ISSN 0494-7304 0132-053x

TARTU ÜLIKOOLI
TOIMETISED

УЧЕНЫЕ ЗАПИСКИ ТАРТУСКОГО УНИВЕРСИТЕТА

ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS

937

METHODS FOR SOLUTION OF INTEGRAL EQUATIONS AND ILL-POSED PROBLEMS

Matemaatika- ja mehhaanika-alaseid töid

Труды по математике и механике

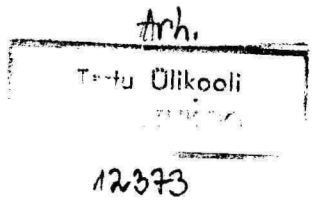
TARTU 1992

TARTU ÜLIKOOLI TOIMETISED
УЧЕНЫЕ ЗАПИСКИ ТАРТУСКОГО УНИВЕРСИТЕТА
ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS
Alustatud 1893.a. VIHIK 937 ВЫПУСК Основаны в 1893.г.

METHODS FOR SOLUTION OF INTEGRAL EQUATIONS AND ILL-POSED PROBLEMS

Matemaatika- ja mehhaanika-alaseid töid
Труды по математике и механике

TARTU 1992



Acta et commentationes Universitatis Tartuensis, 937, 1992, 3-14.

SOLUTION OF LARGE SYSTEMS ARISING BY DISCRETIZATION OF MULTIDIMENSIONAL WEAKLY SINGULAR INTEGRAL EQUATIONS

Gennadi Vainikko

Discretizing a linear integral equation on a bounded region $G \subset \mathbb{R}^n$, one obtains a linear system of equations of order $l_h \asymp h^{-n}$. A direct solution of such systems is possible only in case of rough discretizations with $h = h_*$ not too small. Using two grid iteration methods, it is possible to solve those systems for much finer discretizations with $h << h_*$. Thereby a significant economy of computing time can be achieved: instead of $O(l_h^3)$ arithmetical operations on Gauss type direct methods, an approximate solution of a suitable accuracy can be found in $O(l_h^2)$ operations, in some cases even in $O(l_h \log l_h)$ operations.

1. **Integral equation.** In this paper, we shall deal with an integral equation

$$u(x) = \int_G K(x,y)u(y)dy + f(x), \qquad x \in G, \tag{1}$$

where $G \subset \mathbb{R}^n$ is an open bounded set with a piecewise smooth boundary ∂G. The following assumptions (A1)- (A4) are made.

(A1) Kernel $K(x,y)$ is twice continuously differentiable on $(G \times G) \setminus \{x=y\}$ and there exsists a real number ν ($\nu < n$) such that for any $x, y \in G$, $x \neq y$, and any multi-indeces $\alpha = (\alpha_1, \dots, \alpha_n)$ and $\beta = (\beta_1, \dots, \beta_n)$ with $|\alpha| + |\beta| \le 2$,

$$\left| \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \dots \left(\frac{\partial}{\partial x_n}\right)^{\alpha_n} \left(\frac{\partial}{\partial x_1} + \frac{\partial}{\partial y_1}\right)^{\beta_1} \dots \left(\frac{\partial}{\partial x_n} + \frac{\partial}{\partial y_n}\right)^{\beta_n} K(x,y) \right| \le$$
$$\le b \begin{cases} 1, & \nu + |\alpha| < 0 \\ 1 + |\log |x-y||, & \nu + |\alpha| = 0 \\ |x-y|^{-\nu-|\alpha|}, & \nu + |\alpha| > 0 \end{cases}, \quad b = \text{const}. \tag{2}$$

Here the following usual conventions are adopted:

$$|\alpha| = \alpha_1 + \dots + \alpha_n \quad \text{for} \quad \alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}_+^n,$$
$$|x| = (x_1^2 + \dots + x_n^2)^{1/2} \quad \text{for} \quad x = (x_1, \dots, x_n) \in \mathbb{R}^n.$$

(A2) The homogeneous integral equation corresponding to (1) has in L(G) only the trivial solution.

