^2 J + b\,\mathrm{I}$, где $a^2 > 0$ и b – постоянные и I – единичный оператор. В этом случае выполнено условие коэрцитивности (2). Приближенные решения задачи (I) ищем в виде разложений по собственным элементам канонического изоморфизма J.

Изоморфизм J является самосопряженным положительно определенным оператором в H. Предложим, что любое ограниченное в V множество компактно в H. Тогда J^{-1} вполне непрерывен и собственные элементы φ_p $(p = 1, 2, \ldots)$ оператора J образуют последовательность, полную в V, H и V' (см. [3], стр. 195). Упорядочим собственные значения λ_p оператора J: $0 < \lambda_1 \le \lambda_2 \le \ldots$. Собственные элементы φ_p считаем ортонормированными в H, т.е. $(\varphi_p, \varphi_q) = \delta_{pq}$, где δ_{pq} – символ Кронекера. Тогда $(\varphi_p, \varphi_q)_V = (J\varphi_p, \varphi_q) = \lambda_p \delta_{pq}$ и $(\varphi_p, \varphi_q)_{V'} = \lambda_p^{-1} \delta_{pq}$.

При численном решении задачи (I) заменяем $g = (f, h)$ на $g_\delta = (f_\delta, h_\delta)$, где

$$f_\delta = \sum_{p=1}^{N} (f, \varphi_p)\varphi_p, \qquad h_\delta = \sum_{p=1}^{N} (h, \varphi_p)\varphi_p .$$

При этом N выбираем так, чтобы было

$$\|g - g_\delta\|^2_{\mathcal{H}_2} = \int_0^T \|f - f_\delta\|^2_{V'} dt + \|h - h_\delta\|^2_H \le \delta^2,$$

где δ – уровень точности задания данных.

В таком случае задача (6) для определения $w = B^* g_\delta$ принимает вид

$$-J^{-1} w'' + Jw = \sum_{p=1}^{N} \left[-\frac{1}{\lambda_p}(f', \varphi_p) + \left(a^2 + \frac{b}{\lambda_p}\right)(f, \varphi_p) \right] \varphi_p,$$

$$J^{-1} w'(0) = \sum_{p=1}^{N} \frac{1}{\lambda_p} (f(0), \varphi_p)\varphi_p, \quad J^{-1} w'(T) = \sum_{p=1}^{N} \left[\frac{1}{\lambda_p}(f(T), \varphi_p) + (h, \varphi_p)\right]\varphi_p .$$

Ищем решение последней задачи в виде

$$w = \sum_{p=1}^{N} w_p(t)\varphi_p .$$

Подставив последнее разложение в задачу, получим для определения коэффициентов w_p краевую задачу

$$-w_p'' + \lambda_p^2 w_p = -(f', \varphi_p) + (a^2\lambda_p + b)(f, \varphi_p),$$

$$w_p'(0) = (f(0), \varphi_p), \qquad w_p'(T) = (f(T), \varphi_p) + \lambda_p (h, \varphi_p).$$

Решением этой задачи является

$$w_p = \frac{\operatorname{ch}\lambda_p(T-t)}{\operatorname{sh}\lambda_p T} \int_0^t (f(s), \varphi_p)\left[\operatorname{sh}\lambda_p s + \left(a^2 + \frac{b}{\lambda_p}\right)\operatorname{ch}\lambda_p s\right] ds -$$

$$- \frac{\operatorname{ch}\lambda_p t}{\operatorname{sh}\lambda_p T} \int_t^T (f(s), \varphi_p)\left[\operatorname{sh}\lambda_p (T-s) - \left(a^2 + \frac{b}{\lambda_p}\right)\operatorname{ch}\lambda_p (T-s)\right] ds +$$

$$+ \frac{\operatorname{ch}\lambda_p t}{\operatorname{sh}\lambda_p T}(h, \varphi_p) \qquad (p = 1, 2, \dots, N). \tag{7}$$

Возьмем начальное приближение в виде

$$u_0 = \sum_{p=1}^{N} u_{0p}(t)\varphi_p .$$

Тогда и последовательные приближения, определяемые формулой (3), выражаются в виде

$$u_n = \sum_{p=1}^{N} u_{np}(t)\varphi_p \qquad (n = 1, 2, \dots, N).$$

При помощи формулы (7) получаем для вычисления коэффициентов последнего разложения рекуррентную формулу

$$u_{n+1,p}(t) = u_{np}(t) + \alpha^2 \Big\{ w_p(t) - u_{np}(t) - \frac{1}{\lambda_p}\left[(a^2\lambda_p + b)^2 - \lambda_p^2\right] \times$$

$$\times \left[\frac{ch\,\lambda_p(T-t)}{sh\,\lambda_p T} \int_0^t u_{np}(s)\, ch\,\lambda_p s\, ds + \frac{ch\,\lambda_p t}{sh\,\lambda_p T} \int_t^T u_{np}(s)\, ch\,\lambda_p(T-s)\, ds \right] -$$

$$-\left(a^2 + \frac{b}{\lambda_p} + 1\right) \frac{ch\,\lambda_p t}{sh\,\lambda_p T}\, u_{np}(T) + \left(a^2 + \frac{b}{\lambda_p}\right) \frac{ch\,\lambda_p(T-t)}{sh\,\lambda_p T}\, u_{np}(0) \Big\}. \qquad (8)$$

Последняя формула принимает особенно простой вид, если $a^2 = 1$ и $b = 0$ (см. [5]).

3. Рассмотрим еще следующий пример обратной задачи теплопроводности:

$$\frac{\partial u}{\partial t} = a^2 \frac{\partial^2 u}{\partial x^2} - bu + f(x,t),$$
$$u(x,T) = h(x), \quad u(0,t) = u(1,t) = 0. \qquad (9)$$

Предположим, что $a^2 = const > 0$, $b = const$, $h \in L_2(0,1)$ и $f \in L_2[(0,1) \times (0,1)]$. Тогда можно взять за пространства $H = L_2(0,1)$ и $V = H_c^1(0,1)$, в которых метрики определены скалярными произведениями

$$(u,v) = \int_0^1 u v\, dx, \qquad (u,v)_V = \int_0^1 \frac{\partial u}{\partial x}\frac{\partial v}{\partial x}\, dx.$$

В таком случае $Ju = -\partial^2 u/\partial x^2$ и задачу (9) можно рассматривать как задачу (I), в которой $A = a^2 J + bI$. Так как $\|u\|_H \le \frac{1}{2}\|u\|_V$, то при помощи формулы (5) получаем оценку

$$\|B\| \le \left[2 + \left(a^2 + \frac{|b|}{4}\right)^2 + \frac{1}{4T}\right]^{1/2}.$$

Собственными значениями и функциями оператора J являются

$$\lambda_p = (\pi p)^2, \quad \varphi_p = \sqrt{2}\, \sin \pi p x \quad (p = 1, 2, \ldots).$$

При реализации изложенного алгоритма на ЭВМ надо учесть то, что вместе с ростом p значения гиперболических функции в формулах (7) и (8) довольно быстро растут и могут вызвать переполнение. Во избежание переполнения можно пользоваться приближенными формулами типа

$$\frac{ch\,\lambda_p(T-t)}{sh\,\lambda_p T}\, sh\,\lambda_p s \approx \frac{e^{\lambda_p(T-t)} e^{\lambda_p s}}{2 e^{\lambda_p T}} = \frac{1}{2} e^{-\lambda_p(t-s)}$$
$$(0 < s \le t < T).$$

Литература

I. В а й н и к к о Г.М., Оценки погрешности метода последовательных приближений для некорректных задач. Автоматика и телемеханика, I980, № 3, 84-92.

2. Л а т т е с Р., Л и о н с Ж.-Л., Метод квазиобращения и его приложения. Москва, I980.

3. М и х л и н С.Г., Вариационные методы в математической физики. Москва, I957.

4. О б е н Ж.-П., Приближенное решение эллиптических краевых задач. Москва, I977.

5. Т а м м е Э., О решении обратной задачи теплопроводности. Тезисы конференции "Теоретические и прикладные вопросы математики". Тарту, I980, I73-I75.

6. M i l l e r, K., Efficient numerical methods for backward solution of parabolic problems with variable coefficients. Res. Notes Math., 1975, № 1, 54-64.

Поступило
25 I 1982

THE SOLUTION OF THE HEAT EQUATION BACKWARDS IN TIME WITH THE SUCCESSIVE APPROXIMATION METHOD

E. Tamme

Summary

The heat equation backwards in time is an ill-posed problem. In this paper the application of the successive approximation method for the numerical solution of this problem is considered.

Уч.зап.Тартуск.ун-та,1982,633,41-49

ДВУХШАГОВЫЕ α-ПРОЦЕССЫ И ИХ ПРИМЕНЕНИЕ ДЛЯ РЕШЕНИЯ НЕКОРРЕКТНЫХ ЗАДАЧ

Л.Сарв

I. Рассмотрим уравнение

$$Au = f, \tag{I}$$

где $A \in \mathcal{L}(U,F)$ -линейный ограниченный оператор из гильбертова пространства U в гильбертово пространство F. Пусть нулевое пространство $N(A)$, вообщем, нетривиально и область значений $R(A)$. вообщем, не замкнута. Следовательно, задача. (I) может оказатся некорректно поставленной [2,6]. Предполагается, что $f \in R(A) \subset F$. Задачу (I) будем решать следующим нелинейным итерационным методом:

$$\begin{gathered} u_{n+1} = u_n - \varepsilon_{1,n} A^* r_n + \varepsilon_{2,n} A^*AA^* r_n \\ \varepsilon_{1,n} = (b_{\alpha+4}\, b_{\alpha+1} - b_{\alpha+3}\, b_{\alpha+2})/(b_{\alpha+4}\, b_{\alpha+2} - b^2_{\alpha+3}), \\ \varepsilon_{2,n} = (b_{\alpha+3}\, b_{\alpha+1} - b^2_{\alpha+2})/(b_{\alpha+4}\, b_{\alpha+2} - b^2_{\alpha+3}). \end{gathered} \tag{2}$$

Здесь $r_n = Au_n - f$, $b_{\nu,n} = b_\nu = (B^\nu r_n, r_n)$, $B = AA^*$, $\alpha \geq -1$ целое число. Заметим, что параметры $\varepsilon_{1,n}$ и $\varepsilon_{2,n}$ определяются таким образом, чтобы момент порядка α от очередной невязки был бы минимальным: $(B^\alpha r_{n+1}, r_{n+1}) \to \min$. Для $\alpha = -1$ последнее эквивалентно условию $(\psi_{n+1}, \psi_{n+1}) \to \min$, где $\psi_{n+1} = u_{n+1} - u_*$; u_* - ближайшее к u_0 решение уравнения (I). Итак, при $\alpha = -1$ метод (2) является двухшаговым методом минимальных ошибок, при $\alpha = 0$ двухшаговым методом наискорейшего спуска и при $\alpha = 1$ двухшаговым методом минимальных невязок. Для $\alpha = 0$ метод (2) был предложен в [3]. Идея введения параметра α для одношагового случая была предложена в [4].

Исходя из задачи (I) рассмотрим случай, когда вместо оператора A и правой части f даны их приближения $A_\eta \in \mathcal{L}(U,F)$ и $f_\delta \in F$, такие что $\|A - A_\eta\| \leq \eta$ и $\|f - f_\delta\| \leq \delta$, $\eta \in (0,\eta_0)$, $\delta \in (0,\delta_0)$. Для решения приближенной задачи также применяем метод (2). Понятно, что в таком случае в (2) вместо A фигурирует A_η а вместо f - f_δ.

Приведём основные результаты, которые справедливы для $\alpha \geq 0$. Относительно начального приближения u_0 будем предполагать, что u_0-u_* не является линейной комбинацией двух или менее собственных элементов оператора A^*A.

Теорема I. При решении уравнения (I) методом (2) в случае точно заданных данных имеем $\lim_{n\to\infty}\|u_n-u_*\|=0$.

Если при этом $u_0-u_*=(A^*A)^p v$, $p>0$, $\|v\|\le q$

то $\|u_n-u_*\|\le C_{p,q}n^{-p}$.

Следующие теоремы относится к случаю, когда A и f даны приближенно.

Теорема 2. Если $n=n(\delta,\eta)$ выбрать так, чтобы $n\to\infty$ и $(\delta+\eta)n^{1/2}\to 0$ при $\delta,\eta\to 0$ то

$$\|u_n-u_*\|\to 0 \qquad \text{при} \quad \delta,\eta\to 0. \tag{3}$$

Если при этом известно, что

$$u_0-u_*=(A^*A)^p v \quad p>0,\ v\in U,\ \|v\|\le q, \tag{4}$$

и выбрать $n=c_{p,q}(\delta+\eta)^{-2/(2p+1)}$, то можно оценить ошибку

$$\|u_n-u_*\|\le c'_{p,q}(\delta+\eta)^{2p/(2p+1)} \tag{5}$$

Сформулируем теперь два правила останова для итераций(2) по невязке [I].

Правило останова ($П_I$): зададим числа $b_1>1$ и $b_*>\|u_*\|$; итерации остановим на таком $n=n(\delta,\eta)$ для которого впервые

$$\|r_n\|\le b_1\delta+b_*\eta.$$

Правило останова ($П_2$): зададим числа $b_1>1$, $b_2>1$ и $\alpha>0$; итерации остановим на таком $n=n(\delta,\eta)$ для которого впервые выполнено хотя бы одно из неравенств

$$\|r_n\|\le b_1\delta+b_2\|u_n\|\eta,\quad n\ge \alpha/(b_1\delta+b_2\|u_n\|\eta)^2.$$

Теорема 3. Для правил останова ($П_I$) и ($П_2$) при $\delta,\eta\to 0$ справедливо

$$(\delta+\eta)n^{1/2}\to 0,\quad \|u_n-u_*\|\to 0. \tag{6}$$

Если выполнено (4), то

$$n(\delta,\eta)\le c_{p,q}(\delta+\eta)^{-2/(2p+1)}, \tag{7}$$

$$\|u_n-u_*\|\le c'_{p,q}(\delta+\eta)^{2p/(2p+1)}. \tag{8}$$

2. Пусть A и f заданы точно. Для доказательства теорем этого пункта нам понадобится

Лемма I. Итерационный метод (2) имеет следующие свойства:

$$(B^{\alpha+1} r_n, r_{n+1}) = (B^{\alpha+2} r_n, r_{n+1}) = 0\,, \tag{9}$$

$$(B^{\alpha} r_n, r_{n+2}) = (B^{\alpha} r_n, r_{n+1}) = (B^{\alpha} r_{n+1}, r_{n+1})\,, \tag{I0}$$

$$(B^{\alpha+1} r_n, r_{n+2}) = (\varepsilon_{2,n+1}/\varepsilon_{2,n})(B^{\alpha+1} r_{n+1}, r_{n+1})\,, \tag{II}$$

$$(B^{\alpha} r_n, r_n) - (B^{\alpha} r_{n+1}, r_{n+1}) = \varepsilon_{2,n}(B^{\alpha+1} r_n, r_n)/\beta_{1,n}\,, \tag{I2}$$

$$\varepsilon_{2,n} - \beta_{0,n}\beta_{1,n} \geqslant 0\,, \tag{I3}$$

$$1/\|B\|^2 \le \varepsilon_{1,n}/\|B\| \le \varepsilon_{2,n} \le \gamma\,, \tag{I4}$$

$$b_1 - \varepsilon_{1,n} b_2 + \varepsilon_{2,n} b_3 \geqslant 0\,, \tag{I5}$$

$$\varepsilon_{2,n} \le (b_3 b_1 - b_2^2)/(b_4 b_2 - b_3^2)\,, \tag{I6}$$

$$\|r_n\|^2 - \|r_{n+1}\|^2 \geqslant c\,(B r_n, r_n)\,, \tag{I7}$$

где $\beta_{0,n} = b_{\alpha+1}/b_{\alpha+2}$, $\beta_{1,n} = b_{\alpha+1}(b_{\alpha+3} b_{\alpha+1} - b_{\alpha+2}^2)/$
$/(b_{\alpha+2}^2 - 2 b_{\alpha+1} b_{\alpha+2} b_{\alpha+3} + b_{\alpha+4} b_{\alpha+1}^2)$,

и γ являетоя константой, не зависящий от n.

Доказательство. Из (2) получим

$$r_{n+1} = r_n - \varepsilon_{1,n} B r_n + \varepsilon_{2,n} B^2 r_n\,. \tag{I8}$$

Равенства (9), (I0), (II) и (I2) можно теперь непосредственно получить из (I8). Докажем, например, (II):

$$(B^{\alpha+1} r_n, r_{n+2}) = (B^{\alpha+1} r_n, r_{n+1} - \varepsilon_{1,n+1} B r_{n+1} +$$
$$+ \varepsilon_{2,n+1} B^2 r_{n+2}) = \varepsilon_{2,n+1}(B^{\alpha+1} r_n, B^2 r_{n+1}) = \varepsilon_{2,n+1}(B^2 r_n, B^{\alpha+1} r_{n+1}) =$$
$$= (\varepsilon_{2,n+1}/\varepsilon_{2,n})(r_{n+1} - r_n + \varepsilon_{1,n} B r_n, B^{\alpha+1} r_{n+1}) =$$
$$= (\varepsilon_{2,n+1}/\varepsilon_{2,n})(B^{\alpha+1} r_{n+1}, r_{n+1}).$$

Отметим следующий факт. Если при данном u_n провести два шага одношагового α -процесса [4] и невязки для них обозначить через ρ_{n+1} и ρ_{n+2}, то

$$\rho_{n+1} = r_n - \beta_{0,n} B r_n\,,$$

$$\rho_{n+2} = \rho_{n+1} - \beta_{1,n} B \rho_{n+1} = r_n - (\beta_{0,n} + \beta_{1,n}) B r_n + \beta_{0,n}\beta_{1,n} B^2 r_n\,.$$

Из последнего соотношения в соответствии с выбором $\varepsilon_{1,n}$ и

$\mathcal{E}_{2,n}$ имеем

$$(B^{\alpha} r_{n+1}, r_{n+1}) \le (B^{\alpha} \rho_{n+2}, \rho_{n+2}). \qquad (19)$$

Непосредственный контроль показывает что

$$\mathcal{E}_{1,n}/\mathcal{E}_{2,n} = (\beta_{0,n} + \beta_{1,n})/(\beta_{0,n}\, \beta_{1,n}). \qquad (20)$$

Если выписать (19) и учесть (20), то получим (13).

Первое неравенство в (14) получаем следующим образом

$$b_{\alpha+4} b_{\alpha+2} - b_{\alpha+3}^2 = b_{\alpha+4} \int_0^{\|B\|} \lambda^{\alpha+2} d\|E_\lambda r_n\|^2 - b_{\alpha+3} \int_0^{\|B\|} \lambda^{\alpha+3} d\|E_\lambda r_n\|^2 =$$

$$= \int_0^{\|B\|} \lambda (b_{\alpha+4} \lambda^{\alpha+1} - b_{\alpha+3} \lambda^{\alpha+2}) d\|E_\lambda r_n\|^2 \le$$

$$\le \|B\| (b_{\alpha+4} \int_0^{\|B\|} \lambda^{\alpha+1} d\|E_\lambda r_n\|^2 - b_{\alpha+3} \int_0^{\|B\|} \lambda^{\alpha+2} d\|E_\lambda r_n\|^2) =$$

$$= \|B\| (b_{\alpha+4} b_{\alpha+1} - b_{\alpha+3} b_{\alpha+2}),$$

где E_λ – спектральная функция оператора B.

Второе неравенство в (14) доказывается аналогично.

Далее обозначим $a_n = \|B^{(\alpha+1)/2} r_n\|$. Из (12) находим

$$\mathcal{E}_{2,n+1}/\mathcal{E}_{2,n} \le a_{n+2}\, a_n / a_{n+1}^2. \qquad (21)$$

Перемножая неравенства (21) с $n = 0, 1, \dots, k-1$ получим

$$\mathcal{E}_{2,k} \le \mathcal{E}_{2,0} (a_{k+1}/a_k)(a_0/a_1). \qquad (22)$$

С другой стороны, учитывая (12) имеем

$$0 \le (B^{\alpha}(r_k - r_{k+2}), r_k - r_{k+2}) = (B^{\alpha} r_k, r_k) - 2(B^{\alpha} r_{k+1}, r_{k+1}) +$$

$$+ (B^{\alpha} r_{k+2}, r_{k+2}) = \mathcal{E}_{2,k} (B^{\alpha+1} r_k, r_k)/\beta_{1,k} - \mathcal{E}_{2,k+1} (B^{\alpha+1} r_{k+1}, r_{k+1})/\beta_{1,k+1} =$$

$$= \mathcal{E}_{2,k}\, a_k^2/\beta_{1,k} - \mathcal{E}_{2,k+1}\, a_{k+1}^2/\beta_{1,k+1}.$$

Последнее, вместе с (22) и (13) даёт

$$\mathcal{E}_{2,k} \le \mathcal{E}_{2,0}^2 [\beta_{1,k}/(\mathcal{E}_{2,k+1} \beta_{1,k})] \cdot (a_0^2/a_1^2) \le \mathcal{E}_{2,0}^2 [1/(\beta_{0,k+1}\, \beta_{1,k})] \cdot$$

$$\cdot (a_0^2/a_1^2) \le \|B\|^2 \mathcal{E}_{2,0} (B^{\alpha+1} r_0, r_0)/(B^{\alpha+1} r_1, r_1) = \gamma.$$

Здесь мы учитываем, что $(B^{\alpha+1} r_1, r_1) \ne 0$. Действительно иначе r_0 является линейной комбинацией двух собственных элементов оператора AA^*, а это исключено по предположению. Тем самым (14) доказано.