(A3) $f \in C^{2,\nu}(G)$, i.e. f is twice continuously differentiable on G and, for any $x \in G$ and any multiindex $\alpha \in \mathbb{N}_+^n$ with $|\alpha| \le 2$,

$$|D^\alpha f(x)| \le c_f \begin{cases} 1, & |\alpha| < n-\nu \\ 1+|\log \varrho(x)|, & |\alpha| = n-\nu \\ \varrho(x)^{n-\nu-|\alpha|}, & |\alpha| > n-\nu \end{cases}, \quad c_f = \text{const},$$

where $\varrho(x) = \inf_{y \in \partial G} |x-y|$ is the distance from x to ∂G.

(A4) For any $x^1, x^2 \in G$,

$$|f(x^1) - f(x^2)| \le c_f' \begin{cases} d_G(x^1,x^2), & \nu < n-1 \\ d_G(x^1,x^2)[\,1+|\log d_G(x^1,x^2)|\,], & \nu = n-1 \\ d_G(x^1,x^2)^{n-\nu}, & \nu > n-1 \end{cases}$$

where $d_G(x^1,x^2)$ is defined as the infimum of lengths of polygonal paths in G joining points x^1 and x^2; if x^1 and x^2 belong to different connectivity components of G, define $d_G(x^1,x^2) = \infty$.

Note that kernels $K(x,y) = a(x,y)|x-y|^{-\nu}$ $(0 < \nu < n)$ and $K(x,y) = a(x,y)\log|x-y|$ $(\nu = 0)$ satisfy (A1) if a(x,y) is twice continuously differentiable on $(G \times G) \setminus \{x=y\}$ and its derivatives are bounded or, more generally, e.g. in case $0 < \nu < n$,

$$\left| \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \dots \left(\frac{\partial}{\partial x_n}\right)^{\alpha_n} \left(\frac{\partial}{\partial x_1} + \frac{\partial}{\partial y_1}\right)^{\beta_1} \dots \left(\frac{\partial}{\partial x_n} + \frac{\partial}{\partial y_n}\right)^{\beta_n} a(x,y) \right| \le$$

$$\le b' |x-y|^{-|\alpha|} \qquad (|\alpha| + |\beta| \le 2), \qquad b' = \text{const}.$$

A further example of a kernel satisfying (A1) is originated from the radiation transfer theory and is known as Peierls kernel

$$K(x,y) = \frac{1}{4\pi} e^{-\tau(x,y)} |x-y|^{-2} \sigma_s(y) \qquad (n=3,\ \nu=2)$$

where

$$\tau(x,y) = |x-y| \int_0^1 \sigma(tx + (1-t)y)\,dt$$

is the optical distance between points $x, y \in G$ (set G is assumed to be convex in this example); extinction coefficient $\sigma : \bar{G} \to \mathbb{R}$ and scattering coefficient $\sigma_s : \bar{G} \to \mathbb{R}$ are assumed to be twice continuously differentiable.

A more general example of a kernel satisfying (A1) is given by

$$K(x,y) = \varkappa(x,y,|x-y|)$$

where $\varkappa : G \times G \times \mathbb{R}_+ \to \mathbb{R}$ is a twice continuously differentiable function such that, for $|\alpha| + |\beta| + k \le 2$,

$$|D_x^\alpha\ D_y^\beta\ \frac{\partial^k}{\partial r^k}\ \varkappa(x,y,r)| \le b'' r^{\nu-k} \quad , \qquad 0 < \nu < n, \quad b'' = \text{const}.$$

2. Subdivisions of G. Let us denote, for a set $G' \subset G$,

$$d_G\text{-}\operatorname{diam} G' = \sup_{x,y \in G'} d_G(x,y).$$

For a $h>0$, introduce an "approximate subdivision" of G into measurable sets ("cells") $G_{jh} \subset \mathbb{R}^n$ $(j=1, \dots, l_h)$ such that

$$G_{jh} \cap G \neq \Phi, \quad G_{ih} \cap G_{jh} = \Phi \ (i \neq j)$$
$$\operatorname{diam} G_{jh} \leq h, \quad d_G\text{-}\operatorname{diam}(G_{jh} \cap G) \leq c_1 h \ (j=1, \dots, l_h),$$
$$(\bar{G} \setminus \bar{G}_h) \cup (\bar{G}_h \setminus \bar{G}) \subset \{x \in \mathbb{R}^n : \varrho(x) \leq c_2 h^2\}$$

where

$$G_h = \bigcup_{j=1}^{l_h} G_{jh}$$

and the constants c_1 and c_2 do not depend on h. Choose points $\xi_{jh} \in G_{jh} \cap G$ $(j=1, \dots, l_n)$ as follows:

$\xi_{jh} = (\operatorname{mes} G_{jh})^{-1} \int_{G_{jh}} y\,dy$ (=centroid of G_{jh}) if $\operatorname{dist}(\operatorname{co} G_{jh}, \partial G) \geq h$,

$\xi_{jh} \in G_{jh} \cap G$ is an arbitrary point if $\operatorname{dist}(\operatorname{co} G_{jh}, \partial G) < h$.