Теперь докажем (15). Для $\alpha = 0$ и $\alpha = 1$ неравенство (15) непосредственно следует из (2) и (10). Так как $b_{\alpha+4} b_{\alpha+2} - b^2_{\alpha+3} \geqslant 0$ по неравенству моментов [4], то (15) эквивалентно неравенству

$$b_1(b_{\alpha+4} b_{\alpha+2} - b^2_{\alpha+3}) - b_2(b_{\alpha+4} b_{\alpha+1} - b_{\alpha+3} b_{\alpha+2}) + b_3(b_{\alpha+3} b_{\alpha+1} - b^2_{\alpha+2}) \geqslant 0 . \qquad (23)$$

Выпишем первое из слагаемых

$$b_1(b_{\alpha+4} b_{\alpha+2} - b^2_{\alpha+3}) = \int_0^{\|B\|} \nu \, d\|E_\nu r_n\|^2 \Big[\int_0^{\|B\|} \lambda^{\alpha+4} d\|E_\lambda r_n\|^2 \cdot \int_0^{\|B\|} \mu^{\alpha+2} d\|E_\mu r_n\|^2 - \int_0^{\|B\|} \lambda^{\alpha+3} d\|E_\lambda r_n\|^2 \cdot \int_0^{\|B\|} \mu^{\alpha+3} d\|E_\mu r_n\|^2 \Big] = \int_0^{\|B\|}\int_0^{\|B\|}\int_0^{\|B\|} \lambda^{\alpha+4}\mu^{\alpha+2}\nu \, d\|E_\lambda r_n\|^2 \cdot d\|E_\mu r_n\|^2 d\|E_\nu r_n\|^2 - \int_0^{\|B\|}\int_0^{\|B\|}\int_0^{\|B\|} \lambda^{\alpha+3}\mu^{\alpha+3}\nu \, d\|E_\lambda r_n\|^2 d\|E_\mu r_n\|^2 \cdot d\|E_\nu r_n\|^2 = (1/6)\int_0^{\|B\|}\int_0^{\|B\|}\int_0^{\|B\|} \big[\lambda^{\alpha+4}\mu^{\alpha+2}\nu + \lambda^{\alpha+4}\mu\nu^{\alpha+2} + \lambda^{\alpha+2}\mu^{\alpha+4}\nu + \lambda^{\alpha+2}\mu\nu^{\alpha+4} + \lambda\mu^{\alpha+4}\nu^{\alpha+2} + \lambda\mu^{\alpha+2}\nu^{\alpha+4} - 2\lambda^{\alpha+3}\mu^{\alpha+3}\nu - 2\lambda^{\alpha+3}\mu\nu^{\alpha+3} - 2\lambda\mu^{\alpha+3}\nu^{\alpha+3} \big] \cdot d\|E_\lambda r_n\|^2 d\|E_\mu r_n\|^2 d\|E_\nu r_n\|^2 .$$

Если аналогичным же образом представить и два следующих слагаемых в (23), а подинтегральную функцию разложить на множители, то получим

$$\int_0^{\|B\|}\int_0^{\|B\|}\int_0^{\|B\|} \Big[\lambda\mu\nu(\lambda-\mu)^2(\lambda-\nu)^2(\mu-\nu)^2 \sum_{\substack{i,j,k \geqslant 0 \\ i+j+k=2\alpha-2}} \lambda^i \mu^j \nu^k \Big] d\|E_\lambda r_n\|^2 d\|E_\mu r_n\|^2 \cdot d\|E_\nu r_n\|^2 \geqslant 0 .$$

Доказательство (16) аналогичное и мы его проводить не будем.

Для доказательства (17) покажем сперва, что

$$\varepsilon_{1,n}(b_1 - \varepsilon_{1,n} b_2 + \varepsilon_{2,n} b_3) - \varepsilon_{2,n}(b_2 - \varepsilon_{1,n} b_3 + \varepsilon_{2,n} b_4) \geqslant 0 . \qquad (24)$$

В силу (13), (15) и (20) справедливо

$$b_3/b_2 \leqslant \varepsilon_{1,n}/\varepsilon_{2,n} \leqslant [b_3 + (b_1/\varepsilon_{2,n})]/b_2 . \qquad (25)$$

Обозначим $x = \varepsilon_{1,n}/\varepsilon_{2,n}$. Перепишем (25) в виде

$$[2b_3 + (b_1/\varepsilon_{2,n}) - \sqrt{(b_1/\varepsilon_{2,n})^2}]/2b_2 \leqslant x \leqslant [2b_3 + (b_1/\varepsilon_{2,n}) + \sqrt{(b_1/\varepsilon_{2,n})^2}]/2b_2 . \qquad (26)$$

Из (26) вытекает, что для любого $A \geqslant 0$ число x удовлетворяет неравенству

$$[2b_3+(b_1/\varepsilon_{2,n})-\sqrt{(b_1/\varepsilon_{2,n})+A}\,]/2b_2 \leq x \leq$$
$$\leq [2b_3+(b_1/\varepsilon_{2,n})+\sqrt{(b_1/\varepsilon_{2,n})+A}\,]/2b_2 \,. \tag{27}$$

Положим $A=4[(b_3^2-b_4 b_2)\varepsilon_{2,n}+b_3 b_1-b_2^2]/\varepsilon_{2,n}$; $A \geqslant 0$ в силу (I6). Теперь (27) означает, что x удовлетворяет неравенству

$$b_2 x^2 - x[2b_3+(b_1/\varepsilon_{2,n})]+(b_2/\varepsilon_{2,n})+b_4 \leq 0 \,.$$

Если учесть, что $x=\varepsilon_{1,n}/\varepsilon_{2,n}$ то отсюда непосредственно вытекает (24).

Докажем теперь (I7). В силу (I3), (I8), (20), (24) и неравенства моментов получаем

$$\|r_n\|^2-\|r_{n+1}\|^2=\varepsilon_{1,n}b_1-\varepsilon_{2,n}b_2+\varepsilon_{1,n}(b_1-\varepsilon_{1,n}b_2+\varepsilon_{2,n}b_3)-\varepsilon_{2,n}(b_2-$$
$$-\varepsilon_{1,n}b_3+\varepsilon_{2,n}b_4)\geqslant\varepsilon_{1,n}b_1-\varepsilon_{2,n}b_2=\varepsilon_{2,n}b_1[(\varepsilon_{1,n}/\varepsilon_{2,n})-(b_2/b_1)]\geqslant$$
$$\geqslant\varepsilon_{2,n}b_1[(\varepsilon_{1,n}/\varepsilon_{2,n})-(1/\beta_{0,n})]=\varepsilon_{2,n}b_1/\beta_{1,n}\geqslant\beta_{0,n}b_1\geqslant$$
$$\geqslant b_1/\|B\|=(Br_n,r_n)/\|B\| \,.$$

Этим доказательство леммы 2 завершено.

<u>Доказательство теоремы I</u>. Пусть F_λ спектральная функция оператора $A^*A=\mathcal{D}$ Для любого $t\in(0,\|\mathcal{D}\|)$ имеем

$$\|\Psi_n\|\leqslant\|F_t\Psi_n\|+\|A\Psi_n\|/t^{1/2}=$$
$$=\Big[\int_0^t(1-\varepsilon_{1,n}\lambda+\varepsilon_{2,n}\lambda^2)^2 d\|F_\lambda\Psi_{n-1}\|^2\Big]^{1/2}+\|r_n\|/t^{1/2} \,.$$

Если теперь фиксировать $t_0=1/(\|B\|\cdot\gamma)$ и взять $t\in(0,t_0)$, то $(1-\varepsilon_{1,n}t+\varepsilon_{2,n}t^2)^2\leq 1$ и следовательно

$$\|\Psi\|\leqslant\|F_t\Psi_0\|+\|r_n\|/t^{1/2} \,. \tag{28}$$

Аналогично получим, что для

$$\|r_n\|\leqslant\|E_t r_0\|+(Br_n,r_n)^{1/2}/t^{1/2}\leqslant$$
$$\leqslant t^{1/2}\|\Psi_0\|+(Br_n,r_n)^{1/2}/t^{1/2} \,.$$

Минимизируя правую часть последней оценки по t, имеем

$$\|r_n\|\leqslant(Br_n,r_n)^{1/4}/\|\Psi_0\|^{1/2} ,$$

или

$$(Br_n,r_n)\geqslant c_1\|r_n\|^4 \,.$$

что вместе с (I7) даёт $\|r_n\|^2-\|r_{n+1}\|^2\geqslant c_2\|r_n\|^4$. Анализ этой оценки приводит к неравенству $\|r_n\|\leqslant c_3 n^{-1/2}$. Подставляя последнее в (28), получим

$$\|\Psi_n\|\leqslant\|F_t\Psi_0\|+c_3/(nt)^{1/2}$$

Учитывая, что $\Psi_0\perp N(A)$, можно для любого $\varepsilon>0$ фиксировать

$t_* \in (0, t_0)$ такой что $\|F_{t_*}\Psi_0\| \leq \varepsilon/2$. Затем фиксируем $N = [4c_3^2/\varepsilon^2 t_*]+1$ и получим, что $\|\Psi_n\| \leq \varepsilon$ для $n \geq N$. Пусть теперь $u_0 - u_* = (A^*A)^p v$. Из (28) тогда получим

$$\|\Psi_n\| \leq t^p\|v\| + \|z_n\|/t^{1/2}.$$

После минимизации по t имеем

$$\|\Psi_n\| \leq c_4\|z_n\|^{2p/(2p+1)}. \tag{29}$$

Действуя аналогично для $\|z_n\|$, имеем

$$(Bz_n, z_n) \geq c_5\|z_n\|^{(4p+4)/(2p+1)}. \tag{30}$$

По теореме Лагранжа о среднем, примененной к $f(x) = x^{2p+1}$, получаем оценку

$$\|z_n\|^2 - \|z_{n+1}\|^2 \leq (2p+1)\|z_n\|^{4p/(2p+1)}\left[\|z_n\|^{2/(2p+1)} - \|z_{n+1}\|^{2/(2p+1)}\right]. \tag{31}$$

Неравенства (30) и (31) вместе с (17) дают

$$\|z_n\|^{2/(2p+1)} - \|z_{n+1}\|^{2/(2p+1)} \geq c_6\|z_n\|^{4/(2p+1)}. \tag{32}$$

В свою очередь (32), как можно показать, влечет оценку

$$\|z_n\| \leq c_7\, n^{-(2p+1)/2}$$

и отсюда в силу (29) следует $\|\Psi_n\| \leq c_{p,q}\, n^{-p}$. Теорема I доказана.

3. Рассмотрим теперь приближенную задачу. Пусть $B_\eta = A_\eta A_\eta^*$, $\mathcal{D}_\eta = A_\eta^* A_\eta$, E_λ и F_λ – спектральные функции операторов B_η и $\mathcal{D}_\eta$ соответственно.

Лемма 2. При решении приближённой задачи с использованием метода (2) найдутся константы c', c'' и t_0 такие что

$$\|z_n\|^2 - \|z_{n+1}\|^2 \geq c'(B_\eta z_n, z_n), \tag{33}$$

а для всех $t \in (0, t_0)$,

$$\|F_t\Psi_n\| \leq \|F_t\Psi_{n-1}\| + c'' t^{1/2}\|A_\eta u_* - f_\delta\|, \tag{34}$$

$$\|E_t z_n\| \leq \|E_t z_{n-1}\|. \tag{35}$$

Доказательство. Сперва заметим, что из (2) следует

$$z_{n+1} = z_n - \varepsilon_{1,n} B_\eta z_n + \varepsilon_{2,n} B_\eta^2 z_n, \tag{36}$$

где $z_n = A_\eta u_n - f_\delta$. Сравнивая теперь (36) с (18) можно показать, что все утверждения леммы I переносятся на случай приближенно заданых данных. По условию $u_0 - u_*$ не является линейной комбинацией двух собственных элементов оператора A^*A. Нетрудно заметить, что при достаточно малых δ, η, начальная ошибка $u_0 - u_*$ не будет линейной комбинацией двух собственных элементов и для оператора $A_\eta^* A_\eta$ и тогда итерации существуют при всех n, т.е. $r_n \neq 0$, $n = 1, 2, \ldots$. Отсюда следует ещё тот факт, что константу γ, фигурирующую в (I4), можно здесь считать не зависящей от δ и η. Итак, (33) получается точно так как и (I7). Из (36) следует, что для $t_0 = 1/(\|A\| + \eta_0)^2 \gamma$ верна оценка (35). Действительно,

$$\|E_t r_n\| = \left[\int_0^t (1 - \lambda\varepsilon_{1,n} + \lambda^2\varepsilon_{2,n})^2 d\|E_\lambda r_{n-1}\|^2\right]^{1/2} \leq \|E_t r_{n-1}\|,$$

так как $(1 - \lambda\varepsilon_{1,n} + \lambda^2\varepsilon_{2,n})^2 \leq 1$ для $\lambda \in (0, t_0)$. Из (2) имеем

$$\|F_t \psi_n\| \leq \left[\int_0^t (1 - \lambda\varepsilon_{1,n-1} + \lambda^2\varepsilon_{2,n-1})^2 d\|F_\lambda \psi_{n-1}\|^2\right]^{1/2} +$$

$$+ t^{1/2}\left[\int_0^t (\varepsilon_{1,n-1} - \varepsilon_{2,n-1}\lambda)\, d\|E_\lambda(f_\delta - A_\eta u_*\|^2\right]^{1/2}.$$

Учитывая, что $(\varepsilon_{1,n-1} - \varepsilon_{2,n-1}\lambda)^2 \leq [((\|A\| + \eta_0)^2\gamma]^2$ при $\lambda \in [0, t_0]$, из последнего получим (34). Лемма I доказана.

Для доказательства теорем 2 и 3 теперь достаточно сослаться на теорему 3 из [5]. В заключение автор выражает глубокую благодарность Г.Вайникко за руководство работой.

Литература

1. В а й н и к к о Г.М., Оценки скорости сходимости метода последовательных приближений для некорректных задач. Автоматика и телемеханика, I980, № 3, 84-92.
2. И в а н о в В.К., В а с и н В.В., Т а н а н а В.П., Теория линейных некорректных задач и её приложение. Москва, I978.
3. К а н т о р о в и ч Л.В., О методе наискорейшего спуска. Докл. АН СССР, I974, 56, № 3, 233-236.
4. К р а с н о с е л ь с к и й М.А., В а й н и к к о Г.М. и др., Приближённое решение операторных уравнений. Москва, I969.
5. С а р в Л., Одно семейство нелинейных итерационных методов для решения некорректных задач. Изв. АН Эст.ССР, Физ., мат., I982, 3I, № 3, 26I-269.
6. Т и х о н о в А.Н., А р с е н и н В.Я., Методы решения некорректных задач. Москва, I979.

Поступило
I6 III I982

TWO-STEP α-PROCESSES AND THEIR APPLICATION FOR SOLVING ILL-POSED PROBLEMS

L. Sarv

Summary

Let A be a bounded linear operator between two Hilbert spaces, with the range of A not necessarily closed. Operator equation (1) is examined, where instead of exact A and f their approximations A_η and f_δ are given. For solving (1) the family of iteration methods (2) is used. From (2) we get for $\alpha = 0$ the two-step method of steepest descent and for $\alpha = 1$ the two-step method of minimal deviations.

The regularization of ill-posed problems by iteration methods is usually realized by stopping the iterations on a certain step. In the present paper the theorems of convergence are proved and error estimations are deduced for apriori given stopping index and for stopping according to deviation.

Уч.зап.Тартуск.ун-та,1982,633,50-54

ИТЕРАЦИОННАЯ СХЕМА ГЕРОНА ОБРАЩЕНИЯ ПЛОХО ОБУСЛОВЛЕННЫХ МАТРИЦ

А. Лилла

Рассмотрим систему линейных алгебраических уравнений

$$Au=f, \; f\in R(A), \tag{1}$$

с вырожденной самосопряженной неотрицательной квадратичной матрицей. Система (1) прямыми методами не разрешима. В работе предлагается алгоритм конструирования последовательности матриц B_n, таких что $\|B_n\|\leqslant 2^{2n}$ и $\|AB_nf-f\|\to 0$ при $n\to\infty$.

Алгоритм имеет следующий вид:

$$\begin{aligned} &\text{I} \quad Y_0=I \\ &\text{II} \quad Y_{k+1}=(Y_k+AY_k^{-1})/2 \\ &\text{III} \quad B_k=Y_k^{-1}Y_k^{-1} \qquad k=1,2,\ldots \end{aligned} \tag{2}$$

Для удобства дальнейших исследований предположим что, $\|A\|\leqslant 1$, хотя все полученные результаты распространяются на случай матрицы A произвольной нормы.

Докажем следующие свойства матриц:

1° матрица Y_k симметрична;

2° $Y_k>0$ и она имеет с A одинаковые собственные векторы;

3° $\|Y_k^{-1}\|\leqslant 2^k$ и $Y_k\to\sqrt{A}$ при $k\to\infty$.

Представим матрицу A в виде произведения

$$A=C\Lambda C^*,$$

где C унитарная матрица, а $\Lambda=\mathrm{diag}(a_1,a_2,\ \ \ ,a_m)$ - диагональная матрица, a_i - собственные значения матрицы A.

Свойства 1° и 2° докажем методом полной индукции. При $k=1$ имеем

$$Y_1=(I+AI)/2=(CC^*+C\Lambda C^*CC^*)/2=C[(I+\Lambda)/2]C^*.$$

Обозначим диагональную матрицу $(I+\Lambda)/2=f_1(\Lambda)$. Она положительная, как сумма положительной и неотрицательной матриц. Матрица Y_1 обладает свойствами 1° и 2° Предположим, что матрица $Y_K = C f_K(\Lambda) C^*$ удовлетворяет свойства I° и 2° и покажем, что тогда Y_{K+1} также удовлетворяет I° и 2°. Действительно,

$$Y_{K+1} = (Y_K + A Y_K^{-1})/2 = (C f_K(\Lambda) C^* + C \Lambda C^* C f_K^{-1}(\Lambda) C^*)/2 =$$
$$= C\left[(f_K(\Lambda) + \Lambda f_K^{-1}(\Lambda))/2\right] C^*.$$

Матрица Y_{K+1} симметрична, имеет с A одинаковые собственные векторы, положительная, как сумма положительной и неотрицательной матриц, $f_{K+1}(\Lambda) = (f_K(\Lambda) + \Lambda f_K^{-1}(\Lambda))/2$. Свойства I° и 2° доказаны.

Свойство 3° эквивалентно требованию, что

$$f_K^{-1}(\Lambda) \leq 2^K \quad \text{и} \quad f_K(\Lambda) \to \sqrt{\Lambda} \quad \text{при} \quad K \to \infty.$$

Заметим, что вышеописанный алгоритм формально совпадает с методом Ньютона для решения уравнения $y^2 - a = 0$; он известен также как метод Герона для вычисления $\sqrt{a}$. Так как по этому алгоритму вычисляют диагональные элементы матриц $f_K(\Lambda)$ то они при $K \to \infty$ сходятся к $\sqrt{\Lambda}$ и следовательно $Y_K \to \sqrt{A}$. Формула Герона преобразует отрезок $[0,1]$ в $(0,1]$ непрерывно и монотонно и на K-ом шаге $\sqrt{0}$ преобразуется в 2^{-K}. Следовательно, $f_K(\Lambda) \geq 2^{-K}$ или $f_K^{-1}(\Lambda) \leq 2^K$, чем свойство 3° доказано.

Исследуем скорость сходимости $f_K(\Lambda) \to \sqrt{\Lambda}$. Метод Ньютона для решения уравнения $g(x) = 0$ сходится (см. [2]), если $h = K\eta \leq 1/2$ где $\eta = |g/g'|$, $K = |g''/g'|$. В данном случае $g = y^2 - a$, $g' = 2y$, $g'' = 2$ и $h = |(y^2 - a)/2y^2| \leq 1/2$ при всех $a \in [0,1]$. При этом сходимость совершается со скоростью

$$|Y_K - \sqrt{a}| \leq 2^{-K}(2h)^{2^K}.$$

Диагональным элементам a_i матрицы Λ соответствуют величины $q_i = 2h - 2h_i$. Тогда сходимость $f_K(\Lambda) \to \sqrt{\Lambda}$ характеризуется диагональной матрицей скоростей Q_K с элементами $q_i^{2^K}$ и именно $|f_K(\Lambda) - \sqrt{\Lambda}| \leq 2^{-K} Q_K$.

Исследуем сходимость $AB_K f - f \to 0$. Имеем

$$AB_K f - f = (AB_K A - A)u = C(\Lambda f_K^{-2}(\Lambda)\Lambda - \Lambda)C^* u.$$

Остается проверить сходимость $\|\Lambda f_K^{-2}(\Lambda)\Lambda - \Lambda\| \to 0$. Обозначим диагональные элементы $f_K^{-2}(\Lambda)$ через $1/\Lambda_{K,i}^2$ и проверим сходимость по-компонентно:

$$\left|\frac{a_i^2}{\lambda_{k,i}^2}-a_i\right|=\left|\frac{(\sqrt{a_i}-\lambda_{k,i})(\sqrt{a_i}+\lambda_{k,i})}{\lambda_{k,i}^2}\right|a_i\leqslant 2a_i\left|\frac{\sqrt{a_i}-\lambda_{k,i}}{\lambda_{k,i}}\right|\leqslant$$
$$\leqslant\sqrt{a_i}\,2^{-k+1}q_i^{2^k}.$$

При этом $|\Lambda f_k^{-2}(\Lambda)\Lambda-\Lambda|\leqslant 2^{-k+1}Q_k\sqrt{\Lambda}$ или $\|AB_k f-f\|\leqslant$ $\leqslant \gamma_k 2^{-k+1}\|u\|$, где $\gamma_k=\max\limits_{0\leqslant a\leqslant 1}\sqrt{a}\,(1-a)^{2^k}<1$.

Приведем результаты численных экспериментов, в которых схема (2) сравнивается со схемой (см. [4])

$$\text{I}\quad B_0=I$$
$$\text{II}\quad B_{k+1}=B_k+B_k(I-AB_k)\quad k=1,2,\ldots \tag{3}$$

Одновременно исследована устойчивость решения относительно малых возмущений правой части уравнения (I). Приведем результаты двух экспериментов. В первом случае матрица A получена при дискретизации методом прямоугольников по 30 точкам ядра интегрального уравнения

$$\int_0^1 K(s,t)u(s)ds=f(t),$$

где

$$K(s,t)=\begin{cases} s(1-t) & 0\leqslant s\leqslant t\leqslant 1,\\ t(1-s) & 0\leqslant t\leqslant s\leqslant 1.\end{cases}$$

Во втором случае методом средних прямоугольников по 29 точкам дискретизировано ядро интегрального уравнения

$$\int_0^1\left\{\frac{1}{2}(s+t)+st+\frac{1}{3}\right\}u(s)ds=t+\frac{7}{12}.$$

Эти примеры взяты соответственно из [3] и [I]. Правые части уравнений (I) f сконструированы по решению $u(t)\equiv 1$, $f:=Au$ и умышленно испорчены добавками $\pm\delta$, $\delta=0.001$. В следующих таблицах приведены нормы погрешности $\|u-B_k f_\delta\|$ и невязки $\|f_\delta-AB_k f_\delta\|$; при этом использована норма $\|u\|=$ $=\left(\frac{1}{m}\sum\limits_{i=1}^{m}|u_i|^2\right)^{1/2}$. Из результатов видно что, при равной устойчивости схема (2) сходится два раза быстрее, чем схема (3). Так как они требуют при вычислении примерно равное количество операций, то схема (2) более экономная (метод квадратного корня обращения положительной матрицы требует $\approx n^3/2$ операций). В этих примерах при схеме (2) наиболее точный результат получается при остановке на невязке уровня I.5δ, а при схеме (3) – при остановке на невязке уровня I.2δ.

Пример первый

Шаг итерации	Схема (2)		Схема (3)	
	Погрешность	Невязка	Погрешность	Невязка
I	0.304	0.02230	0.343	0.02727
2	0.I98	0.00862	0.30I	0.02I79
3	0.I25	0.00I77	0.245	0.0I4I4
4	0.I78	0.00055	0.I89	0.00663
5	0.309	0.000I2	0.I45	0.00275
6	0.35I	0.0	0.II9	0.00I33
7	0.353	0.0	0.I30	0.00077
8	0.353	0.00002	0.I89	0.00048
9	0.353	0.000I7	0.27I	0.00023
I0	0.353	0.00I5I	0.333	0.00006

Пример второй

Шаг итерации	Схема (2)		Схема (3)	
	Погрешность	Невязка	Погрешность	Невязка
I	0.248	0.00634	0.252	0.02848
2	0.232	0.00I69	0.248	0.00I32
3	0.I88	0.00I44	0.242	0.00I74
4	0.262	0.00I07	0.23I	0.00I69
5	I.006	0.00098	0.2I2	0.00I59
6	4.02	0.00098	0.I84	0.00I42
7	I6.I	0.00I2I	0.I73	0.00I2I
8	64	0.I90	0.257	0.00I04
9	272	I04	0.503	0.00099
I0	8694I	$I0^6$	I.006	0.00098

Литература

I. Б а к у ш и н с к и й А.Б., Об одном численном методе решения интегральных уравнений Фредгольма I рода. Ж. вычисл. мат. и мат. физ., 1965, 5, № 4, 744-749.

2. К а н т о р о в и ч А.В., А к и л о в Г.П., Функциональный анализ. Москва, 1977.

3. К р ы л о в В.И., Б о б к о в В.В., М о н а с т ы р н ы й П.И., Вычислительные методы высшей математики. Минск, 1975.

4. У и л к и н с о н Дж. Х., Р а й н и ш К., Справочник алгоритмов на языке АЛГОЛ. Линейная алгебра. Москва, 1976.

Поступило
2 IУ 1982

HERON'S ITERATION METHOD FOR THE INVERSION OF ILL-CONDITIONED MATRIX

A. Lilla

Summary

This article deals with the inversion of the matrix by method (2). This method is compared with method (3).Comparative calculation experiments are carried out.

Уч.зап.Тартуск.ун-та,1982,633,55-66

О СХОДИМОСТИ РАЗНОСТНОГО МЕТОДА В СИЛЬНОЙ НОРМЕ ДЛЯ НЕЛИНЕЙНОЙ ЗАДАЧИ ЭЛЛИПТИЧЕСКОГО ТИПА

М. Фишер

Введение

Для дифференциального уравнения второго порядка со слабыми нелинейностями в коэффициентах изучаются вопросы сходимости разностного метода и метода итерации в метрике соболевского пространства H^2. Результаты основываются на неравенствах коэрцитивности.

Для линейного эллиптического оператора второго порядка под неравенством коэрцитивности понимают неравенства вида

$$\|Au\|_0 \geqslant \gamma \|u\|_2 - c\|u\|_1 ,$$

где $\|u\|_k$ означает норму в $H^k = W^{k,2}$. Для задачи Дирихле такое неравенство хорошо известно (см. [8]). Для соответствующей разностной задачи неравенства коэрцитивности установлены в [6,11]. В [7,10] приведены неравенства коэрцитивности при некоторых сильных ограничениях для общих нелинейных разностных эллиптических операторов.

Для общих нелинейных эллиптических задач порядка $2n$ сходимость разностного метода в метрике дискретного соболевского пространства H^n, изучалось многими авторами (см. [4,9,13]). Этими результатами будем в дальнейшем пользоваться.

§ I. Решаемая задача и разностная схема

Рассмотрим задачу Дирихле

$$Au \equiv -\sum_{i,j=1}^{s} \frac{\partial}{\partial x_i}\left(a_{ij}(x,u)\frac{\partial u}{\partial x_j}\right) = f, \quad u \in H_0^2, \quad f \in H^0 = L_2(\Omega), \qquad (I)$$

$s = 2$ или $s = 3$, $u = u(x)$, $a_{ij}(x,u) = a_{ji}(x,u)$,

где

$$H_0^2 = \{u : u \in H^2(\Omega),\ u|_{\partial\Omega} = 0\},$$

$H^k(\Omega) = W^{k,2}(\Omega)$ - пространство Соболева,

$$\Omega = \{0 < x_i < 1,\ i = 1, \dots, s\},$$

$\partial\Omega$ - его граница, $\bar{\Omega}$ - его замыкание.

Предполагается, что рассматриваемый оператор эллиптический, т.е. для каждого $a>0$ существует такое число $\tau_a>0$, что при всех $t_i\in R$, $x\in\Omega$, $u\in[-a,a]$

$$\sum_{i,j=1}^{s} a_{ij}(x,u)t_i t_j \geqslant \tau_a \sum_{i=1}^{s} t_i^2 . \qquad (2)$$

Обозначим

$$\bar\Omega_h=\{x: x_i=k_i h,\ k_i=0,1,\dots,n;\ i=1,\dots,s\},\quad h=\frac{1}{n},$$

$$\Omega_h=\bar\Omega_h\cap\Omega,\quad \partial\Omega_h=\bar\Omega_h\cap\partial\Omega .$$

В качестве дискретного аналога задачи (I) рассмотрим разностную задачу

$$A_h u_h \equiv -\frac{1}{2}\sum_{i,j=1}^{s}\left[\partial_i(a_{ij}(x,u_h)\bar\partial_j u_h)+\bar\partial_i(a_{ij}(x,u_h)\partial_j u_h)\right]=f_h, \qquad (3)$$

$$u_h\in H_0^2(\Omega_h),\quad f_h\in L_2(\Omega_h),$$

где

$$H_0^2(\Omega_h)=\{u_h\in H^2(\bar\Omega_h),\ u_h|_{\partial\Omega_h}=0\},$$

$H^k(\bar\Omega_h)=W^{k,2}(\bar\Omega_h)$, $W^{k,p}(\bar\Omega_h)$-пространство Соболева сеточных функций $u_h:\bar\Omega_h\to R$. с нормой

$$\|u_h\|_{k,p}=\Big(\sum_{|\alpha|\le k} h^s\sum_{x\in\Omega_h^\alpha}|\partial^\alpha u_h(x)|^p\Big)^{1/p},\ 1\le p<\infty,\ k\geqslant 0,$$

где Ω_h^α такая максимальная подсетка сетки $\bar\Omega_h$, что разность $\partial^\alpha u_h=\partial_1^{\alpha_1}\dots\partial_s^{\alpha_s}u_h$ не использует значения u_h вне $\bar\Omega_h$. Через ∂_i и $\bar\partial_i$, как обычно, обозначаем соответственно правые и левые разностные соотношения в i-том направлении. Далее будем использовать сокращенное обозначение $\|u_h\|_k=\|u_h\|_{k,2}$.

§2. Теоремы вложения

При доказательстве неравенства коэрцитивности играют важную роль некоторые дискретные аналоги теорем вложения Соболева. Сформулируем результат, который позволяет их вывести.

Пусть $\Phi_h: E_h\to E_h$ линейный разностный оператор порядка $2r$, где E_h - пространство сеточных функций, заданных на Ω_h, D_h^α разностная аппроксимация оператора $D^\alpha=\frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1}\dots\partial x_s^{\alpha_s}}$, α - s-мерный мультииндекс,

$$\mathcal{I}_1(\beta,\varrho)=\Big\{\Big(\frac1p,\frac1q\Big):\frac1q>\frac1p-\frac{2r\beta-\varrho}{s},\ 0\le\frac1p,\frac1q\le 1\Big\},$$

$$\mathcal{I}_2(\beta,\varrho)=\Big\{\Big(\frac1p,\frac1q\Big):\frac1q=\frac1p-\frac{2r\beta-\varrho}{s},\ \frac{2r\beta-\varrho}{s}<\frac1p<1\Big\},$$

$$\bar{\mathcal{I}}_1(\beta,\varrho)=\Big\{\Big(\frac1p,\frac1q\Big):\frac1p\geqslant\frac1q,\Big(\frac1p,\frac1q\Big)\in\mathcal{I}_1(\beta,\varrho)\Big\}.$$

При этих предположениях имеет место (см. [5])

Теорема I. Пусть в комплексной плоскости существует сектор $S_\vartheta = \{\lambda : \vartheta < \arg\lambda < 2\pi - \vartheta\}$ $(0<\vartheta<\pi)$ такой, что при $\lambda \in S_\vartheta$ определен оператор $(\Phi_h - \lambda)^{-1}$ и справедлива оценка

$$\| D_h^\alpha (\Phi_h - \lambda)^{-1} \|_{L_{p,h} \to L_{q,h}} \le c(\vartheta)(1+|\lambda|)^{-1+\frac{|\alpha|}{2r} + \frac{s}{2r}\left(\frac{1}{p} - \frac{1}{q}\right)} \qquad (4)$$

при $\alpha = 0$ и при некотором α $(0<|\alpha|<2r)$ и при всех $\left(\frac{1}{p}, \frac{1}{q}\right) \in \bar{\mathcal{I}}_1(1, |\alpha|)$. Тогда оператор $D_h^\alpha \Phi_h^{-\beta}$ $\left(\frac{|\alpha|}{2r} < \beta < \min\left\{1, \frac{s+|\alpha|}{2r}\right\}\right)$ равномерно по h ограничен из $L_{p,h}$ в $L_{q,h}$, где $\left(\frac{1}{p}, \frac{1}{q}\right) \in \mathcal{I}_1(\beta, |\alpha|) \cap \mathcal{I}_2(\beta, |\alpha|)$.

Рассмотрим оператор

$$J_h u_h = -\sum_{i=1}^{s} \partial_i \bar{\partial}_i u_h, \quad u_h \in L_2(\Omega_h). \qquad (5)$$

Известно, что для сеточных функций, равных нулю на $\partial\Omega_h$ норма $\|u_h\|_2$ эквивалентна норме $\|J_h u_h\|_0$ (см. [6]), а норма $\|u_h\|_1$ эквивалентна норме $\|J_h^{1/2} u_h\|_0 = (J_h u_h, u_h)$. При помощи оператора J_h можно определить интерполяционные пространства $H^{2\beta}$ $\left(\frac{1}{2}<\beta<1\right)$ с нормой

$$\|u_h\|_{2\beta} := \|J_h^\beta u_h\|_0 .$$

В [5] доказано, что при всех $|\alpha| = 0, 1$ оператор (5) $(r=1)$ удовлетворяет оценке (4) при всех $\left(\frac{1}{p}, \frac{1}{q}\right) \in \bar{\mathcal{I}}_1(1, |\alpha|)$. Из теоремы I вытекают следующие нужные нам в дальнейшем дискретные аналоги теорем вложения Соболева:

$$\|u_h\|_C \le c\|u_h\|_{k'}, \quad k'>1 \text{ при } s=2, \quad k' > \frac{3}{2} \text{ при } s=3, \qquad (6)$$

$$\|u_h\|_{1,4} \le c\|u_h\|_{k'}, \quad k' \geq \frac{3}{2} \text{ при } s=2, \quad k' \geq \frac{7}{4} \text{ при } s=3. \qquad (7)$$

При сколь угодно малом $\varepsilon > 0$ имеет место также неравенство

$$\|u_h\|_{k'} \le c\left(\varepsilon^{2-k'}\|u_h\|_2 + \varepsilon^{1-k'}\|u_h\|_1\right), \quad 1<k'<2. \qquad (8)$$

§3. Неравенство коэрцитивности

Теорема 2. Пусть функции $a_{ij}(x,u) = a_{ji}(x,u)$ дифференцируемы и

$$|a_{ij}(x,u)|, \left|\frac{\partial a_{ij}(x,u)}{\partial x_l}\right|, \left|\frac{\partial a_{ij}(x,u)}{\partial u}\right|, \left|\frac{\partial^2 a_{ij}(x,u)}{\partial x_l \partial u}\right|, \left|\frac{\partial^2 a_{ij}(x,u)}{\partial u^2}\right| \le d_a, \qquad (9)$$

$l=1,\dots,s$, $\forall x \in \Omega_h$, $\forall u \in [-a, a]$, d_a – положительная постоянная, зависящая от a. Пусть выполнено условие эллиптичности (2). Тогда для любых $u_h, v_h \in H_0^2(\Omega_h)$, $\|u_h\|_2, \|v_h\|_2 \le a$ имеет место

$$(A_h u_h - A_h v_h, J_h(u_h - v_h)) \geq \frac{\tau_a}{2}\|u_h - v_h\|_2^2 - c_a\|u_h - v_h\|_1^2 , \qquad (10)$$

а вместе с тем

$$((A_h + c_a I_h)u_h - (A_h + c_a I_h)v_h, J_h(u_h - v_h)) \geq \frac{\tau_a}{2}\|u_h - v_h\|_2^2 , \qquad (11)$$

$$\|A_h u_h - A_h v_h\|_0 \geq \frac{\tau_a}{2}\|u_h - v_h\|_2 - c_a\|u_h - v_h\|_1 , \qquad (12)$$

где I_h - единичный оператор , $c_a = c \cdot d_a \cdot a^{\frac{4}{2-\kappa'}} \cdot \tau_a^{\frac{\kappa'}{\kappa'-2}}$, $c > 0$, $\kappa' \in [\frac{3}{2}, 2)$ при $s = 2$, $\kappa' \in [\frac{7}{4}, 2)$ при $s = 3$.

Доказательство. Рассмотрим скалярное произведение

$$(A_h u_h - A_h v_h, J_h z_h) = \frac{1}{2}(\varphi_h + \psi_h, J_h z_h) , \; z_h = u_h - v_h ,$$

где

$$\varphi_h = -\sum_{i,j=1}^{s}\left[\partial_i(a_{ij}(x,u_h)\bar\partial_j z_h) + \bar\partial_i(a_{ij}(x,u_h)\partial_j z_h)\right],$$

$$\psi_h = -\sum_{i,j=1}^{s}\left[\partial_i(a_{ij}(x,u_h) - a_{ij}(x,v_h))\bar\partial_j v_h) + \bar\partial_i(a_{ij}(x,u_h) - a_{ij}(x,v_h))\partial_j v_h)\right].$$

Нетрудно представить φ_h в форме

$$\varphi_h = \varphi_{h1} + \varphi_{h2} + \varphi_{h3} ,$$

где

$$\varphi_{h1} = -\sum_{i,j=1}^{s} a_{ij}(x,u_h)(\partial_i\bar\partial_j z_h + \bar\partial_i\partial_j z_h)$$

$$\varphi_{h2} = -\sum_{i,j=1}^{s}\Big[\frac{\partial a_{ij}(\xi^+, \theta u_h + (1-\theta)u_h^{+1i})}{\partial u}\partial_i u_h (\bar\partial_j z_h)^{+1i} + $$
$$+ \frac{\partial a_{ij}(\xi^-, \theta u_h^{-1i} + (1-\theta)u_h)}{\partial u}\bar\partial_i u_h (\partial_j z_h)^{-1i}\Big],$$

$$\varphi_{h3} = -\sum_{i,j=1}^{s}\Big[\frac{\partial a_{ij}(\xi^+, \theta u_h + (1-\theta)u_h^{+1i})}{\partial x_i}(\bar\partial_j z_h)^{+1i} + $$
$$+ \frac{\partial a_{ij}(\xi^-, \theta u_h^{-1i} + (1-\theta)u_h)}{\partial x_i}(\partial_j z_h)^{-1i}\Big], \; 0 < \theta < 1 ,$$

$$\xi^+ = \theta x + (1-\theta)(x + he_i) , \; \xi^- = \theta(x - he_i) + (1-\theta)x ,$$
$$(u_h)^{+1i} = u_h(x + he_i), \; (u_h)^{-1i} = u_h(x - he_i), \; e_i = (\delta_{i1}, \ldots, \delta_{is}) .$$

Оценим скалярное произведение $(\varphi_{h1}, J_h z_h)$, используя формулы суммирования по частям и условие эллиптичности (2). Мы получаем

$$(\varphi_{h1}, J_h z_h) \geq 2\tau_a\|z_h\|_2^2 - c d_a\|z_h\|_1\|z_h\|_2 - c d_a\|u_h\|_{1,4}\|z_h\|_{1,4}\|z_h\|_2 .$$

Непосредственно находим, что

$$|(\varphi_{h2}, J_h z_h)| \leq c d_a\|u_h\|_{1,4}\|z_h\|_{1,4}\|z_h\|_2 ,$$
$$|(\varphi_{h3}, J_h z_h)| \leq c d_a\|z_h\|_1\|z_h\|_2 .$$

Итак

$$(\varphi_h, J_h z_h) \geq 2\tau_a\|z_h\|_2^2 - c d_a\|z_h\|_1\|z_h\|_2 - c d_a\|u_h\|_{1,4}\|z_h\|_{1,4}\|z_h\|_2 .$$

Для оценки скалярного произведения $(\psi_h, J_h z_h)$ запишем ψ_h в виде

$$\psi_h = \psi_{h1} + \psi_{h2} + \psi_{h3},$$

где

$$\psi_{h1} = \sum_{i,j=1}^{s} [(a_{ij}(x,u_h) - a_{ij}(x,v_h))\partial_i\bar{\partial}_j v_h + (a_{ij}(x,u_h) - a_{ij}(x,v_h))\bar{\partial}_i\partial_j v_h],$$

$$\psi_{h2} = \sum_{i,j=1}^{s} \Big[\frac{\partial a_{ij}(\xi^+, \theta u_h + (1-\theta)u_h^{+1_i})}{\partial u}\partial_i u_h - \frac{\partial a_{ij}(\xi^+, \theta v_h + (1-\theta)v_h^{+1_i})}{\partial v}\partial_i v_h\Big](\bar{\partial}_j v_h)^{+1_i} +$$

$$+ \sum_{i,j=1}^{s} \Big[\frac{\partial a_{ij}(\xi^-, \theta u_h^{-1_i} + (1-\theta)u_h)}{\partial u}\bar{\partial}_i u_h - \frac{\partial a_{ij}(\xi^-, \theta v_h^{-1_i} + (1-\theta)v_h)}{\partial v}\bar{\partial}_i v_h\Big](\partial_j v_h)^{-1_i},$$

$$\psi_{h3} = \sum_{i,j=1}^{s} \Big[\frac{\partial a_{ij}(\xi^+, \theta u_h + (1-\theta)u_h^{+1_i})}{\partial x_i} - \frac{\partial a_{ij}(\xi^+, \theta v_h + (1-\theta)v_h^{+1_i})}{\partial x_i}\Big](\bar{\partial}_j v_h)^{+1_i} +$$

$$+ \sum_{i,j=1}^{s} \Big[\frac{\partial a_{ij}(\xi^-, \theta u_h^{-1_i} + (1-\theta)u_h)}{\partial x_i} - \frac{\partial a_{ij}(\xi^-, \theta v_h^{-1_i} + (1-\theta)v_h)}{\partial x_i}\Big](\partial_j v_h)^{-1_i}.$$

При помощи равенства

$$a_{ij}(x,u_h) - a_{ij}(x,v_h) = \frac{\partial a_{ij}(x, \theta u_h + (1-\theta)v_h)}{\partial u} z_h, \quad 0<\theta<1$$

получим

$$|(\psi_{h1}, J_h z_h)| \le c d_a \|z_h\|_C \|v_h\|_2 \|z_h\|_2 .$$

Преобразуем ψ_{h2}:

$$\psi_{h2} = -\sum_{i,j=1}^{s} \Big[\frac{\partial a_{ij}(\xi^+, \theta v_h + (1-\theta)v_h^{+1_i})}{\partial v}\partial_i z_h (\bar{\partial}_j v_h)^{+1_i} -$$

$$- \Big(\frac{\partial a_{ij}(\xi^+, \theta u_h + (1-\theta)u_h^{+1_i})}{\partial u} - \frac{\partial a_{ij}(\xi^+, \theta v_h + (1-\theta)v_h^{+1_i})}{\partial v}\Big)\partial_i u_h (\bar{\partial}_j v_h)^{+1_i}\Big] -$$

$$- \sum_{i,j=1}^{s} \Big[\frac{\partial a_{ij}(\xi^-, \theta v_h^{-1_i} + (1-\theta)v_h)}{\partial v}\bar{\partial}_i z_h (\partial_j v_h)^{-1_i} -$$

$$- \Big(\frac{\partial a_{ij}(\xi^-, \theta u_h^{-1_i} + (1-\theta)u_h)}{\partial u} - \frac{\partial a_{ij}(\xi^-, \theta v_h^{-1_i} + (1-\theta)v_h)}{\partial v}\Big)\bar{\partial}_i u_h (\partial_j v_h)^{-1_i}\Big].$$

Так как

$$\frac{\partial a_{ij}(x,u)}{\partial u} - \frac{\partial a_{ij}(x,v)}{\partial v} = \frac{\partial^2 a_{ij}(x, \theta u_h + (1-\theta)v_h)}{\partial u^2}(u_h - v_h), \quad 0<\theta<1,$$

то

$$|(\psi_{h2}, J_h z_h)| \le c d_a (\|z_h\|_{1,4} \|v_h\|_{1,4} \|z_h\|_2 + \|z_h\|_C \|u_h\|_{1,4} \|v_h\|_{1,4} \|z_h\|_2).$$

Аналогичным образом находим, что

$$|(\psi_{h3}, J_h z_h)| \le c d_a \|z_h\|_C \|v_h\|_1 \|z_h\|_2 .$$

Следовательно,

$$|(\psi_h, J_h z_h)| \le c d_a \|z_h\|_2 (\|z_h\|_C \|v_h\|_2 + \|z_h\|_{1,4} \|v_h\|_{1,4} + \|z_h\|_C \|u_h\|_{1,4} \|v_h\|_{1,4}),$$

и из формул (6), (7) получим

$$(A_h u_h - A_h v_h, J_h z_h) \geq \tau_a \|z_h\|_2^2 - c d_a \|z_h\|_1 \|z_h\|_2 -$$

$$- c d_a \|z_h\|_{k'} \|z_h\|_2 (\|u_h\|_2 + \|v_h\|_2 + \|u_h\|_2 \|v_h\|_2).$$

Учитывая, что $\|u_h\|_2, \|v_h\|_2 \leq a$ при помощи (8) установим

$$(A_h u_h - A_h v_h, J_h z_h) \geq \tau_a \|z_h\|_2^2 - c d_a \|z_h\|_2 \|z_h\|_1 - c d_a a^2 \|z_h\|_2 (\varepsilon^{2-k'} \|z_h\|_2 + \varepsilon^{1-k'} \|z_h\|_1) \geq$$

$$\geq \tau_a \|z_h\|_2^2 - c d_a a^2 \varepsilon^{2-k'} \|z_h\|_2^2 - c d_a a^2 \varepsilon^{1-k'} \|z_h\|_2 \|z_h\|_1.$$

Используя ε-неравенство, получим из последнего

$$(A_h u_h - A_h v_h, J_h z_h) \geq (\tau_a - c d_a a^2 \varepsilon^{2-k'}) \|z_h\|_2^2 - c d_a a^2 \varepsilon^{-k'} \|z_h\|_1^2.$$

Выбирая $\varepsilon > 0$ так, что $c d_a a^2 \varepsilon^{2-k'} \leq \frac{\tau_a}{2}$ получим требуемую оценку (I0). Теорема доказана.