Here coA is the convex hull of a set $A \subset \mathbb{R}^n$ and $\operatorname{dist}(A_1, A_2) = \inf_{x^1 \in A_1, x^2 \in A_2} |x^1 - x^2|$. Now we can introduce a cubature formula:

$$\int_G u(y)dy \approx \sum_{j=1}^{l_h} w_{jh}\, u(\xi_{jh}), \qquad w_{jh} = \operatorname{mes} G_{jh}. \tag{3}$$

3. Discretization of the integral equation. We introduce the following three discretizations of equation (1):

$$u_{ih} = \sum_{j=1}^{l_h} t_{ijh}\, u_{jh} + f(\xi_{ih}) \quad (i=1,\dots,l_h), \quad t_{ijh} = \int_{G_{jh}} K(\xi_{ih}, y)dy; \tag{4}$$

$$u_{ih} = \sum_{j=1}^{l_h} t'_{ijh}\, u_{jh} + f(\xi_{ih}) \quad (i=1,\dots,l_h), \quad t'_{ijh} = \begin{cases} K(\xi_{ih}, \xi_{jh}) w_{jh}, & i \neq j \\ 0, & i=j; \end{cases} \tag{4'}$$

$$u_{ih} = \sum_{j=1}^{l_h} t''_{ijh}\, u_{jh} + f(\xi_{ih}) \quad (i=1,\dots,l_h),$$

$$t''_{ijh} = \begin{cases} K(\xi_{ih}, \xi_{jh}) w_{jh}, & i \neq j \\ \displaystyle\int_{G_h} K(\xi_{ih}, y)dy - \sum_{\substack{k=1 \\ k \neq i}}^{l_h} K(\xi_{ih}, \xi_{kh}) w_{kh}, & i=j. \end{cases} \tag{4''}$$

In methods (4) and (4") it may happen that the kernel K(x,y) must be integrated over $G_{jh} \not\subset G$ or $G_h \equiv \bigcup_{1 \leq j \leq l_h} G_{jh} \not\subset G$. We assume that

$K(x,y)$ is extended with respect to y on G_h so that estimate (2) remains valid for $|\alpha|=|\beta|=0$:

$$|K(x,y)| \le b \begin{cases} 1 & , \ \nu<0 \\ 1+|\log|x-y||, & \nu=0 \\ |x-y|^{-\nu} & , \ \nu>0 \end{cases}, \quad (x\in G, y\in G_h).$$

System (4) corresponds to a collocation method: denoting by φ_{jh} the characteristic function of G_{jh}, approximating the solution u of equation (1) by a piecewise constant function

$$\bar{u}_h = \sum_{j=1}^{l_h} u_{jh}\varphi_{jh}$$

and collocating the equation in the points ξ_{ih} $(i=1, \dots, l_h)$ with G_h as the domain of integration instead of G, we obtain system (4).

System (4') corresponds to a classical cubature formula method: approximating the integral in (1) by means of cubature formula (3) and collocating in the points ξ_{ih} $(i=1, \dots, l_h)$ we obtain system (4') if we reject the terms where $i=j$ (the kernel $K(x,y)$ is not defined for $x=y$). On the other hand, (4') may be considered as an approximation to (4):

$$t_{ijh} = \int_{G_{jh}} K(\xi_{ih},y)dy \approx K(\xi_{ih},\xi_{jh})w_{jh} = t'_{ijh} \quad (i\ne j).$$

System (4") corresponds to the Kantorovich-Krylov modification of the cubature formula method.

4. <u>Discretization error</u>. Methods (4) and (4') are investigated in [6,7] and method (4") in [4]. Let us present the main results of these works.