§ 4. Сходимость разностного метода

При исследовании сходимости разностного метода пользуемся методикой, приведенной в [2].

Введем оператор $p_h \in \mathcal{L}(H^k(\Omega), H^k(\Omega_h))$, положив

$$(p_h u)(x) = h^{-s} \int_{\omega_h(x)} u(y)\, dy \quad , \; x \in \Omega_h,$$

где $\omega_h(x)$ элементарная ячейка объема h^s с центром в точке x:

$$\omega_h(x) = \{ y = (y_1, \dots, y_s) \in \Omega, \, x_j - \tfrac{h}{2} < y_j \leq x_j + \tfrac{h}{2}, \, j = 1, \dots, s \}.$$

Эти операторы являются для пар $H^0(\Omega)$ и $H^0(\Omega_h)$, $H^1(\Omega)$ и $H^1(\Omega_h)$, $H^2(\Omega)$ и $H^2(\Omega_h)$ связывающими, т.е. удовлетворяют условиям (см. [I2])

$$\|p_h u\|_0 \to \|u\|_0, \; \forall u \in H^0(\Omega), \; \|p_h u\|_1 \to \|u\|_1, \; \forall u \in H^1(\Omega),$$

$$\|p_h u\|_2 \to \|u\|_2, \; \forall u \in H^2(\Omega) \quad , \; n \in N.$$

Через N обозначено здесь множество натуральных чисел.

Известно (см. [7]), что если коэффициенты уравнения (I) $a_{ij}(x,u) \cdot \Omega \times R \to R$ непрерывные, вещественные функции своих аргументов, удовлетворяющие условию

$$|a_{ij}(x,u)| \leq c \tag{I3}$$

и условию монотонности

$$\sum_{i,j=1}^{s} (a_{ij}(x,p_0) p_j - a_{ij}(x,q_0) q_j)(p_i - q_i) \geq m \sum_{i=1}^{s} (p_i - q_i)^2, \quad m > 0, \tag{I4}$$

$$\forall x \in \Omega, \; \forall p_i, q_i \in R, \quad i = 0, 1, \dots, s$$

и если

$$\|f_h - p_h f\|_0 \to 0 \; , \; f \in H^0(\Omega), \; f_h \in H^0(\Omega_h), \tag{I5}$$

то как уравнение (I) так и уравнение (3) имеют единственное решение соответственно $u^* \in H_0^1(\Omega)$ и $u_h^* \in H_0^1(\Omega_h)$, причем

$$\| u_h^* - p_h u^* \|_1 \to 0$$

Теорема 3. Пусть выполнены условия теоремы 2 и условия (I3), (I4), (I5). Если решение задачи (I) $u^* \in H_0^2(\Omega)$, то имеет место сходимость

$$\| u_h^* - p_h u^* \|_2 \to 0 \tag{I6}$$

Доказательство. Напишем неравенство коэрцитивности(I2) для $u_h = u_h^*$, $v_h = p_h u_h^*$

$$\| A_h u_h^* - A_h p_h u^* \|_0 \geq \frac{\tau_a}{\varkappa} \| u_h^* - p_h u^* \|_2 - c_a \| u_h^* - p_h u^* \|_1 .$$

Так как известно, что $\| u_h^* - p_h u^* \|_1 \to 0$, то для доказательства утверждения (I6) достаточно показать сходимость $\| A_h u_h^* - A_h p_h u^* \|_0 \to 0$. Но

$$\| A_h u_h^* - A_h p_h u^* \|_0 \leq \| f_h - p_h f \|_0 + \| A_h p_h u^* - p_h A u^* \|_0$$

и остается показать, что $\| A_h p_h u^* - p_h A u^* \|_0 \to 0$.

Убедимся сперва, что на плотном в $H_0^2(\Omega)$ множестве $H_0^2(\Omega) \cap C^\infty(\Omega)$ имеет место сходимость

$$\| \partial^\alpha p_h u - p_h D^\alpha u \|_0 \to 0 , \forall \alpha , |\alpha| \leq 2 . \tag{I7}$$

Действительно, для $u \in C^\infty(\Omega)$

$$\| \partial^\alpha p_h u - p_h D^\alpha u \|_0^2 = \| h^{-s} \int_{\omega_h(x)} \partial^\alpha u(y) dy - h^{-s} \int_{\omega_h(x)} D^\alpha u(y) dy \|_0^2 =$$

$$= h^{-2s} h^s \sum_{x \in \Omega_h^*} \left| \int_{\omega_h(x)} [\partial^\alpha u(y) - D^\alpha u(y)] dy \right|^2 \leq$$

$$\leq \sum_{x \in \Omega_h^*} \int_{\omega_h(x)} | \partial^\alpha u(y) - D^\alpha u(y) |^2 dy$$

Но $\partial^\alpha u(y) - D^\alpha u(y) = O(h)$ и следовательно имеет место (I7). Вместе с тем $\| A_h p_h u - p_h A u \|_0 \to 0$ для $u \in C^\infty(\Omega)$. Принимая во внимание равномерную по h ограниченность оператора A_h : $H_0^2(\Omega_h) \to H^0(\Omega_h)$ получаем сходимость и для всякого $u \in H_0^2(\Omega)$.

§5. О гладкости решения уравнения (I)

Для исследования разрешимости уравнения (I) в норме $H_0^2(\Omega)$ используем лемму из [II].

Лемма. Пусть оператор $\mathcal{A} : E \to F$ (E, F - банаховы пространства) дифференцируем по Фреше в шаре $\| u - u^0 \| \leq \delta_0$ и пусть оператор $\mathcal{A}'(u^0) \in \mathcal{L}(E,F)$ имеет обратный $[\mathcal{A}'(u^0)]^{-1} \in \mathcal{L}(F,E)$, $\| \mathcal{A}'(u^0) \| \leq \tau$, $\| [\mathcal{A}'(u^0)]^{-1} \| \leq \varkappa$.

Пусть с некоторым $q\ (0\le q<1)$ выполнены неравенства

$$\sup_{\|u-u^0\|\le\delta_0}\|\mathcal{A}'(u)-\mathcal{A}'(u^0)\|\le\frac{q}{\varkappa},\quad \|\mathcal{A}u^0-f\|\le\frac{\delta_0(1-q)}{\varkappa}.$$

Тогда уравнение $\mathcal{A}u=f$ имеет в шаре $\|u-u^0\|\le\delta_0$ единственное решение u^*, причем

$$\frac{\alpha}{1+q}\le\|u^*-u^0\|\le\frac{\alpha}{1-q},$$

где

$$\alpha=\|[\mathcal{A}'(u^0)]^{-1}(\mathcal{A}u^0-f)\|,\quad \frac{\|\mathcal{A}u^0-f\|}{\tau}\le\alpha\le\|\mathcal{A}u^0-f\|.$$

Пусть выполнены условия теоремы 2, т.е. (2) и (9). Из условия (9) следует дифференцируемость по Фреше оператора $A:H_0^2(\Omega)\to H^0(\Omega)$ в шаре $\|u\|_2\le a$, причем

$$A'(u)v=-\sum_{i,j=1}^{s}\frac{\partial}{\partial x_i}\left(a_{ij}(x,u(x))\frac{\partial v}{\partial x_j}\right)-\sum_{i,j=1}^{s}\frac{\partial}{\partial x_i}\left(\frac{\partial a_{ij}(x,u(x))}{\partial u}\frac{\partial u}{\partial x_j}v\right)$$

и

$$A'(0)v=-\sum_{i,j=1}^{s}\frac{\partial}{\partial x_i}\left(a_{ij}(x,0)\frac{\partial v}{\partial x_j}\right),\quad \forall v\in H_0^2(\Omega),\ \|v\|_2\le a.$$

Так как из условия (2) вытекает, что

$$(A'(0)v,v)\ge\tau_a\|v\|_1^2,\quad \forall v\in H_0^2(\Omega),\ \|v\|_2\le a,$$

то существует $[A'(0)]^{-1}\in\mathcal{L}(H^0,H_0^2)$.

Для каждого $v\in H^0$ имеем

$$\|A'(0)v\|_0=\left\|\sum_{i,j=1}^{s}\frac{\partial}{\partial x_i}\left(a_{ij}(x,0)\frac{\partial v}{\partial x_j}\right)\right\|_0=$$

$$=\left\|\sum_{i,j=1}^{s}\left[a_{ij}(x,0)\frac{\partial^2 v}{\partial x_i\partial x_j}+\frac{\partial a_{ij}(x,0)}{\partial x_i}\frac{\partial v}{\partial x_j}\right]\right\|_0\le 2d_a\|v\|_2,$$

поэтому

$$\|A'(0)\|_{\mathcal{L}(H_0^2,H^0)}\le 2d_a.$$

Аналогично, из

$$\|[A'(u)-A'(0)]v\|_0\le 6d_a\|u\|_2\|v\|_2$$

находим, что

$$\sup_{\|u\|_2\le a}\|A'(u)-A'(0)\|_{\mathcal{L}(H_0^2,H^0)}\le 6ad_a.$$

Пусть $\|[A'(0)]^{-1}\|_{\mathcal{L}(H^0,H_0^2)}\le\varkappa$ и пусть число a выбрано настолько малым, что при некотором q, $0<q<1$, $6ad_a\le\frac{q}{\varkappa}$. Тогда по лемме при таких f, для которых $\|f\|_0\le\frac{a(1-q)}{\varkappa}$, имеет уравнение (I) в шаре $\|u\|_2\le a$ единственное решение $u^*\in H_0^2(\Omega)$.

§ 6. Метод итераций

Для решения задачи (3) можно использовать итерационный процесс

$$u_h^{l+1} = u_h^l - tJ_h^{-1}(A_h u_h^l - f_h), \quad l = 0,1,\dots \tag{18}$$

который при любом $t \in (0, \frac{2m}{M^2})$ и при любом начальном приближений $u_h^0 \in H_0^1(\Omega_h)$ сходится к решению u_h^* задачи (3) в норме $H^1(\Omega_h)$ со скоростью геометрической прогрессии со знаменателем $q_1 = q_1(t) = (1 - 2mt + M^2t^2)^{1/2} < 1$, если только оператор $A_h: H_0^1(\Omega_h) \to H^{-1}(\Omega_h)$ удовлетворяет условиям (см. [3]):

1^0 $\langle A_h u_h - A_h v_h, u_h - v_h \rangle \geq m \|u_h - v_h\|_1^2$,

2^0 $\|A_h u_h - A_h v_h\|_{-1} \leq M \|u_h - v_h\|_1, \forall u_h, v_h \in H_0^1(\Omega_h)$.

То, что 1^0, 2^0 в самом деле имеет место, следует из условия монотонности (14) и из условия

$$|a_{ij}(x,p_0)p_j - a_{ij}(x,q_0)q_j|^2 \leq M \sum_{i=0}^{2} (p_i - q_i)^2, \quad p_i, q_i \in R, \; x \in \Omega, \tag{19}$$

В [3] доказано, что итерационный метод (18) не выводит из шара $K = \{u_h : u_h \in H_0^1(\Omega_h) : \|u_h\|_1 \leq \frac{3}{m}\|f_h\|_{-1}\}$. Следовательно, достаточно требовать выполнения условий 1^0, 2^0 лишь для $u_h, v_h \in K$.

Для исследования сходимости итерационного процесса (18) в метрике $H^2(\Omega_h)$ перепишем (18) в виде

$$u_h^{l+1} = B_h u_h^l + tJ_h^{-1}(f_h + c_a u_h^l), \quad l = 0,1,\dots, \tag{18'}$$

где

$$B_h u_h^l = u_h^l - tJ_h^{-1}(A_h + c_a I_h) u_h^l .$$

Так как оператор $A_h : H_0^2(\Omega_h) \to H^0(\Omega_h)$ при выполнении условия (9) дифференцируем по Фреше в шаре $\|u_h\|_2 \leq a$, то

$$\|A_h u_h - A_h v_h\|_0 \leq M_a \|u_h - v_h\|_2, \tag{20}$$

$$\forall u_h, v_h \in H_0^2(\Omega_h), \; \|u_h\|_2, \|v_h\|_2 \leq a,$$

где

$$M_a = \sup_{\|u_h\|_2 \leq a} \|A_h'(u_h)\|_{\mathcal{L}(H_0^2, H^0)} = (2+3a)d_a .$$

Теперь, из неравенства коэрцитивности (11) с $\tau_a > m$ и из (20) вытекает, что оператор B_h является сжимающим, т.е.

$$\|B_h u_h - B_h v_h\|_2 \leq q \|u_h - v_h\|_2, \; \forall u_h, v_h \in H_0^2(\Omega_h), \|u_h\|_2, \|v_h\|_2 \leq a \tag{21}$$

где $q = q(t) = [1 - mt + t^2(M_a + c_a)^2]^{1/2} < 1$ при $t \in (0, \frac{m}{(M_a+c_a)^2})$.

Рассмотрим шар $S_a = \{u_h \in H_0^2(\Omega_h) \cdot \|u_h\|_2 \leq a = \frac{3}{m}\|f_h\|_0\}$. Оценим разность двух последовательных приближений в метрике $H^2(\Omega_h)$, выбирая за начальное приближение $\overset{\circ}{u}_h = 0$. Тогда $\|u_h^1\|_2 = t\|f_h\|_0$ и

$$\|u_h^{l+1} - u_h^l\|_2 \leq q\|u_h^l - u_h^{l-1}\|_2 + tc_a\|u_h^l - u_h^{l-1}\|_0 \leq$$

$$\leq q^l\|u_h^1 - \overset{\circ}{u}_h\|_2 + t^2 c_a\|f_h\|_0 \sum_{j=0}^{l-1} q^j q_1^{l-j-1} \leq \tag{22}$$

$$\leq \|f_h\|_0 (tq^l + t^2 c_a \sum_{j=0}^{l-1} q^j q_1^{l-j-1}) =$$

$$= \|f_h\|_0 (tq^l + t^2 c_a \frac{q^l - q_1^l}{q - q_1}), \quad l = 1, 2, \ldots,$$

$$t \in (0, b), \quad b = \min\left(\frac{m}{(M_a+c_a)^2}, \frac{2m}{M^2}\right).$$

Здесь мы использовали неравенство (21) и сходимость метода (18) в метрике $H^1(\Omega_h)$.

Выясним, при каких условиях любое приближение, найденное по (18) остается в шаре S_a. Для этого оценим

$$\|\overset{\circ}{u}_h\|_2 + \|u_h^1 - \overset{\circ}{u}_h\| + \sum_{l=1}^{\infty}\|u_h^{l+1} - u_h^l\|_2 \leq t\|f_h\|_0 (1 + \sum_{l=1}^{\infty} q^l + tc_a \sum_{l=1}^{\infty} \frac{q^l - q_1^l}{q - q_1}) =$$

$$= t\|f_h\|_0 \left[1 + \frac{1}{1-q} + \frac{tc_a}{q-q_1}\left(\frac{1}{1-q} - \frac{1}{1-q_1}\right)\right] =$$

$$= \|f_h\|_0 \left[t\frac{2-q}{1-q} + t^2 c_a \frac{1}{(1-q)(1-q_1)}\right].$$

Имеем

$$q(t) = 1 - \frac{m}{2}t + \frac{t^2}{2}\left[(M_a + c_a)^2 - \frac{1}{4}m^2\right] + O(t^3),$$

$$q_1(t) = 1 - mt + \frac{t^2}{2}(M^2 - m^2) + O(t^3).$$

При достаточно малых t найдется такое $\varepsilon > 0$, что

$$q(t) = 1 - \frac{m}{2}t(1-\varepsilon), \quad q_1(t) \leq 1 - mt(1-\varepsilon).$$

В результате получим

$$\frac{t}{1-q}(2-q) + c_a\frac{t^2}{(1-q)(1-q_1)} \leq \frac{1}{1-\varepsilon}\left(\frac{2}{m} + t\right) + \frac{2c_a}{m^2(1-\varepsilon)^2}.$$

Для того, чтобы любое приближение u_h^l осталось в шаре S_a надо потребовать, чтобы

$$\frac{1}{1-\varepsilon}\left(\frac{2}{m} + t\right) + \frac{2c_a}{m^2(1-\varepsilon)^2} \leq \frac{3}{m} \quad \text{или} \quad t \leq \frac{1}{m}\left(1 - 3\varepsilon - \frac{2c_a}{m(1-\varepsilon)}\right) = b'.$$

Но так как t может принимать только положительные значения, то должно выполняться

$$\frac{2c_a}{m} < (1-3\varepsilon)(1-\varepsilon)$$

или

$$\frac{2c_a}{m} < 1 - 4\varepsilon.$$

Это, в свою очередь, влечет за собой ограничения на свободный член f_h. Именно, при $a=\frac{3}{m}\|f_h\|_0$ должно быть

$$\frac{2ca}{m}=2c\,(3\|f_h\|_0)^{\frac{4}{2-k'}}\,m^{\frac{6}{k+2}}\le 1-4\varepsilon. \qquad (23)$$

Значит, при достаточно малых по норме f_h любое приближение u_h^l остается в шаре S_a.

Из неравенства (22) следует, что

$$\|u_h^{l+p}-u_h^l\|_2\le\sum_{i=l}^{l+p-1}\|u_h^{i+1}-u_h^i\|_2\le$$

$$\le t\|f_h\|_0\Big[q^l\sum_{i=0}^{p-1}q^i+\frac{tca}{q-q_1}\Big(q^l\sum_{i=0}^{p-1}q^i-q_1^l\sum_{i=0}^{p-1}q_1^i\Big)\Big]=$$

$$=t\|f_h\|_0\Big[q^l\frac{1-q^p}{1-q}+\frac{tca}{q-q_1}\Big(q^l\frac{1-q^p}{1-q}-q_1^l\frac{1-q_1^p}{1-q_1}\Big)\Big].$$

Так как $q,q_1<1$, то

$$\|u_h^{l+p}-u_h^l\|_2\le t\|f_h\|_0\Big[q^l\frac{1}{1-q}+\frac{tca}{|q-q_1|}\Big(q^l\frac{1}{1-q}+q_1^l\frac{1}{1-q_1}\Big)\Big]$$

и последовательность $\{u_h^l\}$ является в $H_0^2(\Omega_h)$ фундаментальной.

Значит, если свободный член f_h таков, что выполняется (23), то задача (3) имеет в $S_a\subset H_0^2(\Omega_h)$ единственное решение u_h^*, и имеет место следующая оценка скорости сходимости метода итераций (I8):

$$\|u_h^*-u_h^l\|_2\le t\|f_h\|_0\Big[q^l\frac{1}{1-q}+\frac{tca}{|q-q_1|}\Big(q^l\frac{1}{1-q}+q_1^l\frac{1}{1-q_1}\Big)\Big],\ l=1,2,\dots,$$
$$t\in(0,\bar t),\ \bar t=\min(b,b').$$

Литература

I. Вайнберг М.М., Вариационный метод и метод монотонных операторов. Москва, I972.

2. Вайникко Г., Анализ дискретизационных методов. Тарту, I976.

3. Гаевский Х., Грегер К., Захариас К., Нелинейные операторные уравнения и операторные дифференциальные уравнения. Москва, I978.

4. Гудович Н.Н., О применении разностного метода к решению нелинейных эллиптических уравнений. Докл. АН СССР, I968, I79, № 6, I257-I266.

5. Гурова И.Н., Дробные степени разностных операторов и их приложения в теории разностных схем. Канд. диссертация. Воронеж, I980.

6. Дьяконов Е.Г., Разностные методы решения краевых задач. Вып. I. Москва, I97I.

7. Д ь я к о н о в Е.Г., О сходимость одного итерационного процесса. Успехи матем. наук, 1966, 21, № 1, 179-182.

8. Л а д ы ж е н с к а я О.А., У р а л ь ц е в а Н.Н., Линейные и кыазилинейные уравнения эллиптического типа. Москва, 1973.

9. Л а п и н А.В., Л я ш к о А.Д., Исследование метода сеток для нелинейных эллиптических уравнений любого порядка. Изв. вузов. Математика, 1974, № 10, 37-43.