<u>Theorem</u> 1. Let assumptions (A1)-(A4) be fulfilled and let the subdivision of G and the choice of collocation points ξ_{jh} satisfy the conditions of Section 2. Then there exists a $h_0>0$ such that, for $0<h<h_0$, system (4) has a unique solution (u_{ih}), and

$$\max_{1\le i\le l_h} |u_{ih} - u(\xi_{ih})| \le \text{const}\,(\varepsilon_{\nu h})^2, \qquad \varepsilon_{\nu h} = \begin{cases} h & , \ \nu<n-1, \\ h(1+|\log h|) & , \ \nu=n-1, \\ h^{n-\nu} & , \ \nu>n-1, \end{cases} \quad (5)$$

where u is the (unique) solution to (1), $u\in C^{2,\nu}(G)$.

<u>Theorem</u> 2. Let the conditions of Theorem 1 be fulfilled and let

$$|\xi_{ih} - \xi_{jh}| \ge c_0 h \quad (i\ne j) \qquad (6)$$

with a constant $c_0>0$ not depending on h. Then there exists a $h'_0>0$ such that, for $0<h<h'_0$, system (4') has a unique solution (u_{ih}), and

$$\max_{1\le i\le l_h} |u_{ih}-u(\xi_{ih})| \le \text{const}\ \varepsilon'_{\nu h}\ , \quad \varepsilon'_{\nu h}=\begin{cases} h^2 & ,\ \nu<n-2,\\ h^2(1+|\log h|) & ,\ \nu=n-2,\\ h^{n-\nu} & ,\ \nu>n-2.\end{cases}$$

Theorem 3. Let the conditions of Theorem 2 be fulfilled and let

$$d_G(x^1,x^2) \le \text{const}\,|x^1-x^2| \quad (x^1,x^2\in G). \tag{7}$$

Then there exists a $h_0''>0$ such that , for $0<h<h_0''$, system (4") has a unique solution (u_{ih}), and error estimate (5) holds.

From Theorems 1-3 we see that an accuracy $O(\varepsilon^p_{\nu h})$, $p\ge 2$, is sufficient if we solve systems (4)-(4") approximately.

5. Two grid iterations. Let $h<<h_*$. Introduce approximate subdivisions of G into cells G_{jh} $(j=1,\dots,l_h)$ and $G_{j'h_*}$ $(j'=1,\dots,l_{h_*})$ and corresponding collocation points $\xi_{jh}\in G_{jh}\cap G$ and $\xi_{j'h_*}\in G_{j'h_*}\cap G$ as in Section 2. For simplicity, let the following compatiblity conditions be fulfilled: (i) every cell G_{jh} $(j=1,\dots l_h)$ is contained in a cell ("panel") $G_{j'h_*}$ $(1\le j'\le l_{h_*})$ and , conversely, every panel $G_{j'h_*}$ $(j'=1,\dots,l_{h_*})$ is a union of some cells G_{jh} $(1\le j\le l_h)$; (ii) every collocation point $\xi_{j'h_*}$ $(j'=1,\dots,l_{h_*})$ occurs as a collocation point for a cell $G_{jh}\subset G_{j'h_*}$, i.e. $\Xi_{h_*}\subset\Xi_h$ where $\Xi_h=\{\xi_{jh}\}_{j=1,\dots,l_h}$. Introduce the following Banach spaces and operators:

$E=BC(G)$, space of bounded continuous functions $u:G\to\mathbb{R}$,

$$\|u\|=\sup_{x\in G}|u(x)|;$$

$E_h=C(\Xi_h)$, space of grid functions $u_h:\Xi_h\to\mathbb{R}$, $\|u\|=\max\limits_{\xi_{ih}\in\Xi_h}|u_h(\xi_{ih})|$;

$E_{h_*}=C(\Xi_{h_*})$;

$p_h\in L(E,E_h)$, restriction operator, $(p_h u)(\xi_{ih})=u(\xi_{ih})$ for $u\in E$, $\xi_{ih}\in\Xi_h$;

$p_{h_*}\in L(E,E_{h_*})$, similar restriction operator,

$p_{h_*h}\in L(E_h,E_{h_*})$, restriction operator, $(p_{h_*h}u_h)(\xi_{j'h_*})=u_h(\xi_{j'h_*})$ for $u_h\in E_h$, $\xi_{j'h_*}\in\Xi_{h_*}$;