10. Л а п и н А.В., Исследование разностных схем в норме $W_2^{(2)}$ для квазилинейных эллиптических уравнений. Ж. вычисл. мат. и мат. физ., 1974, 14, № 6, 1516-1525.

11. С о б о л е в с к и й П.Е., Т и у н ч и к М.Ф., О разностном методе приближенного решения эллиптических уравнений. Тр. мат. фак. Воронежск. ун-та, 1971, вып. 4, 117-127.

12. Т а м м е Э., О регулярной сходимости разностных аппроксимаций задачи Дирихле. Изв. АН Эст.ССР. Физ., мат., 1977, 26, № 1, 3-8.

13. S c h u m a n n, R., Z e i d l e r, E., The finite difference method for quasilinear elliptic equations of order $2m$. Numer. Funct. Anal. and Optimiz., 1979, 1, № 2, 161-194

Поступило
I IУ 1982

ÜBER DIE KONVERGENZ DER DIFFERENZENMETHODE IN DER STARKEN NORM FÜR NICHTLINEARE ELLIPTISCHE DIFFERENTIALGLEICHUNG

M. Fischer

Zusammenfassung

Für die Differentialgleichung zweiter Ordnung mit den schwachen Nichtlinearitäten in den Koeffizienten untersucht man die Fragen der Konvergenz der Differenzenmethode und der Iterationsmethode in der Metrik des Sobolev-Raumes H^2. Die Resultate berühen sich auf den Ungleichungen der Koerzitivität.

Уч.зап.Тартуск.ун-та,1982,633,67-74

ИТЕРАТИВНОЕ РЕШЕНИЕ ИНТЕГРАЛЬНОГО УРАВНЕНИЯ СО СЛАБО ОСОБЕННЫМ ЯДРОМ

П. Уба

В статье рассматривается интегральное уравнение, ядро которого имеет логарифмическую или степенную особенность. Предлагается итеративный процесс коллокационного метода с кусочно-полиномиальной аппроксимацией решения на неравномерной сетке. Итеративный метод основан на использовании другой, более грубой аппроксимации того же интегрального уравнения.

1. Рассмотрим интегральное уравнение

$$u(t) = \int_0^b \varkappa(|t-s|)\, u(s)\, ds + f(t). \tag{1.1}$$

Пусть $f \in C^m[0,b]$, $\varkappa \in C^{m-1}(0,b]$, $m \geq 2$, причем

$$|\varkappa(t)| \leq c(|\ln t| + 1), \quad |\varkappa^{(k)}(t)| \leq c\, t^{-k}, \quad k = 1, \ldots, m-1 \tag{1.2}$$

или

$$|\varkappa^{(k)}(t)| \leq c\, t^{-k-\alpha}, \quad 0 < \alpha < 1, \quad k = 0, 1, \ldots, m-1. \tag{1.3}$$

Предположим, что соответствующее однородное интегральное уравнение имеет только нулевое решение. Тогда (см. [6]) уравнение (1.1) имеет единственное решение u и $u \in C[0,b] \cap C^m(0,b)$. В случае степенной особенности (1.3) справедливы оценки

$$|u^{(k)}(t)| \leq c_0 (t^{-k-\alpha+1} + (b-t)^{-k-\alpha+1}), \quad k = 1, \ldots, m, \quad c_0 = const; \tag{1.4}$$

в случае логарифмической особенности (1.2) неравенство (1.4) справедливо при $k = 2, \ldots, m$ с $\alpha = 0$ и $|u'(t)| \leq c_0(|\ln t| + |\ln(b-t)|)$. На основании этой информации построим коллокационный метод с кусочно-полиномиальной аппроксимацией решения на неравномерной сетке (см. [7]). В качестве узлов аппроксимации берем

$$t_i = (b/2)(i/n)^r \quad (i = 0, 1, \ldots, n), \quad t_{n+i} = b - t_{n-i} \quad (i = 1, \ldots, n), \tag{1.5}$$

где число r ($r \geq 1$) характеризует степень неравномерности сетки. На стандартном отрезке $[-1, 1]$ определим какие-нибудь точки интерполяции τ_k ($k = 1, \ldots, l$):

$$-1 \leq \tau_1 < \tau_2 < \ldots < \tau_\ell \leq 1 \qquad (I.6)$$

и при помощи линейного преобразования перенесём их на отрезки $[t_i, t_{i+1}]$: $\tau_{ik} = t_i + (\tau_k + 1)(t_{i+1} - t_i)/2$ $(k=1,\ldots,\ell;\ i=0,\ldots,2n-1)$.

Приближенное решение $u_n(t)$ уравнения (I.I) строим в виде кусочно-полиномиальной функции степени $\ell - 1$ на сетке (I.5) (в точках t_i, $i=1,\ldots,2n-1$ она, вообще говоря, разрывна). Потребуем, чтобы u_n удовлетворяло уравнению (I.I) в точках интерполяции:

$$u_n(\tau_{ik}) = \int_0^b \varkappa(|\tau_{ik} - s|)\, u_n(s)\, ds + f(\tau_{ik})$$

$$(k=1,\ldots,\ell;\ i=0,1,\ldots,2n-1). \qquad (I.7)$$

Из [7] известно, что при достаточно больших n задача (I.7) однозначно определяет приближение $u_n(t)$ и если $\tau = \mu/(1-\alpha)$, $\mu \leq \ell$, то $\sup\limits_{0\leq t\leq b} |u_n(t) - u(t)| \leq const \cdot n^{-\mu}$; в точках интерполяции имеет место сверхсходимость.

Вид системы (I.7) конкретизируется выбором базиса в подпространстве кусочно-полиномиальных функций. В дальнейшем рассмотрим u_n на каждом частичном отрезке в виде

$$u_n(t) = \sum_{\nu=1}^{\ell} a_{j\nu} \cdot \varphi_{j\nu}(t), \quad t_j \leq t \leq t_{j+1} \qquad (I.8)$$

где $\varphi_{j\nu}$ фундаментальные многочлены Лагранжа порядка $\ell - 1$ (т.е. $\varphi_{j\nu}(\tau_{jk}) = \delta_{\nu k}$, $k,\nu = 1,\ldots,\ell$). Теперь условия (I.7) приобретают вид системы уравнений

$$a_{ik} = \sum_{j=0}^{2n-1} \sum_{\nu=1}^{\ell} \int_{t_j}^{t_{j+1}} \varkappa(|\tau_{ik} - s|)\, \varphi_{j\nu}(s)\, ds \cdot a_{j\nu} + f(\tau_{ik}) \qquad (I.9)$$

относительно неизвестных a_{ik} $(k=1,\ldots,\ell;\ i=0,1,\ldots,2n-1)$. Если для некоторых t_i справедливо $t_i = \tau_{i,1} = \tau_{i-1,\ell}$, то соответствующее уравнение входит в систему (I.9) только один раз.

Как правило, при малых значениях n (в дальнейшем обозначим их через m) решение системы (I.9) не представляет трудностей, а при больших значениях n возникающие трудности значительны.

2. Если через T обозначить интегральный оператор, то уравнение (I.I) можно рассматривать как операторное уравнение $u = Tu + f$ в банаховом пространстве $C[0,\ell] \equiv E$. С помощью формулы

$$(P_n u)(t) = \sum_{k=1}^{\ell} u(\tau_{ik})\, \varphi_{ik}(t), \quad t_i \leq t \leq t_{i+1}, \quad i = 0,1,\ldots,2n-1$$

определим интерполяционный проектор $P_n : E \to E_n$, сопоставляющий любой $u \in E$ её кусочно-полиномиальный интерполянт $u_n \in E_n$, где E_n – конечномерное пространство всех функций вида (I.8) с нормой $\|u_n\| = \sup\limits_{0\leq t\leq b} |u_n(t)|$. Задачу (I.7) можно переписать в равносильной форме

$$u_n = P_n T u_n + P_n f, \qquad (2.1)$$

где оператор T естественным образом распространен на кусочно-непрерывные функции. Для определённости будем считать, что функции из E_n непрерывны справа. Помимо пространства E_n введем пространство E_m, где точки аппроксимации и интерполяции обозначим через $t_i^{(m)}$ и $\tau_{ij}^{(m)}$ соответственно, а фундаментальные многочлены Лагранжа через $\varphi_{ij}^{(m)}$ $(i=0,1,\dots,2m-1;\ j=1,\dots,\ell)$. Оператор P_m применим и к $u_n \in E_n$, причем $P_m u_n \in E_m$ и представляется в виде

$$(P_m u_n)(t) = \sum_{k=1}^{\ell} u_n(\tau_{ik}^{(m)})\varphi_{ik}^{(m)}(t), \quad t_i^{(m)} \le t \le t_{i+1}^{(m)}, \quad i=0,1,\dots,2m-1.$$

Следующие две теоремы, доказательство которых приводится в пункте 4, представляют основные результаты данной статьи.

<u>Теорема 1</u>. При любом $u_n^{(0)} \in E_n$ итерационный процесс

$$u_n^{(\mu+1)} = u_n^{(\mu)} + \tau_n^{(\mu)} + P_n(I - P_m T)^{-1}(P_m T) P_m \tau_n^{(\mu)}, \qquad (2.2)$$
$$\tau_n^{(\mu)} = P_n f - (I - P_n T) u_n^{(\mu)}$$

при $\mu \to \infty$ и достаточно больших n и m сходится к решению u_n^* уравнения (2.1).

<u>Теорема 2</u>. При любом $u_n^{(0)} \in E_n$ итерационный процесс

$$u_n^{(\mu+1)} = u_n^{(\mu)} + \tau_n^{(\mu)} + P_n(I - P_m T)^{-1} P_m (P_n T) \tau_n^{(\mu)}, \qquad (2.3)$$
$$\tau_n^{(\mu)} = P_n f - (I - P_n T) u_n^{(\mu)}$$

при $\mu \to \infty$ и достаточно больших n и m сходится к решению u_n^* уравнения (2.1).

Заметим, что каждый шаг процесса (2.2) или (2.3) требует решения задачи $z_m = P_m T z_m + g_m$, где, в зависимости от метода, $g_m = (P_m T) P_m \tau_n^{(\mu)}$ или $g_m = P_m (P_n T) \tau_n^{(\mu)}$. Практически решается алгебраическая система уравнений типа (1.9), порядок которой определяется числом m. Если $m \ll n$, то при малом числе шагов количество арифметических операции сильно уменьшается по сравнению с решением системы (1.9).

3. После некоторых простых преобразований можно (2.2) и (2.3) привести к виду, аналогичному описанному в [5] первому и второму итерационным методам, где все расчёты проводятся в функциональном пространстве E. Практическое выполнение этих операций связано затруднениями. По сравнению с формулами из [5], в последнее слагаемое в (2.2) и (2.3) добавлены операторы $P_n : E_m \to E_n$ и $P_m : E_n \to E_m$, обеспечивающие практическое

выполнение расчётов в конечномерных пространствах E_n и E_m. Определение m и n, гарантирующее сходимость процессов, весьма неудобное. К. Аткинсон в [5] утверждает, что для сходимости процесса достаточно одной правильной значащей цифры в $u_n^{(0)}$.

В расчетах элементы пространств E_n и E_m рассматриваются как векторы с координатами $u_{ik} = u_n(\tau_{ik})$ и $u_{jv}^{(m)} = u_m(\tau_{jv}^{(m)})$ $(i=0,1,\dots,2n-1;\ j=0,1,\dots,2m-1;\ k,v=1,\dots,\ell)$, а оператор P_nT (аналогично и P_mT) представляет собою матрицу с элементами

$$b_{ij,vk} = \int_{t_j}^{t_{j+1}} \varkappa(|\tau_{ij}-s|)\,\varphi_{vk}(s)\,ds,\quad i,v=0,1,\dots,2n-1,\ j,k=1,\dots,\ell.$$

При переходе из одного пространства в другое используются фундаментальные многочлены Лагранжа

$$\varphi_{ik}(t) = \prod_{\substack{\mu=1\\ \mu\neq k}}^{\ell} (t-\tau_{i\mu})/(\tau_{ik}-\tau_{i\mu}),\quad t_i \le t \le t_{i+1}$$

и, например, для $P_m: E_n \to E_m$ координаты вектора $u_m = P_m u_n$ вычисляются по формуле

$$u_{jv}^{(m)} = \sum_{k=1}^{\ell} u_{ik}\,\varphi_{ik}(\tau_{jv}^{(m)}),\quad t_i \le \tau_{jv}^{(m)} \le t_{i+1},\quad j=0,1,\dots,2m-1,\ v=1,\dots,\ell.$$

В итоге вычисления в (2.2) и (2.3) сводятся к операциям с матрицами и векторами.

4. Напомним некоторые определения из [I]. Пусть даны банаховы пространства $\mathcal{U}, \mathcal{V}$ и $\mathcal{U}_n, \mathcal{V}_n$ $(n=1,2,\dots)$, а также системы $\varphi = \{\varphi_n\}$ и $\psi = \{\psi_n\}$ линейных ограниченных операторов $\varphi_n: \mathcal{U} \to \mathcal{U}_n$ и $\psi_n: \mathcal{V} \to \mathcal{V}_n$ таких, что при $n \to \infty$ для любых $x \in \mathcal{U}$ и $y \in \mathcal{V}$ справедливо $\|\varphi_n x\| \to \|x\|$ и $\|\psi_n y\| \to \|y\|$. Операторы φ_n и ψ_n будем называть связывающими.

Определение. Последовательность $\{x_n \mid x_n \in \mathcal{U}_n\}$ φ-сходится (или дискретно сходится) к $x \in \mathcal{U}$ если $\|x_n - \varphi_n x\|_{\mathcal{U}_n} \to 0$ при $n \to \infty$; обозначим $x_n \xrightarrow{\varphi} x$. Последовательность операторов $\{A_n \mid A_n : \mathcal{U}_n \to \mathcal{V}_n\}$ $\varphi\psi$-сходится (или дискретно сходится) к оператору $A: \mathcal{U} \to \mathcal{V}$, если для любой φ-сходящейся последовательности $\{x_n\}$ имеет место соотношение $x_n \xrightarrow{\varphi} x \Rightarrow A_n x_n \xrightarrow{\psi} Ax$; обозначим $A_n \xrightarrow{\varphi\psi} A$. Последовательность операторов $\{A_n\}$ компактно сходится к оператору A, если $A_n \xrightarrow{\varphi\psi} A$ и выполнено условие: $\|x_n\| \le const \Rightarrow \{A_n x_n\}$ ψ-компактна.

Свойства дискретной и компактной сходимости, указанные в [I] и [2], ниже считаются известными.

Учитывая, что $\|P_n u - u\| \to 0$ при $n \to \infty$ (см. [7]), легко видеть, что проектор $P_n: E \to E_n$ является связывающим оператором.

Предложение. Последовательность операторов $P_m: E_n \to E_m$ дискретно сходится к тождественному оператору $I: E \to E$ по отношению к связывающим операторам $P_n: E \to E_n$ и $P_m: E \to E_m$.

Для доказательства отметим, что нормы $\| P_n \|$ равномерно ограничены (см. [7]) и если $u_n \xrightarrow{\mathcal{P}} u$ то

$$\| P_m u_n - P_m I u \| \le \| P_m \| (\| u_n - P_n u \| + \| P_n u - u \|) \to 0 .$$

Лемма. Последовательность операторов $P_n T : E_n \to E_n$ компактно сходится к оператору $T : E \to E$.

Доказательство. Нормы $\| P_n T \|$ равномерно ограниченны. Пусть $u_n \xrightarrow{\mathcal{P}} u$. Тогда из сходимости

$$\| P_n T u - P_n T u_n \| \le \| P_n T \| (\| u - P_n u \| + \| P_n u - u_n \|) \to 0$$

следует дискретная сходимость $P_n T \xrightarrow{\mathcal{PP}} T$. Пусть $\| u_n \| \le const$, $u_n \in E_n$. В силу полной непрерывности T последовательность $\{ T x_n \}$ относительна компактна в E и следовательно $\{ P_n T x_n \}$ $\mathcal{P}$-компактна. Лемма доказана.

Сформулируем один результат о сходимости итерационных методов, который в основном имеется в [2]. Наша формулировка незначительно отличается от формулировки [2]; доказательство повторяет рассуждения [2], поэтому опускается.

Пусть A и B линейные операторы в пространстве $\mathcal{U}$ и пусть $\varphi_n \in \mathcal{L}(\mathcal{U}, \mathcal{U}_n)$ связывающие операторы. Пусть для уравнения $Au + Bu = f$ построена последовательность дискретизаций

$$A_n u_n + B_n u_n = f_n , \tag{4.I}$$

где $u_n, f_n \in \mathcal{U}_n$ и $A_n, B_n \in \mathcal{L}(\mathcal{U}_n, \mathcal{U}_n)$. Полагаем, что A и A_n имеют ограниченные обратные и будем считать заданными операторы $Q_{nm} \in \mathcal{L}(\mathcal{U}_m, \mathcal{U}_n)$ и $R_{nm} \in \mathcal{L}(\mathcal{U}_n, \mathcal{U}_m)$. Если а) оператор $A^{-1}B$ вполне непрерывен; б) оператор $I + A^{-1}B$ обратим; в) последовательность операторов $A_n^{-1} B_n$ компактно сходится к $A^{-1}B$; г) операторы Q_{nm} и R_{nm} φ-сходятся к $I : \mathcal{U} \to \mathcal{U}$, то при достаточно больших n и m сходится к решению u_n^* уравнения (4.I) итерационный процесс:

$$u_n^{(\mu+1)} = v_n^{(\mu+1)} + Q_{nm} w_m^{(\mu+1)}, \quad A_n v_n^{(\mu+1)} = f_n - B_n u_n^{(\mu)}$$
$$A_m w_m^{(\mu+1)} + B_m w_m^{(\mu+1)} = B_m R_{nm} (u_n^{(\mu)} - v_n^{(\mu+1)}). \tag{4.2}$$

(Здесь индексы n и m подчеркивают принадлежность пространствам $\mathcal{U}_n$ и $\mathcal{U}_m$ соответственно.)

Доказательство теоремы I. Сделаем в (4.I) и (4.2) следующие замены: $A = I$, $A_n = I$, $B = -T$, $B_n = -P_n T$, $R_{nm} = P_m$, $f_n = P_n f$, $Q_{nm} = P_n$, $\varphi_n = P_n$. Отметим, что свойства а) и б) обеспечены постановкой задачи (I.I), а свойства в) и г) доказаны в лемме и в предложении соответственно. Переписывая (4.2) в новых обозначениях, получим $u_n^{(\mu+1)} = v_n^{(\mu+1)} + P_n w_m^{(\mu+1)}$, $v_n^{(\mu+1)} = P_n f + (P_n T) u_n^{(\mu)}$, $(I - P_m T) w_m^{(\mu+1)} = -(P_m T) P_m (u_n^{(\mu)} - v_n^{(\mu+1)})$, откуда после замены $v_n^{(\mu+1)}$ и $w_m^{(\mu+1)}$ в первом уравнении их значениями получается метод (2.2).

Для доказательства теоремы 2 учтем, что $A_n = I$, $A = I$ и заменим в (4.2) последнее равенство на $\omega_m^{(\mu+1)} + B_m \omega_m^{(\mu+1)} = R_{nm} B_n (u_n^{(\mu)} - \vartheta_n^{(\mu+1)})$. Аналогично доказательству сходимости процесса (4.2) показывается сходимость полученного метода к решению задачи (4.I). Остальное повторяет доказательства теоремы I.

5. Для численного апробирования процессов (2.2) и (2.3) решалось уравнение

$$u(t) = \frac{1}{4}\int_0^{0.3} E(|t-s|)u(s)ds + \frac{1}{8}e^{2t-0.6}, \quad E(\tau) = \int_\tau^\infty \frac{e^{-s}}{s}ds, \quad (5.I)$$

при нескольких значениях n и ℓ сначала методом Гаусса, а затем итерационными методами (2.2) и (2.3). Исходное $u_n^{(0)}$ определяли $u_n^{(0)} := P_n u_m^*$, где u_m^* решение задачи с грубой аппроксимацией. Для получения одинаковой точности при равных n требовалось методом (2.2) не более двух шагов итераций даже при малых значениях m (например $m = 3$, $n = 40$, $\ell = 2$). Методом (2.3) достигалась такая же точность уже после первого шага. По сравнению с методом (2.2) решение методом (2.3) сопровождается увеличением количества арифметических операций на каждом шагу итерации, но зато уменьшается число итерационных шагов. Но затраты времени ЭВМ, как отмечено и в [5], в большинстве случаев остаются примерно равными.

Еще замечено, что при решении методом (2.2) после третьего шага итерации шестая значащая цифра уже не изменялась, а после седьмого шага достигалась машинная двойная точность (вычисления проведены в ВЦ ТГУ на ЭВМ ЕС-I022 с использованием двойной точности).

Некоторые из результатов приведены в таблице, в которой в целях сравнения добавлены результаты, полученные другими авторами другими методами. Символом N обозначено число точек интерполяции на отрезке $[0, 0.3]$. Отметим, что в двух последних разделах аргументы $t = 0.0$, $t = 0.15$ и $t = 0.3$ являются точками интерполяции, а значения в точках $t = 0.1$ и $t = 0.2$ вычислены по соответствующей кусочно-полиномиальной функции.

Приближенное решение уравнения (5.I)

t =	0.0	0.I	0.I5	0.2	0.3
Метод инвариантного погружения (см. [3])					
	0.082559	0.I02958	—	0.I2285I	0.I40966
Метод механических квадратур (см. [4])					
N=24	0.082634	0.I02950	—	0.I22780	0.I40960
N=96	0.082578	0.I02958	—	0.I22824	0.I40965
Кусочно-линейная коллокация на равномерной сетке (см. [4])					
N=I2	0.082558	0.I02960	—	0.I22837	0.I40962
N=24	0.08256I	0.I02963	—	0.I2284I	0.I40968
Кусочно-полиномиальная коллокация на неравномерной сетке; система (I.9) решается методом Гаусса					
ℓ=3, n=3, N=I3	0.0825627	0.I029666	0.II28538	0.I228468	0.I409692
ℓ=3, n=6, N=25	0.0825633	0.I029636	0.II28539	0.I228386	0.I409700
ℓ=2, n=I2, N=25	0.0825629	0.I029656	0.II28536	0.I2284I0	0.I409694
ℓ=2, n=20, N=4I	0.0825632	0.I029655	0.II28538	0.I228423	0.I409700
Кусочно-полиномиальная коллокация на неравномерной сетке; система (I.9) решается итерационными методами					
m=2, n=I0, ℓ=3, N=2I	0.0825633	0.I029654	0.II28539	0.I228437	0.I40970I
m=2, n=5, ℓ=4, N=3I	0.0825632	0.I029656	0.II28538	0.I228439	0.I409700
m=3, n=20, ℓ=2, N=4I	0.082563I	0.I029654	0.II28538	0.I228423	0.I409698
m=3, n=40, ℓ=2, N=8I	0.0825632	0.I02965I	0.II28539	0.I228427	0.I409700
m=5, n=37, ℓ=3, N=I49	0.0825633	0.I029650	0.II28539	0.I228429	0.I40970I
m=4, n=75, ℓ=2, N=I5I	0.0825633	0.I02965I	0.II28539	0.I228429	0.I40970I

Литература

1. Вайникко Г., Анализ дискретизационных методов. Тарту, 1976.
2. Даугавет И.К., Об итеративном решении уравнений, возникающих при компактной аппроксимации операторов. Ж. вычисл. матем. и матем. физ., 1980, 20, № 4, 1046–1049.
3. Касти Дж., Калаба Р., Методы погружения в прикладной математики. "Мир", М., 1976.
4. Педас А.А., О приближённом решении интегральных уравнений со слабой особенностью. Канд. дисс., Тарту, 1978.
5. Atkinson K., Iterative variants of the Nystrom method for the numerical solution of integral equations. Numer. Math., 1973, 22, № 1, 17–31.
6. Vainikko, G., and Pedas, A., The properties of solutions of weakly singular integral equations. J. Austral. Math. Soc., 1981, B 22, 4, 419–430.
7. Vainikko, G. and Uba, P., A piecewise polynomial approximation to solution of an integral equation with weakly singular kernel. J. Austral. Math. Soc., 1981, B 22, 4, 431–438.

Поступило
20 УП 1981

ITERATIVE SOLUTION OF AN INTEGRAL EQUATION WITH WEAKLY SINGULAR KERNEL

P. Uba

Summary

The object of this paper is the numerical solution of integral equation (1.1) with a logarithmically or algebraically singular kernel. The collocation method on a non-uniform grid with piecewise polynomial approximation reduces integral equation (1.1) to the system of equations (1.9), whose operator form is (2.1). The convergence of iterative methods (2.2) and (2.3) to the solution of equation (2.1) is proved. Both iterative methods are based on the use of another, much more rough approximation of the same integral equation. The numerical testing, some results of which are given, demonstrates a rapid convergence and an essentially shorter computing time, compared with the inversion of (1.9).

Уч.зап.Тартуск.ун-та,1982,633,75–87

ОБ УСТОЙЧИВОСТИ РЕШЕНИЯ ОПЕРАТОРНОГО УРАВНЕНИЯ С НЕЛИНЕЙНОСТЬЮ ВТОРОГО ПОРЯДКА

Б. Шифрин

Введение. Ряд задач математической физики (см. [3]) порождает уравнения вида $Ax + \check{K}x = \check{f}$, где оператор $A \in \mathcal{L}(X, E)$ – линейный непрерывно обратимый, $\check{K}: X \to E$ – нелинейный второго порядка (см. ниже); X – некоторое банахово пространство. Полагая $f := A^{-1}\check{f} \in X$, $K := A^{-1}\check{K}$ можно перейти к равносильному уравнению в X:

$$x + Kx = f . \tag{I}$$

В данной работе исследуется корректность задачи (I) и даются оценки устойчивости решения по f, K. Устанавливается и устойчивость соответствующего (I) итерационного процесса. Результаты приложимы, в частности, к модифицированному процессу Ньютона. В заключительной части работы предложен некий общий подход к изучению возмущений операторных уравнений.

§ I. Описание уравнения и его корректная разрешимость

I.I. Нелинейности второго порядка. Пусть $W \subset X$ некоторая окрестность нуля. Назовём оператор $K: W \to X$ нелинейным порядка $l > 1$ в W, если выполнены условия:

i) $\forall r > 0$, $\forall x, y \in W$: $\|x\|, \|y\| \le r$ имеем

$$\|Kx - Ky\| \le c\, r^{l-1} \|x - y\| ; \tag{I.I}$$

ii) $K0 = 0$.

Принципиальным для нас будет именно условие i). Очевидно, множество $\Pi^{(2)} := \Pi^{(2)}(W)$ $(l = 2)$ таких операторов линейно; оно превращается в банахово пространство введением, например, следующей нормы:

$$\|K\|_{(1)} = \sup_{x, y \in W,\, x \ne y} \frac{\|Kx - Ky\|}{(\|x\| + \|y\|)\|x - y\|} . \tag{I.2}$$

Полнота $\Pi^{(2)}$ проверяется стандартным образом. Отметим равносильное определению (I.2) неравенство с минимальной константой $c' = \|K\|_{(1)}$

$$\|Kx - Ky\| \le \|K\|_{(1)} (\|x\| + \|y\|) \|x - y\| . \tag{I.3}$$

Пример I.I. Уравнения типа Навье—Стокса (см. [3]) приводятся к виду (I) с $Kx = B(x,x)$, где $B: X \times X \to X$ – такое отображение, что

$$\text{i)}\ \|B(x^1,y)-B(x^2,y)\| \le \beta\|y\|\,\|x^1-x^2\|,\quad \|B(x,y^1)-B(x,y^2)\| \le \beta\|x\|\,\|y^1-y^2\|,$$

$$\text{ii)}\quad B(0,x)=0 \quad (\forall\ x,y,\ x^1,x^2,\ y^1,y^2 \in X).$$

В частности, B может быть билинейным или линейным по одному из аргументов. Отметим теперь, что из i) следует

$$\|B(x^1,x^1)-B(x^2,x^2)\| \le \beta(\|x^1\|+\|x^2\|)\|x^1-x^2\|.$$

Таким образом, $K \in \Pi^{(2)}$, $\|K\|_{(1)} \le \beta$.

Уравнение (I.2) с $Kx = B(x,x)$ рассмотрено в [3]; результаты о его разрешимости переносятся на общий случай без каких-либо изменений.[1] Формулировке предпошлём некоторые обозначения. Ниже $Ш(0,r) = \{x \in X : \|x\| < r\}$; $\varkappa = \varkappa(\beta,\gamma) := 4\beta\gamma$ $(\forall \beta,\gamma \ge 0)$; $r_1 = r_1(\beta,\gamma)$, $r_2 = r_2(\beta,\gamma)$ $(r_1 \le r_2)$ – корни уравнения $\beta t^2 - t + \gamma = 0$ (предполагается, что они вещественны):

$$r_1 = (1-\sqrt{1-\varkappa})/(2\beta),\quad r_2 = (1+\sqrt{1-\varkappa})/(2\beta). \tag{I.4}$$

Константы $\beta, \gamma > 0$ будут обычно связаны с уравнением (I) условиями типа

$$\|K\|_{(1)} \le \beta,\quad \|f\| \le \gamma. \tag{I.5}$$

Теорема I.I. Пусть уравнение (I) с нелинейным оператором $K \in \Pi^{(2)}(W)$ таково, что при некоторых β, γ, удовлетворяющих (I.5), выполнены условия:

$$W \supset \overline{Ш}(0,\check{r})\ \text{для некоторого}^{2}\ \check{r} \in [r_1, r_2[, \tag{I.6}$$

$$\varkappa := 4\beta\gamma \le 1. \tag{I.7}$$

Тогда уравнение (I) имеет решение $x_* \in \overline{Ш}(0,r_1) \subset \overline{Ш}(0,\check{r})$, единственное в шаре $\overline{Ш}(0,\check{r})$.

Доказательство более сильного утверждения мы наметим в § 2.

I.2. Оценки устойчивости решения. Пусть известно, что $\|K\|_{(1)} \le \beta$ с некоторым $\beta > 0$. Поставим, в свете теоремы I.I, вопрос об устойчивости решения x_* по правой части $f \in \overline{Ш}(0, \frac{1}{4\beta})$.

---

[1] По-существу эти результаты близки к теоремам Л.В.Канторовича о методе Ньютона (см. [I], стр. 680).

[2] Ниже мы полагаем $[r_1, r_2[:= \{r_1\}$ при $r_1 = r_2$.

Теорема I.2. Пусть $f, \tilde{f} \in \overline{Ш}(0, \frac{1}{4\beta})$ где $\beta \geq \|K\|_{(1)}$. Положим $\|f - \tilde{f}\| \leq \delta$. Тогда для решений $x_*, \tilde{x}_*$ уравнения (I) с оператором $K \in \Pi^{(2)}(W)$, $W \supset \overline{Ш}(0, \frac{1}{2\beta})$ и правыми частями $f, \tilde{f}$ справедливы следующие оценки:

1^0 Если $\min\{\|f\|, \|\tilde{f}\|\} < 1/(4\beta)$ то

$$\|x_* - \tilde{x}_*\| \leq \frac{2\delta}{(1-\varkappa)^{1/2} + (1-\tilde{\varkappa})^{1/2}} \quad (\varkappa := 4\beta\|f\|, \tilde{\varkappa} = 4\beta\|\tilde{f}\|). \tag{I.8}$$

2^0 В общем случае

$$\|x_* - \tilde{x}_*\| \leq 2(1/\beta - \delta)^{\frac{1}{2}}\delta^{\frac{1}{2}} \leq 2\beta^{-\frac{1}{2}}\delta^{\frac{1}{2}}. \tag{I.9}$$

Сформулируем более общее утверждение, допустив возмущение и по $K \in \Pi^{(2)}(W)$. Именно, пусть $\tilde{x}_*$ - решение уравнения

$$x + \tilde{K}x = \tilde{f}. \tag{$\tilde{\text{I}}$}$$

Предполагается, что

$$\begin{gathered} K, \tilde{K} \in \Pi^{(2)}(W), \ \|K\|_{(1)} \leq \beta, \ \|\tilde{K}\|_{(1)} \leq \tilde{\beta}, \\ \|f\| \leq \gamma, \ \|\tilde{f}\| \leq \tilde{\gamma}, \quad W \supset Ш(0, \tau_{max}) \\ (\tau_{max} := \max\{\tau_1, \tilde{\tau}_1\}), \ \|f - \tilde{f}\| \leq \delta, \ \|K - \tilde{K}\| \leq \eta. \end{gathered} \tag{I.10}$$

Теорема I.3. Пусть в предположении (I.10) выполнено условие

$$\varkappa_{max} := \max\{\varkappa, \tilde{\varkappa}\} \leq 1.$$

Тогда для решений $x_*, \tilde{x}_*$ уравнений (I), ($\tilde{\text{I}}$) справедливы оценки:

1^0 Если $\min\{\varkappa, \tilde{\varkappa}\} < 1$, то

$$\|x_* - \tilde{x}_*\| \leq \frac{2\delta + 2\tau_{max}^2\eta}{(1-\varkappa)^{1/2} + (1-\tilde{\varkappa})^{1/2}} \leq \frac{\delta + \tau_{max}^2\eta}{(1-\varkappa_{max})^{1/2}}. \tag{I.11}$$

2^0 В общем случае в предположении $\delta \leq 1/(2\beta_{min})$

$$\begin{gathered} \|x_* - \tilde{x}_*\| \leq 2(1/\beta_{min} - \delta)^{1/2}(\delta + \tau_{max}^2\eta)^{1/2} \leq 2\beta_{min}^{-1/2}(\delta + \tau_{max}^2\eta)^{1/2} \leq \\ \leq \beta_{min}^{-3/2}(4\beta_{min}^2\delta + \eta)^{1/2} \\ (\beta_{min} := \min\{\beta, \tilde{\beta}\}). \end{gathered} \tag{I.12}$$

Замечание I.I. В (I.12) учтено, что $\tau_{max} \leq 1/(2\beta_{min})$. Отметим, что

$$\|x_* - \tilde{x}_*\| \leq \tau_1 + \tilde{\tau}_1 \leq 2\tau_{max} \leq 1/\beta_{min}. \tag{I.13}$$

Доказательство теоремы I.3. Случай 1^0. Пусть, например, $\beta_{min} = \beta$. Имеем $x_* = f - Kx_*$, $\tilde{x}_* = \tilde{f} - \tilde{K}\tilde{x}_*$. Вычитая из одного равенства другое и преобразуя, получим

$$x_* - \tilde{x}_* = (f - \tilde{f}) + (K\tilde{x}_* - Kx_*) + (\tilde{K} - K)\tilde{x}_*. \tag{I.14}$$

Отсюда на основе (I.3) и (I.10) вытекает, что

$$\|x_* - \tilde{x}_*\| \leq \delta + \beta(\|x_*\| + \|\tilde{x}_*\|)\|x_* - \tilde{x}_*\| + \eta\|\tilde{x}_*\|^2. \tag{I.15}$$

Здесь мы учтём, что $\|x_*\| \leq \tau_1$, $\|\tilde{x}_*\| \leq \tilde{\tau}_1$, следовательно,

$$\|x_*-\tilde{x}_*\| \le \delta + \beta(\tau_1+\tau_2)\|x_*-\tilde{x}_*\| + \eta\,\tau_{max}^2 =$$
$$= (\delta + \tau_{max}^2\eta) + \beta\left(\frac{1-\sqrt{1-\varkappa}}{2\beta} + \frac{1-\sqrt{1-\tilde{\varkappa}}}{2\tilde{\beta}}\right)\|x_*-\tilde{x}_*\| \le$$
$$\le (\delta + \tau_{max}^2\eta) + \left(1 - \frac{\sqrt{1-\varkappa}+\sqrt{1-\tilde{\varkappa}}}{2}\right)$$

или $\dfrac{\sqrt{1-\varkappa}+\sqrt{1-\tilde{\varkappa}}}{2}\|x_*-\tilde{x}_*\| \le \delta + \tau_{max}^2\eta$. Этим (I.II) доказано.

Случай 2^0. 2^0а. Рассмотрим подслучай, когда считается, что

$$\delta + \tau_{max}^2\eta < 1/\beta - \delta. \qquad \text{(I.I6)}$$

Введем $f_1 = (1-\lambda)f$, $f_2 = (1-\lambda)\tilde{f}$ $(0 \le \lambda \le 1)$. Тогда $\varkappa_1 := 4\beta\|f_1\| \le (1-\lambda)\varkappa \le 1-\lambda$; аналогично $\varkappa_2 = 4\beta\|f_2\| \le 1-\lambda$. Отметим, что $\|f_1 - f_2\| \le (1-\lambda)\|f-\tilde{f}\| \le (1-\lambda)\delta$. Кроме того,

$$\|f - f_1\| \le \lambda\gamma_{max}, \quad \|\tilde{f} - f_2\| \le \lambda\gamma_{max} \quad (\gamma_{max} := \max\{\gamma, \tilde{\gamma}\})$$

Пусть x_1, x_2 – решения, соответствующие f_1, f_2, т.е. выполнено $x_1 + Kx_1 = f_1$, $x_2 + Kx_2 = f_2$ Применяя оценки (I.II) и (I.8), получим

$$\|x_1 - x_2\| \le \frac{2\|f_1-f_2\| + 2\tau_{max}^2\eta}{(1-\varkappa_1)^{1/2} + (1-\varkappa_2)^{1/2}} \le \frac{(1-\lambda)\delta + \tau_{max}^2\eta}{\lambda^{1/2}}, \qquad \text{(I.I7)}$$

$$\|x_* - x_1\| \le \frac{2\lambda\gamma_{max}}{\lambda^{1/2}}, \quad \|\tilde{x}_* - x_2\| \le \frac{2\lambda\gamma_{max}}{\lambda^{1/2}}. \qquad \text{(I.I8)}$$

Поэтому

$$\|x_*-\tilde{x}_*\| \le \|x_*-x_1\| + \|x_1-x_2\| + \|x_2-\tilde{x}_*\| \le$$
$$\le (\tau_{max}^2\eta + (1-\lambda)\delta + 4\lambda\gamma_{max})\lambda^{-\frac{1}{2}} = (p+q\lambda)\lambda^{-\frac{1}{2}},$$

где

$$p := \tau_{max}^2\eta + \delta, \quad q := 4\gamma_{max} - \delta. \qquad \text{(I.I9)}$$

Функция $\pi(\lambda) = (p+q\lambda)\lambda^{-1/2}$ достигает минимума при $\lambda = \lambda^* = \frac{p}{q}$, $\pi(\lambda^*) = 2(pq)^{1/2}$. Для нас существенно, что в рассматриваемом подслучае 2^0а будет $\lambda^* = \dfrac{\tau_{max}^2\eta + \delta}{4\gamma_{max} - \delta} \le \dfrac{\tau_{max}^2\eta + \delta}{1/\beta_{min} - \delta} < 1$ (см. (I.I6)). Итак, полагая $\lambda = \lambda^*$ в (I.I9), получаем требуемое, именно:

$$\|x_*-\tilde{x}_*\| \le 2(4\gamma_{max}-\delta)^{\frac{1}{2}}(\delta+\tau_{max}^2\eta)^{\frac{1}{2}} \le 2(1/\beta_{min}-\delta)^{\frac{1}{2}}(\delta+\tau_{max}^2\eta)^{\frac{1}{2}}. \qquad \text{(I.20)}$$

Подслучай 2^0б. Пусть теперь $\delta + \tau_{max}^2\eta \ge 1/\beta - \delta$ $(\beta = \beta_{min})$. Тогда

$$2(1/\beta_{min}-\delta)^{\frac{1}{2}}(\delta+\tau_{max}^2\eta)^{\frac{1}{2}} \ge 2(1/\beta_{min} - \delta) \ge 2(1/\beta_{min} - 1/(2\beta_{min})) = 1/\beta_{min} \ge$$
$$\ge \|x_* - \tilde{x}_*\|$$

(см. (I.I3)). Мы снова получили (I.I2). Теорема полностью доказана.

Замечание I.2. Справедлива также оценка (при $\eta \le 8\beta^3_{min}$):

$$\|x_* - \bar{x}_*\| \le 2(4\gamma_{max} - \delta)^{\frac{1}{2}}\delta^{\frac{1}{2}} + 2(4\beta_{max} - \tau^2_{max}\eta)^{\frac{1}{2}}(\tau^2_{max}\eta)^{\frac{1}{2}} \tag{I.21}$$

или, несколько огрубленно

$$\|x_* - \bar{x}_*\| \le 4\gamma^{\frac{1}{2}}_{max}\delta^{\frac{1}{2}} + 2\beta^{-1}_{min}\beta^{\frac{1}{2}}_{max}\eta^{\frac{1}{2}} \le 2\beta^{-\frac{1}{2}}_{min}\delta^{\frac{1}{2}} + 2\beta^{-1}_{min}\beta^{\frac{1}{2}}_{max}\eta^{\frac{1}{2}}. \tag{I.22}$$

Для вывода (I.21) нужно рассмотреть отдельно возмущение по f (см. (I.20) при $\eta = 0$) и отдельно по K, что делается вполне аналогично (вводим $K_1 = (1-\mu)K$, $K_2 = (1-\mu)\bar{K}$ и т.д.). Итоговая оценка (I.21) получается как сумма. Своеобразную симметричность оценки (I.21) нетрудно пояснить на одномерном примере (возмущение обычного квадратного уравнения), что является убедительным в силу § 2.

§ 2. Итерационный процесс с возмущением

2.I. Эквивалентная норма в $\Pi^{(2)}$. Пусть $K \in \Pi^{(2)}(W)$ – оператор с нелинейностью второго порядка. Положим

$$\|K\|_{(2)} = \sup_{x, y \in W,\ x \ne y} \frac{\|Kx - Ky\|}{(2\|x\| + \|x-y\|)}. \tag{2.I}$$

Сравним (2.I) с определением (I.4) нормы $\|K\|_{(1)}$. Ясно, что $\|x\| + \|y\| \le 2\|x\| + \|x-y\| \le 3(\|x\| + \|y\|)$, так что нормы эквивалентны:

$$\|K\|_{(2)} \le \|K\|_{(1)} \le 3\|K\|_{(2)}.$$

Лемма 2.I. Норма $\|K\|_{(2)}$ есть такая минимальная константа $c = c_K$, что $\forall p, q \ge 0$, $p \le q$ выполнено условие

$$\|x\| \le p,\ \|x - y\| \le q - p \implies \|Ky - Kx\| \le c_K(q^2 - p^2). \tag{2.2}$$

Доказательство вытекает из сравнения (2.2) с неравенством

$$\|Ky - Kx\| \le \|K\|_{(2)}(2\|x\| + \|x-y\|)\|x-y\|, \tag{2.3}$$

где, согласно (2.I), $c = \|K\|_{(2)}$ есть минимальная константа, при которой справедливо неравенство такого типа.

Целесообразность введения нормы $\|K\|_{(2)}$ показана ниже.

2.2. Итерационный процесс и его сходимость. Рассмотрим итерационный процесс

$$x^0 = 0, \quad x^n = Tx^{n-1} := f - Kx^{n-1} \quad (n = 1, 2, \ldots). \tag{2.4}$$

Для анализа (2.4) введем мажорирующую последовательность (см. [3]). Именно, пусть $\|K\|_{(2)} \le \beta$, $\|f\| \le \gamma$. Полагая $Ty \equiv f - Ky$ $(\forall y \in W)$, отметим, что

$$\|Ty\| \le \|Ky\| + \|f\| \le \|K\|_{(2)} \|y\|^2 + \|f\| \le \le \beta \|y\|^2 + \gamma = \varphi(\|y\|). \tag{2.5}$$

Исходя из (2.5) определим $\varphi(t) := \beta t^2 + \gamma$,

$$t_n = \varphi(t_{n-1}), \quad t_0 = 0. \tag{2.6}$$

Лемма 2.2. Последовательность t_n мажорирует x^n, т.е.:

$$\|x^n\| \le t_n, \tag{2.7}$$

$$\|x^{n+1} - x^n\| \le t_{n+1} - t_n \quad (n = 0, 1, 2, \dots). \tag{2.8}$$

Доказательство. В силу возрастания $\varphi(t)$ $(t \ge 0)$ справедлива следующая (доказываемая по индукции) импликация:

$$0 \le q_0 \le t_0, \quad 0 \le q_n \le \varphi(q_{n-1}) \Longrightarrow q_n \le t_n.$$

В нашем случае $q_0 = \|x^0\| = t_0 = 0$, $q_n := \|x^n\| = \|Tx^{n-1}\| \le \le \varphi(\|x^n\|) = \varphi(q_{n-1})$. Тем самым (2.7) доказано.