$p_{hh_*}\in L(E_{h_*},E_h)$, piecewise constant prolongation operator, $(p_{hh_*}u_{h_*})(\xi_{jh})=u_{h_*}(\Pi_h\xi_{jh})$ for $u_{h_*}\in E_{h_*}$, $\xi_{jh}\in\Xi_h$ where $\Pi_h\xi_{jh}=\xi_{j'h_*}$ if $G_{jh}\subset G_{j'h_*}$;

$T\in L(E,E)$, the integral operator of equation (1),

$$(Tu)(x)=\int_G K(x,y)u(y)\,dy \quad \text{for } u\in E,\ x\in G;$$

$T_h\in L(E_h,E_h)$, approximation to T corresponding to method (4),

$$(T_h u_h)(\xi_{ih})=\sum_{j=1}^{l_h}\int_{G_{jh}} K(\xi_{ih},y)\,dy\ u_h(\xi_{jh}) \quad \text{for } u_h\in E_h,$$

$\xi_{ih}\in\Xi_h$;

$T_h' \in L(E_h, E_h)$, approximation to T corresponding to method (4');

$T_h'' \in L(E_h, E_h)$, approximation to T corresponding to method (4").

Sytems (4), (4') and (4") can be represented, respectively, as linear equations in finite dimensional space E_h:

$$u_h = T_h u_h + p_h f, \quad u_h = T_h' u_h + p_h f, \quad u_h = T_h'' u_h + p_h f.$$

We solve these equations using two grid iteration methods. For system (4), the algorithm is the following:

$$\begin{aligned} v_h^k &= u_h^k - T_h u_h^k - p_h f \qquad \text{(the residual of } u_h^k\text{)}, \\ u_h^{k+1} &= u_h^k - v_h^k - p_{hh_*}(I_{h_*} - T_{h_*})^{-1} p_{h_*h} T_h v_h^k, \quad k=0,1,2,\dots . \end{aligned} \tag{8}$$

In case of systems (4') and (4") we must only replace T_h, T_{h_*} by T_h', T_{h_*}' and T_h'', T_{h_*}'' respectively:

$$\begin{aligned} v_h^k &= u_h^k - T_h' u_h^k - p_h f, \\ u_h^{k+1} &= u_h^k - v_h^k - p_{hh_*}(I_{h_*} - T_{h_*}')^{-1} p_{h_*h} T_h' v_h^k, \quad k=0,1,2,\dots \end{aligned} \tag{8'}$$

and

$$\begin{aligned} v_h^k &= u_h^k - T_h'' u_h^k - p_h f, \\ u_h^{k+1} &= u_h^k - v_h^k - p_{hh_*}(I_{h_*} - T_{h_*}'')^{-1} p_{h_*h} T_h'' v_h^k, \quad k=0,1,2,\dots \end{aligned} \tag{8"}$$

On every iteration step one has to solve a system of type (4), (4') or (4") corresponding to the rough subdivision of G.

Methods of type (8), (8') originate from works of H. Brankhage [2] and K. Atkinson [1]. Abstract setting with compact convergence of operators was introduced and examined by I.K. Daugavet [9]. P. Uba [11,8] used these methods in case of one dimensional weakly singular integral equations presenting an analysis of convergence in an abstract setting. A systematical analysis of convergence rate and the amount of arithmetical work for two and multi-grid iteration methods was undertaken by W. Hackbusch [3]. His abstract results could be applied to methods (8)-(8"), too, but this way were more complicated and the results were weaker compared with a direct analysis presented in Sections 6-8. We follow [5] where case $\nu = n-2$ was considered.

6. <u>Convergence rate</u>. For convergence analysis, rewrite (8) in an equivalent form

$$u_h^{k+1} = T_{h,h_*} u_h^k + f_{h,h_*}, \quad k=0,1,2,\dots \tag{9}$$

where

$$f_{h,h_*} = p_h f + p_{hh_*}(I_{h_*} - T_{h_*})^{-1} p_{h_*h} T_h p_h f \in E_h \tag{10}$$

$$\begin{aligned} T_{h,h_*} &= \left[I_h - p_{hh_*}(I_{h_*} - T_{h_*})^{-1} p_{h_*h}(I_h - T_h)\right] T_h = (I_h - p_{hh_*} p_{h_*h}) T_h + \\ &+ p_{hh_*}(I_{h_*} - T_{h_*})^{-1}(p_{h_*h} T_h - T_{h_*} p_{h_*h}) T_h \in L(E_h, E_h). \end{aligned} \tag{11}$$