Далее, пусть для $1 \le m \le n-1$ выполнено (2.8), т.е. $\|x^{m+1} - x^m\| \le t_{m+1} - t_m$ (для $m = 0$ это верно, ибо совпадает с условием $\|f\| \le \gamma$). В частности, при $m = n-1$ будет

$$\|x_{n-1}\| \le t_{n-1}, \quad \|x_n\| \le t_n, \quad \|x^n - x^{n-1}\| \le t_n - t_{n-1}.$$

По лемме 2.I тем самым (см. оценку (2.2))

$$\|Kx^n - Kx^{n-1}\| \le \beta(t_n^2 - t_{n-1}^2) = \varphi(t_n) - \varphi(t_{n-1}) = t_{n+1} - t_n.$$

Таким образом, $\|x^{n+1} - x^n\| = \|Kx^n - Kx^{n-1}\| \le t_{n+1} - t_n$. Этим (2.8) и лемма в целом доказаны.

Наложим условие $\varkappa := 4\beta\gamma \le 1$. Тогда нетрудно показать, что $t_n \to t^*$ $(n \to \infty)$, где t^* – меньший корень уравнения $t = \varphi(t)$ или $\beta t^2 - t + \gamma = 0$ (см. [I], [3]). Но тогда из (2.8) несложно вытекает фундаментальность $\{x^n\}_{n=1}^{\infty}$. Следовательно, $\exists x_* \in X$;

$$x^n \to x_*, \quad \|x^n - x_*\| \le t^* - t_n \quad (n = 0, 1, 2, \dots), \tag{2.9}$$

Ясно, что x_* – решение исходного уравнения (можно перейти к пределу в (2.4)). Единственность этого решения в шаре Ш $(0, \tau_2)$ нетрудно доказать (см. аналогичное рассуждение в [3]). Тем самым установлена

Лемма 2.3. Результаты теоремы I.I остаются справедливыми при ослабленном предположении

$$4\|K\|_{(2)} \|f\| \le 1. \tag{2.10}$$

Решение x_* является пределом итерационного процесса (2.4), сходящегося с оценкой (2.9).

Оценки для $t^* - t_n$ приведены в [I] $(\varkappa < 1)$ и в [3] $(\varkappa = 1)$.

2.3. <u>Возмущенный итерационный процесс</u> (соответствующий возмущенному уравнению) выглядит так:

$$\tilde{x}^0 = 0 \ , \ \tilde{x}^n = \tilde{f} - \tilde{K}(\tilde{x}^{n-1}) \ (n = 1, 2, \dots) \ . \qquad (2.11)$$

Предполагается, что

$$\varkappa := 4\|\tilde{K}\|_{(2)}\|\tilde{f}\| \le 1 \ , \quad \|K - \tilde{K}\|_{(2)} \le \eta \ , \ \|f - \tilde{f}\| \le \delta \ . \qquad (2.12)$$

Пусть известны некоторые $\beta \ge \|K\|_{(2)}$, $\gamma \ge \|f\|$. Обозначим $\bar{\beta} = \beta + \eta$, $\bar{\gamma} = \gamma + \delta$. Наложим более сильное ограничение

$$\varkappa = 4\beta\gamma < \bar{\varkappa} := 4(\beta + \eta)(\gamma + \delta) \le 1 \ , \quad W \supset \overline{Ш}(0, r_1(\bar{\beta}, \bar{\gamma})) . \qquad (2.13)$$

В этом случае имеет вещественные корни и уравнение

$$t = \bar{\varphi}(t) := \bar{\beta} t^2 + \bar{\gamma} \quad (\text{ниже} \quad \bar{t}^* := r_1(\bar{\beta}, \bar{\gamma})) \qquad (2.14)$$

<u>Теорема 2.1</u>. Пусть t_n, $\bar{t}_n$ – мажорирующие последовательности для процессов (2.4), (2.11):

$$t_0 = \bar{t}_0 = 0 \ , \quad t_n = \varphi(t_{n-1}) \ , \quad \bar{t}_n = \bar{\varphi}(\bar{t}_{n-1}) \ .$$

Тогда справедливы оценки

$$\|\tilde{x}^n - x^n\| \le \bar{t}_n - t_n \ , \qquad (2.15)$$

$$\|\tilde{x}_* - x_*\| \le \bar{t}^* - t^* \quad (x^* = \lim x^n \ , \ \tilde{x} = \lim \tilde{x}^n) . \qquad (2.16)$$

Доказательство вытекает из результатов § 4.

Явная формула $\bar{t}_n - t_n := \Delta = \Delta(\beta, \gamma, \delta, \eta)$ является громоздкой при $\delta, \eta \ne 0$. Во всяком случае, оценка (1.11) (применимая и к возмущению обычного квадратного уравнения) дает:

$$\|\tilde{x}_* - x_*\| \le \bar{t}^* - t^* \le \frac{2\delta + 2(\bar{t}^*)^2 \eta}{(1 - \varkappa)^{1/2} + (1 - \bar{\varkappa})^{1/2}} \ . \qquad (2.17)$$

Здесь (см. (2.13)) $\varkappa < 1$. Ниже мы рассмотрим более общий случай $\varkappa \le 1$ (он не описывается теоремой 2.1).

Введем предположения, аналогичные (1.10):

$$\|K\|_{(2)} \le \beta, \ \|\tilde{K}\|_{(2)} \le \tilde{\beta} \ , \ \|f\| \le \gamma \ , \ \|\tilde{f}\| \le \tilde{\gamma} \ , \ W \supset Ш(0, r_{max}),$$
$$\|f - \tilde{f}\| \le \delta \ , \ \|K - \tilde{K}\|_{(2)} \le \eta \ . \qquad (2.18)$$

<u>Теорема 2.1.</u> Пусть в предположении (2.18) выполнено условие

$$\varkappa_{max} = \max\{\varkappa, \tilde{\varkappa}\} \le 1 \ .$$

Тогда для решений x_*, $\tilde{x}_*$ уравнений (1),(1) справедливы оценки:

1^0 Если $\varkappa_{max} < 1$ то

$$\|x_* - \tilde{x}_*\| \le \frac{3(\delta + r_{max}^2 \eta)}{(1 - \varkappa_{max})^{1/2}} \ . \qquad (2.19)$$

2^0 В общем случае

$$\|x_* - \tilde{x}_*\| \leq 6\,\gamma_{max}^{\frac12}\,\delta^{\frac12} + 6\,\beta_{max}^{\frac12}\,(\tau_{max}^{2}\,\eta)^{\frac12} \leq$$
$$\leq 3\,\beta_{min}^{-\frac12}\,\delta^{\frac12} + 3\,\beta_{min}^{-1}\,\beta_{max}^{\frac12}\,\eta^{\frac12}. \qquad (2.20)$$

(Эти оценки формально того же типа, что и (I.I4), (I.I6), но константы возросли.)

Доказательство достаточно провести для случая $\eta=0$ $\delta>0$ (случай $\delta=0$, $\eta>0$ рассматривается аналогично. Итоговая оценки получается как сумма). Наметим ход рассуждений. Правые части (2.I9), (2.20) возрастают с ростом β_{max}, γ_{max}, δ,η. Поэтому достаточно доказать эти оценки при минимальных возможных значениях параметров (в (2.I8) будут равенства). Введем $f_1=(1-\lambda)f$, $f_2=(1-\lambda)\tilde{f}$ и соответствующие решения x_1, x_2 (см. аналогичное рассуждение в доказательстве теоремы I.3). Полагая $\delta_1=\lambda\|f\|$ и замечая, что $\bar{\gamma}_1=\|f_1\|+\|f-f_1\|=\|f_1\|+\delta_1=\|f\|$, т.е. что $\bar{\varkappa}_1=\varkappa$ получим согласно (2.I7) (для f_2 рассуждаем аналогично):

$$\|x_* - x_1\| \leq \frac{2\lambda\gamma_{max}}{(1-\varkappa_1)^{\frac12}+(1-\varkappa)^{\frac12}}\,,\quad \|\tilde{x}_* - x_2\| \leq \frac{2\lambda\gamma_{max}}{(1-\varkappa_2)^{\frac12}+(1-\tilde{\varkappa})^{\frac12}}. \qquad (2.21)$$

Пусть, например, $\gamma_{max}=\gamma=\|f\|$. Определим $\lambda:=\frac{\delta}{\gamma_{max}+\delta}$: при таком выборе λ имеем $\bar{\gamma}_1':=\gamma_1+\delta=(1-\lambda)\gamma_{max}+(1-\lambda)\delta=\gamma_{max}$, так что $\bar{\varkappa}_1':=4\beta\bar{\gamma}_1'=4\beta\gamma_{max}=\varkappa$. Тем самым (2.I7) дает

$$\|x_1 - x_2\| \leq \frac{2\delta}{(1-\varkappa_1)^{1/2}+(1-\varkappa)^{1/2}}. \qquad (2.22)$$

Отметим, что при нашем выборе λ выполнено: $2\lambda\gamma_{max}=2(1-\lambda)\delta$. В случае 1^0 для получения (2.I9) осталось просто просуммировать оценки для $\|x_*-x_1\|$, $\|x_1-x_2\|$, $\|x_2-\tilde{x}_*\|$. В общем случае возможно $\varkappa=1$. Поэтому в знаменателях в (2.2I), (2.22) мы будем учитывать только величину $(1-\varkappa_i)^{1/2}\geq(1-(1-\lambda))^{1/2}=\lambda^{1/2}$ ($i=1,2$). Тогда суммирование оценок (с $\lambda=\delta/(\gamma_{max}+\delta)$) дает (2.20).

§ 3. Модифицированный процесс Ньютона

I.I. Выделение нелинейности. Рассмотрим уравнение вида

$$Px=f \quad (P: X\to E\,,\ Px_0=0) \qquad (3.1)$$

с гладким оператором P. К такому виду всегда можно привести уравнение типа $Fx=0$. Без ограничения общности можно считать

$x_0=0$. Итак, условимся что

$$P0=0 \quad (x_0=0) \tag{3.2}$$

Что касается гладкости P, то мы будем предполагать, что выпуклой окрестности нуля W оператор P непрерывно дифференцируем по Фреше, причем:

$$\|P'(x)-P'(y)\| \le L_p \|x-y\| \quad \forall x,y\in W\ ; \tag{3.2}$$

$$\exists\, \Gamma_0=(P'(0))^{-1}\in \mathcal{L}(E,X)\ ,\ \|\Gamma_0\|\le b_0\ . \tag{3.4}$$

Уравнение (3.I) нетрудно представить в форме (I.I), именно $P'(0)x+(Px-P'(0)x)=\check{f}$ или в форме (I.2):

$$x+\Gamma_0(Px-P'(0)x)=f \quad (f:=\Gamma_0\check{f}). \tag{3.5}$$

Лемма 3.I. Пусть отображение $\Phi: X\to X$, $\Phi 0=0$ в W дифференцируемо в W, причем

$$\|\Phi'(x)-\Phi'(y)\|\le L_\Phi\|x-y\| \quad (\forall x,y\in W),\ \Phi'(0)=0. \tag{3.6}$$

Тогда $\Phi\in\Pi^{(2)}(W)$, $\|\Phi\|_{(2)}\le\frac{1}{2}L_\Phi$.

Доказательство. Из условия $\Phi'(0)=0$ имеем

$$\Phi x-\Phi y=\{\Phi x-\Phi y-\Phi'(x)(x-y)\}+(\Phi'(x)-\Phi'(0))(x-y), \tag{3.7}$$

в выпуклой области W справедливо неравенство (см. [1])

$$\|\Phi x-\Phi y-\Phi'(x)(x-y)\|\le\frac{L_\Phi}{2}\|x-y\|^2 \quad (\forall x,y\in W). \tag{3.8}$$

Учитывая (3.8), получаем из (3.8):

$$\|\Phi x-\Phi y\|\le\frac{L_\Phi}{2}\|x-y\|^2+L_\Phi\|x\|\|x-y\|=\frac{L_\Phi}{2}(2\|x\|+\|x-y\|)\|x-y\|.$$

В силу определения (2.I) это как раз и означает, что $\Phi\in\Pi^{(2)}(W)$, $\|\Phi\|_{(2)}\le L_\Phi/2$.

Следствие. При условиях (3.2)-(3.4) уравнение с гладким оператором (уравнение (3.I)) эквивалентно уравнению с линейностью второго порядка.

Действительно, полагая $\Phi x:=\Gamma_0(Px-P'(0)x)$, получаем $\Phi'(x)=\Gamma_0(P'(x)-P'(0))=\Gamma_0P'(x)-I$, так что условия (3.6) выполнены с $L_\Phi=\|\Gamma_0\|L_p\le b_0L_p$.

Итак, к данному уравнению применимы результаты § 2. Условие разрешимости $\varkappa:=4\beta\gamma\le 1$ запишется здесь так:

$4\frac{b_0L_p}{2}\gamma=2b_0L_p\gamma\le 1$ (здесь $\gamma\ge\|f\|$), или $h:=b_0L_p\gamma\le 1/2$. Таким образом, лемма 2.3 совпадает с теоремой Л.В.Канторовича (см. [I,2]) о модифицированном процессе Ньютона, которым и является процесс (2.4).

3.2. Вопросы устойчивости. Ограничимся случаем, когда возможны возмущения правой части (3.I) (дополнительные замечания приведены в § 4). Возмущенное уравнение имеет вид

$$P x = \check{g} \qquad \text{или} \qquad x + \Phi x = g \quad (g = \Gamma_0 \check{g} = \tilde{f}).$$

Итерационный процесс (2.4) (с $K = \Phi$) для исходного и возмущенного уравнений приводится к виду

$$x^{n+1} = x^n - \Gamma_c P x^n + f, \quad \tilde{x}^{n+1} = \tilde{x}^n - \Gamma_0 P \tilde{x}^n + g, \quad x^0 = \tilde{x}^0 = 0. \tag{3.9}$$

В силу изложенного выше, можно утверждать следующее.

Теорема 3.I. Для решения x^* уравнения с гладким оператором (при условиях (3.2)-(3.4)) и для его приближения по методу Ньютона x^n (см. (3.9)) справедливы оценки устойчивости § 2 с $\delta \geq \| \Gamma_c \check{f} - \Gamma_c \check{g} \| = \| f - g \|$, $\quad \varkappa = 2 \ell_0 L_p \gamma$,
$\tilde{\varkappa} = 2 \ell_0 L_p \tilde{\gamma}$, $\bar{\varkappa} = 2 \ell_0 L_p \bar{\gamma}$ $(\gamma \geq \|f\|, \ \bar{\gamma} = \gamma + \delta, \ \tilde{\gamma} \geq \|g\|)$.

§ 4. О возмущении уравнения с известной мажорантой

4.I. Определение 4.I. Мажорантой (обобщенной мажорантой) в шаре $\overline{Ш}(0,r)$ оператора $S: X \to X$ назовём такую непрерывную числовую функцию $\varphi: [0, \lambda] \to R^1$ $(\lambda \geq r)$ что

$$\| S 0 \| \leq \varphi(0), \tag{4.I}$$

$$\| x \| \leq p, \ \| y - x \| \leq q - p \implies \| S y - S x \| \leq \varphi(q) - \varphi(p) \quad (0 \leq p \leq q \leq r). \tag{4.2}$$

Условимся записывать это так: $S \prec \varphi$.

Пример 4.I. Если $K \in \Pi^{(2)}(W)$, $W \supset \overline{Ш}(0,r)$, $\|K\|_{(2)} \leq \beta$ и $f \in X$, $\|f\| \leq \gamma$ то для оператора $T: x \mapsto f - Kx$ в $\overline{Ш}(0,r)$ имеем $T \prec \varphi$, где $\varphi(t) := \beta t^2 + \gamma$ – это следует из леммы 2.I.

Утверждение I. I^0. Если $S \prec \varphi$ в $\overline{Ш}(0,r)$ то

$$\| S x \| \leq \varphi(\|x\|) \qquad (x \in \overline{Ш}(0,r)). \tag{4.3}$$

2^0

$$\left.\begin{array}{l} S_1 \prec \varphi_1, \ S_2 \prec \varphi_2 \\ \alpha \geq 0, \ \beta \geq 0 \end{array}\right\} \implies \alpha S_1 + \beta S_2 \prec (\alpha \varphi_1 + \beta \varphi_2). \tag{4.4}$$

Проверка осуществляется непосредственно. С идеями [I] связано

Утверждение 2. Если S дифференцируемо по Фреше в $W \supset \overline{Ш}(0,r)$ и φ дифференцируема на $[0, \lambda[$ $(\lambda > r)$ то условия (4.I), (4.2) равносильны условиям мажорирования Л.В.Канторовича [1].

4.2. Разрешимость операторного уравнения гарантирована, если разрешимо "мажорантное".[3] Рассмотрим уравнения

$$x = Sx \quad (S \prec \varphi \text{ в } \overline{Ш}(0,r)), \tag{4.5}$$

$$t = \varphi(t). \tag{4.6}$$

Мы считаем $\mathfrak{A} := \{t \in [0,r] : \varphi(t) = t\} \neq \emptyset$; $t^* := \min\limits_{t \in \mathfrak{A}} t$.

Положим $t_{n+1} = \varphi(t_n)$, $t_0 = 0$; $x^{n+1} = Sx^n$, $x^0 = 0$ $(n = 0, 1, 2 \ldots)$.

Утверждение 3. Последовательность $\{x^n\}$ мажорируется последовательностью $\{t_n\}$. При этом $\exists x^* \in \overline{Ш}(0,r)$: $\|x^* - x^n\| \le t^* - t_n$.

Элемент $x^* = \lim x_n$ есть решение уравнения (4.5).

Доказательство близко к рассуждениям п. 2.2.

4.3. Рассмотрим мажорируемое возмущение уравнения (4.5):

$$x = Sx + Rx \equiv \tilde{S}x, \quad R \prec \rho \text{ в } \overline{Ш}(0,r). \tag{4.5'}$$

Тем самым $\tilde{S} \prec \tilde{\varphi} := \varphi + \rho$. Предположим, что уравнение

$$t = \tilde{\varphi}(t) \equiv \varphi(t) + \rho(t) \tag{4.6'}$$

разрешимо в $[0,r]$. Из утверждения 3 следует тогда разрешимость уравнений (4.5') (в $\overline{Ш}(0,r)$), (4.6) (в $[0,r]$) и (4.5) (в $\overline{Ш}(0,r)$), поскольку $\varphi \prec \tilde{\varphi}$ в $[0,r]$. Пусть t^*, $\tilde{t}^*$ наименьшие корни (4.6) и (4.6') на $[0,r]$ и $\tilde{t}_{n+1} = \tilde{\varphi}(\tilde{t}_n)$, $\tilde{t}_0 = 0$. Рассмотрим решения x^*, $\tilde{x}^*$ определённые итерационными процессами

$$x^{n+1} = Sx^n, \; x^0 = 0 \; ; \quad \tilde{x}^{n+1} = \tilde{S}\tilde{x}^n, \; \tilde{x}^0 = 0. \tag{4.7}$$

Теорема 4.I. При сделанных предположениях справедливы оценки

$$\|\tilde{x}^n - x^n\| \le \tilde{t}_n - t_n, \tag{4.8}$$

$$\|\tilde{x}^* - x^*\| \le \tilde{t}^* - t^*. \tag{4.9}$$

Доказательство. I. По индукции убеждаемся, что

$$t_n \le \tilde{t}_n \tag{4.10}$$

Именно: $t_0 = 0 = \tilde{t}_0$, если же $t_{n-1} \le \tilde{t}_{n-1}$, то

$$t_n = \varphi(t_{n-1}) \le \varphi(\tilde{t}_{n-1}) \le \tilde{\varphi}(\tilde{t}_{n-1}) = \varphi(\tilde{t}_{n-1}) + \rho(\tilde{t}_{n-1}) = \tilde{t}_n.$$

2. При $n = 0$ имеем $\|\tilde{x}_0 - x_0\| = |\tilde{t}_0 - t_0| = 0$. Пусть уже известно, что $\|\tilde{x}^n - x^n\| \le \tilde{t}_n - t_n$. Одновременно будет

---

[3] Т.е. принцип мажоранты Л.В.Канторовича сохраняет силу (это можно показать и для теоремы единственности [1]).

$\|\tilde{x}^n\| \leq \tilde{t}_n$, $\|x^n\| \leq t_n$ ($\tilde{t}_n, t_n$ – мажорирующие последовательности). Но тогда по определению 4.I (см. (4.2))

$\|S\tilde{x}^n - Sx^n\| \leq \varphi(\tilde{t}_n) - \varphi(t_n)$. Отсюда

$$\|\tilde{x}^{n+1} - x^{n+1}\| = \|\tilde{S}\tilde{x}^n - Sx^n\| \leq \|R\tilde{x}^n\| + \|S\tilde{x}^n - Sx^n\| \leq$$
$$\leq \rho(\|\tilde{x}^n\|) + \|S\tilde{x}^n - Sx^n\| \leq \rho(\tilde{t}_n) + \varphi(\tilde{t}_n) - \varphi(t_n) =$$
$$= \tilde{\varphi}(\tilde{t}_n) - \varphi(t_n) = \tilde{t}_{n+1} - t_{n+1}.$$

Итак, (4.8) установлено индукцией по n; (4.9) вытекает из (4.8) предельным переходом.

В ситуации § 2 исходное уравнение имело вид $x = Tx := f - Kx$, а возмущенное можно записать в виде

$$x = Tx + Rx := (f - Kx) + \{(\tilde{K} - K)x + (\tilde{f} - f)\}.$$

Теперь легко убедиться, что теорема 2.2 является частным случаем теоремы 4.I (см. пример 4.I).

Пользуесь теоремой 4.I можно исследовать и устойчивость обобщенного модифицированного процесса Ньютона (см. § 3).

$$x^{n+1} = x^n - \Gamma P x^n + f,$$

где $\Gamma \approx \Gamma_0$ (подробнее см. [I]). Здесь можно допустить и возмущение оператора P, однако (в рамках рассматриваемого нами подхода к этим проблемам) – при ограничивающих условиях малости $\|(P-\tilde{P})(0)\|$, $\|(P'-\tilde{P}')(0)\|$, $\|P''-\tilde{P}''\|$. Высказанные в § 4 соображения будут в частности применимы к уравнениям с параметром μ, типа рассмотренных в [I]:

$$Px + \mu Qx = f;$$

именно, оценки типа (4.9) позволяют здесь оценить

$$\|x^*_{\mu_0} - x^*_{\tilde{\mu}}\| \quad (\tilde{\mu} \approx \mu_0).$$

Автор выражает благодарность Г.М.Вайникко за руководство работой.

Литература

1. Канторович Л.В., Акилов Г.И., Функциональный анализ. Москва, 1977.
2. Красносельский М.А., Вайникко Г.М. и др., Приближенное решение операторных уравнений. Москва, 1969.

3. Ш и ф р и н Б.Ф., Метод итераций и условия разрешимости для уравнений типа Навье-Стокса. Изв. АН Эст. ССР, Физ., мат., 1982, 31, № 3, 249-260.

Поступило
10 III 1982

STABILITY OF THE SOLUTION OF THE OPERATOR EQUATION WITH SECOND-ORDER NONLINEARITY

B. Shifrin

Summary

The nonlinear problem $x + Kx = f$ (see (I), (1.1)) proved to be well-posed under some conditions. We estimate the perturbation of the solution, caused by the change of f and K. The estimates of the stability of Newton's method for equation $Px = f$ are also presented. Results of such type can be obtained for the general operator equation provided its majorant (see [1]) is known.

Уч.зап.Тартуск.ун-та,1982,633,88-94

О МЕТОДЕ НЬЮТОНА В СЛУЧАЕ КРАТНЫХ РЕШЕНИЙ

П. Оя

1. Общеизвестно, что метод Ньютона имеет высокую скорость сходимости даже для уравнений в банаховом пространстве, если решение однократное. В случае кратных решений численных уравнений также хорошо известно, что сходимость линейна и определяется кратностью решения. Мы предлагаем правило определения кратности по приближениям Ньютона. Характер сходимости метода Ньютона к кратным решениям нелинейных систем изучен в [2-5], этот вопрос затронут и в настоящей работе.

2. В этом пункте приведем некоторые известные факты о сходимости метода Ньютона для численных уравнений и предложим одну возможность вычисления кратности решения.

Для решения численного уравнения

$$f(x)=0 \tag{1}$$

рассмотрим метод Ньютона

$$x_{n+1}=x_n-\frac{f(x_n)}{f'(x_n)}\ ,\quad n=0,1,\dots\ . \tag{2}$$

Предположим, что уравнение (1) имеет изолированное решение x^*, причем $f(x^*)=f'(x^*)=\dots=f^{(m-1)}(x^*)=0$, $f^{(m)}(x^*)\neq 0$ для некоторого $m\geq 1$. Число m назовем кратностью решения x^*. Все встречающиеся при изложении производные функции f считаем непрерывными в окрестности x^*.

Разлагая $f(x_n)$ и $f'(x_n)$ по формуле Тейлора в точке x^* и обозначая $\varepsilon_n=x_n-x^*$, получим

$$f(x_n)=\frac{f^{(m)}(\xi_n)}{m!}\,\varepsilon_n^m\ ,\quad f'(x_n)=\frac{f^{(m)}(\eta_n)}{(m-1)!}\,\varepsilon_n^{m-1}\ ,$$

где $\xi_n,\eta_n\in(x^*,x_n)$. Тогда из (2) следует

$$\varepsilon_{n+1}=\left(1-\frac{1}{m}\right)\varepsilon_n+\frac{1}{m}\left(1-\frac{f^{(m)}(\xi_n)}{f^{(m)}(\eta_n)}\right)\varepsilon_n\ . \tag{3}$$

Если предполагать, что $\varepsilon_n\to 0$, т.е. метод Ньютона сходится, то $\xi_n,\eta_n\to x^*$ и $f^{(m)}(\xi_n), f^{(m)}(\eta_n)\to f^{(m)}(x^*)$. В случае однократного решения $x^*(m=1)$ получим из (3), что $\varepsilon_{n+1}=o(\varepsilon_n)$, т.е. метод сходится сверхлинейно (быстрее любой геометрической прогрессии), а если f' удовлетворяет условию Липшица, то из (3) вытекает квадратичная сходимость. В случае кратного решения ($m \geq 2$) на основе (3) заключаем, что метод Ньютона сходится асимптотически со скоростью геометрической прогрессии со знаменателем $1-\frac{1}{m}$, если только функция $f^{(m)}$ непрерывна.

Такие же выкладки для метода Ньютона-Шрёдера

$$x_{n+1} = x_n - m\,\frac{f(x_n)}{f'(x_n)}$$

показывают, что непрерывность функции $f^{(m)}$ гарантирует сверхлинейную сходимость, а если $f^{(m)}$ удовлетворяет условию Липшица, то сходимость квадратичная.

Введем обозначения

$$m_n = \frac{1}{1-\dfrac{x_{n+1}-x_n}{x_n-x_{n-1}}}, \qquad \alpha_n = \frac{f^{(m)}(\xi_n)}{f^{(m)}(\eta_n)},$$

где x_n — последовательность приближений, полученных по методу Ньютона, а $f^{(m)}(\xi_n)$ и $f^{(m)}(\eta_n)$, как и раньше, получены при разложении $f(x_n)$ и $f'(x_n)$ по формуле Тейлора. Опираясь на равносильное к (3) соотношение $x_{n+1}-x_n=-\frac{1}{m}\alpha_n$, получаем при помощи непосредственных вычислений равенство

$$m_n - m = \frac{m\left(\dfrac{1}{\alpha_n}-1\right)+m^2\left(\dfrac{1}{\alpha_{n-1}}-\dfrac{1}{\alpha_n}\right)}{1+m\left(\dfrac{1}{\alpha_n}-\dfrac{1}{\alpha_{n-1}}\right)}.$$

Поскольку $\alpha_n \to 1$, то $m_n \to m$. Кроме того, числитель правой части последнего равенства оценивается асимптотически через величину $\frac{(m+3m^2)L}{|f^{(m)}(x^*)|}|\varepsilon_n|$, где L — постоянная Липшица для $f^{(m)}$, и, таким образом, скорость сходимости $m_n \to m$ не уступает скорости сходимости $x_n \to x^*$. Рассмотрим иллюстрирующий

Пример 1. Пусть $f(x)=x(x-2)^4(x^2-6x+10)$. Приведем некоторые промежуточные приближения, найденные по методу Ньютона, исходя из $x_0=2,5$:

$n-1$	x_{n-1}	x_n	x_{n+1}	m_n
0	2,5	2,368421	2,272048	3,737361
4	2,149830	2,111626	2,083309	3,863969
8	2,046566	2,034855	2,026103	3,958298
12	2,014654	2,010984	2,008234	3,987083
16	2,004629	2,003471	2,002603	3,995989
20	2,001464	2,001098	2,000823	3,999349

Из таблицы видно, что $\varepsilon_{n+1} \approx \frac{3}{4}\varepsilon_n$ и $|m_n - m|$ убывает с такой же скоростью, что и $|x_n - x^*|$. Отметим, что кратность решения обнаруживается через несколько шагов метода Ньютона, после этого можно перейти на метод Ньютона - Шрёдера.

3. Рассмотрим систему уравнений

$$F(x) = 0, \tag{4}$$

где $F: \mathbb{R}^n \to \mathbb{R}^n$. Поставим вопрос о перенесении результатов предыдущего пункта на системы (4).

Запишем систему метода Ньютона в виде

$$F'(X_n)(\varepsilon_{n+1} - \varepsilon_n) = -F(X_n), \tag{5}$$

где $\varepsilon_n = X_n - X^*$, X^* — решение системы (4). Хорошо известно (см., напр., [1]), что если матрица $F'(X^*)$ регулярна, то непрерывность F' влечет сверхлинейную сходимость, а условие Липшица для F' — квадратичную сходимость метода Ньютона. В дальнейшем рассмотрим случай, когда $F'(X^*)$ сингулярна. Используемые производные от F считаем непрерывными. Предположим также (это необходимо для применимости метода Ньютона), что для некоторого $r > 0$

$$\det F'(x) \neq 0 \quad \text{при} \quad 0 < \|X - X^*\| < r. \tag{6}$$

Теорема. Пусть

$$F'(X^*) = 0, \ldots, F^{(m-1)}(X^*) = 0 \tag{7}$$

и

$$\det F^{(m)}(X^*)\xi^{m-1} \neq 0 \quad \forall \xi \in \mathbb{R}^n \setminus \{0\}. \tag{8}$$

Тогда при достаточно хорошем начальном приближении сходимость метода Ньютона асимптотически линейна со знаменателем $1 - \frac{1}{m}$.

Доказательство. Учитывая (7), имеем разложения Тейлора

$$F(X_n) = \frac{1}{m!} F^{(m)}(X^*)\varepsilon_n^m + \alpha(X^*; \varepsilon_n),$$

$$F'(X_n) = \frac{1}{(m-1)!} F^{(m)}(X^*)\varepsilon_n^{m-1} + \beta(X^*; \varepsilon_n),$$

где $\dfrac{\|\alpha(X^*; \varepsilon)\|}{\|\varepsilon\|^m} \to 0$, $\dfrac{\|\beta(X^*; \varepsilon)\|}{\|\varepsilon\|^{m-1}} \to 0$ при $\|\varepsilon\| \to 0$.

Поскольку матрица $F^{(m)}(X^*)\varepsilon_n^{m-1}$ регулярна, то система Ньютона (5) приобретает вид

$$\left(I + (m-1)!\left(F^{(m)}(X^*)\varepsilon_n^{m-1}\right)^{-1}\beta(X^*; \varepsilon_n)\right)\varepsilon_{n+1} =$$

$$= \left(1 - \frac{1}{m}\right)\varepsilon_n + (m-1)!\left(F^{(m)}(X^*)\varepsilon_n^{m-1}\right)^{-1}\left(\beta(X^*; \varepsilon_n)\varepsilon_n - \alpha(X^*; \varepsilon_n)\right). \tag{9}$$

Имеет место $\inf\limits_{\|\xi\|=\|X\|=1} \|F^{(m)}(x^*)\,\xi^{m-1}X\| = \gamma > 0$.
Действительно, в случае $\gamma=0$, поскольку непрерывная функция $\|F^{(m)}(x^*)\,\xi^{m-1}X\|$ достигает нижней грани на компакте $\{\xi, X\in\mathbb{R}^n : \|\xi\|=\|X\|=1\}$, существуют ξ_0 и X_0, $\|\xi_0\|=\|X_0\|=1$, такие, что $F^{(m)}(x^*)\,\xi_0^{m-1}X_0=0$. Поскольку $X_0\neq 0$, матрица $F^{(m)}(x^*)\,\xi_0^{m-1}$ сингулярна, а это противоречит предположению (8). Далее,

$$\|F^{(m)}(x^*)\varepsilon_n^{m-1}X\| = \|\varepsilon_n\|^{m-1}\|X\|\,\|F^{(m)}(x^*)\left(\frac{\varepsilon_n}{\|\varepsilon_n\|}\right)^{m-1}\frac{X}{\|X\|}\| \geqslant$$

$$\geqslant \|\varepsilon_n\|^{m-1}\|X\|\,\gamma,$$

следовательно,

$$\|(F^{(m)}(x^*)\,\varepsilon_n^{m-1})^{-1}\| \leqslant \frac{1}{\|\varepsilon_n\|^{m-1}\gamma}.$$

Теперь ясно, что (9) влечет соотношение $\varepsilon_{n+1}\approx\left(1-\frac{1}{m}\right)\varepsilon_n$, знак $\approx$ здесь означает, что опущены бесконечно малые высшего порядка.

Теорема доказана.

Отметим, что условия (7) и (8) обеспечивают выполненность условия (6).

4. Поставим вопрос о существенности условий (7) и (8).

Дадим три определения кратности решения. Число m назовем кратностью решения X^* системы (4), если

1) $\bigcap\limits_{k=1}^{m-1} \mathrm{Ker}\, F^{(k)}(x^*)\neq\{0\}$, $\bigcap\limits_{k=1}^{m} \mathrm{Ker}\, F^{(k)}(x^*)=\{0\}$; (10)

2) $F'(x^*)=0, \dots, F^{(m-1)}(x^*)=0$, $\mathrm{Ker}\, F^{(m)}(x^*)=\{0\}$; (11)

3) выполнены условия (7) и (8).

Понятно, что из (7) и (8) следует (11) и из (11) условие (10). При $n=1$ все три определения совпадают с приведенным в пункте 2, но при $n\geqslant 2$ они различные.

Пример 2. Задана система

$$x_1x_2=0$$
$$-\frac{1}{3}|x_1|^3+\frac{1}{2}x_2^2=0.$$

Система имеет изолированное решение $X^*=(0,0)$, причем

$$F'(X)=\begin{pmatrix} x_2 & x_1\\ -x_1^2\,\mathrm{sgn}\,x_1 & x_2\end{pmatrix}$$

и $\det F'(X)=x_2^2+|x_1|^3=0$ тогда и только тогда, когда $x_1=0, x_2=0$. Условие (6) выполнено. Для $\xi=(\xi_1,\xi_2)$ имеем

$$F''(X^*)\xi = \begin{pmatrix} \xi_2 & \xi_1 \\ 0 & \xi_2 \end{pmatrix},$$

значит, $\operatorname{Ker} F''(X^*) = \{0\}$. Поскольку $F'(X^*)=0$, то решение X^* является двукратным в смысле (11) или (10), но условие (8) не выполнено. Если взять $\varepsilon_n = (\varepsilon, 0)$ $(\varepsilon > 0)$, то, решая соответствующую систему Ньютона, получаем, что $\varepsilon_{n+1} = \frac{2}{3}\varepsilon_n$. В то же время, если $\varepsilon_n = (0, \varepsilon)$, то $\varepsilon_{n+1} = \frac{1}{2}\varepsilon_n$.

Приведенный пример показывает, что кратность m решения в смысле (10) или (11) не гарантирует асимптотически линейной сходимости со знаменателем $1-\frac{1}{m}$. Кроме того, скорость сходимости может зависеть от начального приближения.

Укажем на одну возможность обобщения примера 2. Взяв вместо второго уравнения $-\frac{1}{k} x_1^{k-1}|x_1| + \frac{1}{2} x_2^2 = 0$, где $k \geqslant 4$, получим более гладкие функции F, а скорость сходимости при $\varepsilon_n = (\varepsilon, 0)$ замедляется: $\varepsilon_{n+1} = (1-\frac{1}{k})\varepsilon_n$.

Пример 3. Рассмотрим систему

$$\begin{aligned} 2x_1^3 + x_1^2 x_2 + 2x_1 x_2^2 + x_2^3 + x_1 + \alpha x_2 = 0 \\ x_1^3 + x_1^2 x_2 + x_1 x_2^2 + x_2^3 + x_1 + \alpha x_2 = 0, \end{aligned} \tag{12}$$

имеющую изолированное решение $X^* = (0,0)$ при $\alpha \geqslant 0$. Здесь

$$F'(X) = \begin{pmatrix} 6x_1^2 + 2x_1x_2 + 2x_2^2 + 1 & x_1^2 + 4x_1x_2 + 3x_2^2 + \alpha \\ 3x_1^2 + 2x_1x_2 + x_2^2 + 1 & x_1^2 + 2x_1x_2 + 3x_2^2 + \alpha \end{pmatrix},$$

и, таким образом,

$$\det F'(X) = 3(x_1^2 + x_2^2)^2 + 3\alpha x_1^2 - 2x_1x_2 + \alpha x_2^2 \geqslant$$

$$\geqslant 3(x_1^2 + x_2^2)^2 + (x_1 - x_2)^2 \quad \text{при} \quad \alpha \geqslant 1.$$

Значит, $\det F'(X) = 0$ тогда и только тогда, когда $x_1 = 0$ и $x_2 = 0$. Имеем $F''(X^*) = 0$ и

$$F'''(X)\xi^2 = \begin{pmatrix} 12\xi_1^2 + 4\xi_1\xi_2 + 4\xi_2^2 & 2\xi_1^2 + 8\xi_1\xi_2 + 6\xi_2^2 \\ 6\xi_1^2 + 4\xi_1\xi_2 + 2\xi_2^2 & 2\xi_1^2 + 4\xi_1\xi_2 + 6\xi_2^2 \end{pmatrix}.$$

Из последнего вычислим $\det F'''(X)\xi^2 = 12(\xi_1^2 + \xi_2^2)^2$, следовательно, условие (8) удовлетворяется для $m=3$. В то же время $F'(X^*) \neq 0$ и условие (7) не выполнено. Если $\varepsilon_n = (0, \varepsilon)$, то $\varepsilon_{n+1} = \left(0, \frac{2\varepsilon^3}{3\varepsilon^2 + \alpha}\right)$. Отсюда видно, что тогда $\|\varepsilon_{n+1}\| \approx \approx \frac{2}{\alpha}\|\varepsilon_n\|^3$, т.е. сходимость является кубической.

Пример 3 показывает, что в теореме нельзя отказаться от условия (7), не сделав дополнительных предположений.

Замечание. В рассуждениях работы [4], касающихся сходимости метода Ньютона для систем уравнений, имеются неточности, и утверждения [4] без дополнительных предположений не имеют места. Так, например, для системы (12) при $\varepsilon_n = (0, \varepsilon)$ по приведенным в [4] утверждениям должно быть $\varepsilon_{n+1} = = \frac{1}{3}(-\varepsilon, \varepsilon) + O(\|\varepsilon_n\|^2)$, но это противоречит действительности. Выкладки в [4] проходят при (сильных и неестественных) ограничениях $F'(X_n)N_1 \subset N_1$ и $F'(X_n)R_1 \subset R_1$, где $N_1 = \mathcal{K}er\ F'(X^*)$, $F'(X^*)R_1 = R_1$ и $N_1 \oplus R_1 = \mathbb{R}^n$. Конечно, система (12) при $\varepsilon_n = (0, \varepsilon)$ им не удовлетворяет. Отметим еще, что из утверждений [4] даже при указанных ограничениях не явствует скорость сходимости метода Ньютона для систем уравнений.

5. Функция $\varphi(\xi) = \det F^{(m)}(X^*)\xi^{m-1}$ является $n(m-1)$-линейной от n переменных $\xi_1, \ldots, \xi_n$. Так как $\varphi(-\xi) = = (-1)^{n(m-1)}\varphi(\xi)$, то условие (8) не может быть выполнено, если число $n(m-1)$ нечетное (например, $n=3, m=2$). Действительно, если ξ пробегает сферу $\{\xi : \|\xi\| = 1\}$, то непрерывная функция φ принимает все значения между $\varphi(\xi)$ и $\varphi(-\xi)$, в том числе значение 0, при нечетном $n(m-1)$.

6. Рассмотрим вместо (5) систему метода Ньютона—Шрёдера

$$F'(X_n)(\varepsilon_{n+1} - \varepsilon_n) = -m\,F(X_n).$$

В предположениях (7) и (8) вместо (9) получаем

$$\left(I + (m-1)!\left(F^{(m)}(X^*)\varepsilon_n^{m-1}\right)^{-1}\beta(X^*; \varepsilon_n)\right)\varepsilon_{n+1} =$$
$$= (m-1)!\left(F^{(m)}(X^*)\varepsilon_n^{m-1}\right)^{-1}\left(\beta(X^*; \varepsilon_n)\varepsilon_n - m\,\alpha(X^*; \varepsilon_n)\right).$$

Отсюда видно, что уже в предположениях теоремы метод Ньютона—Шрёдера сходится со сверхлинейной скоростью. Если, например, существует непрерывная в окрестности X^* производная $F^{(m+1)}$, то $\|\alpha(X^*; \varepsilon_n)\| = O(\|\varepsilon_n\|^{m+1})$ и $\|\beta(X^*; \varepsilon_n)\| = = O(\|\varepsilon_n\|^m)$, значит, сходимость квадратичная. Кратность m решения X^* при условиях (7) и (8) определяется по приближениям метода Ньютона X_n как и в пункте 2, используя здесь компоненты X_n.

Литература

1. К р а с н о с е л ь с к и й М.А., В а й н и к к о Г.М., З а б р е й к о П.П., Р у т и ц к и й Я.Б., С т е ц е н к о В.Я., Приближенное решение операторных уравнений. Москва, 1969.
2. D e c k e r, D.W., K e l l e y, C.T., Newton's method at singular points. I. SIAM J. Numer. Anal., 1980, 17, № 1, 66-70.
3. D e c k e r, D.W., K e l l e y, C.T., Newton's method at singular points. II. SIAM J. Numer. Anal., 1980, 17, № 3, 465-471.
4. R a l l, L.B., Convergence of the Newton process to multiple solutions. Numer. Math., 1966, 9, 23-37.
5. R e d d i e n, G.W., On Newton's method for singular problems. SIAM J. Numer. Anal., 1978, 15, № 5, 993-996.

Поступило
30 III 1982

ON NEWTON'S METHOD AT SINGULAR POINTS

P. Oja

Summary

Let x^* be a solution of $f(x)=0$ such that $f(x^*)=f'(x^*)=\ldots=f^{(m-1)}(x^*)=0$, $f^{(m)}(x^*)\neq 0$. It is shown that by means of the Newton's iterations x_n it is possible to determine the multiplicity m: the sequence $m_n=1/(1-(x_{n+1}-x_n)/(x_n-x_{n-1}))$ converges to m. The rapidity of this convergence is characterized by the estimate $|m_n-m|\leq const\,|x_n-x^*|$. For finite systems $F(X)=0$, $F:\mathbb{R}^n\to\mathbb{R}^n$, we establish

Theorem. Let $F(X^*)=0, F'(X^*)=0,\ldots, F^{(m-1)}(X^*)=0$ and $\det F^{(m)}(X^*)X^{m-1}\neq 0$ $\forall X\in\mathbb{R}^n\setminus\{0\}$. Then the convergence of the Newton's method starting from a sufficiently small neighborhood of X^* is asymptotically linear with the common ratio $1-\frac{1}{m}$.

Two examples show that the assumptions of the theorem are essential for obtaining such a convergence.

СОДЕРЖАНИЕ

CONTENTS INHALT