For methods (8') and (8"), the iteration formulae are similar - instead of T_h, T_{h_*}, operators T_h', T_{h_*}' for method (8') and T_h'', T_{h_*}'' for method (8") are involved. Introduce the corresponding operators

$$T'_{h,h_*}, T''_{h,h_*} \in L(E_h, E_h).$$

Theorem 4. Let assumptions (A1) and (A2) be fullfilled and let the subdivision of G corresponding to h and h_* ($h<h_*$) satisfy the conditions of Section 2 and compatibility conditions (i) and (ii), Section 5. Then, for sufficiently small $h_*>0$,

$$\|T_{h,h_*}\| \le c\varepsilon_{vh_*} \tag{12}$$

and

$$\|u_h^k - u_h\| \le \|u_h^0 - u_h\|(c\varepsilon_{vh_*})^k, \quad k=1,2,... \tag{13}$$

where u_h^k is defined by iteration method (8), u_h is the exact solution of system (4), ε_{vh} is defined in (5) and the constant c does not depend on h and h_*.

Under supplementary condition (6), we have $\|T'_{h,h_*}\| \le c\varepsilon_{vh_*}$, and estimate (13) for iterations (8') and the solution to (4') holds.

Under supplementary conditions (6) and (7), we have $\|T''_{h,h_*}\| \le c\varepsilon_{vh_*}$, and estimate (13) for iterations (8") and the exact solution to (4") holds.

Proof. It follows from the conditions of the Theorem that

$$\|(I_h - T_h)^{-1}\| \le \text{const} \quad (0<h<h_0)$$

(see the proof of Theorem 1, Section 4, in [6] or [7]). Note also that $\|p_{hh_*}\| = \|p_{h_*h}\| = 1$. From (11) we see that, to obtain (12), it sufficies to establish the inequalities

$$\|(I_h - p_{hh_*}p_{h_*h})T_h\| \le \text{const}\ \varepsilon_{vh_*}, \tag{14}$$

$$\|(p_{h_*h}T_h - T_{h_*}p_{h_*h})T_h\| \le \text{const}\ \varepsilon_{vh_*}. \tag{15}$$

Let us prove (14). To a $u_h \in E_h$, we coordinate a piecewise constant function $\bar{u}_h = \sum_{j=1}^{l_h} u_h(\xi_{jh})\varphi_{jh}$ (see Section 3). We have $\|\bar{u}_h\|_{L^\infty(\Omega)} = \|u_h\|_{E_h}$ and

$$(T_h u_h)(\xi_{ih}) = \sum_{j=1}^{l_h} \int_{G_{jh}} K(\xi_{ih}, y)\,dy\, u_h(\xi_{jh}) = \int_{G_h} K(\xi_{ih}, y)\bar{u}_h(y)\,dy =$$

$$= \int_G K(\xi_{ih}, y)\bar{u}_h(y)\,dy + \int_{G_h\setminus G} K(\xi_{ih}, y)\bar{u}_h(y)\,dy - \int_{G\setminus G_h} K(\xi_{ih}, y)\bar{u}_h(y)\,dy.$$

It follows from (2) that (see [7])

$$\int_{(G_h\setminus G)\cup(G\setminus G_h)} |K(x,y)|\,dy \le \text{const}\ \varepsilon_{vh^2} \le (\varepsilon_{vh})^2 .$$

$$\int_G |K(x^1,y) - K(x^2,y)|dy \le \text{const} \begin{cases} d_G(x^1,x^2) & , \nu<n-1, \\ d_G(x^1,x^2)(1+|\log d_G(x^1,x^2)|), & \nu=n-1, \\ d_G(x^1,x^2)^{n-\nu} & , \nu>n-1. \end{cases} \quad (16)$$

Therefore,

$$|(T_h u_h)(\xi_{ih}) - (T_h u_h)(\xi_{kh}) \le \text{const}\, \varepsilon^2_{\nu h} \|u_h\| +$$

$$+ \text{const} \begin{cases} d_G(\xi_{ih},\xi_{kh}) & , \nu<n-1 \\ d_G(\xi_{ih},\xi_{kh})(1+|\log d_G(\xi_{ih},\xi_{kh})|), & \nu=n-1 \\ d_G(\xi_{ih},\xi_{kh})^{n-\nu} & , \nu>n-1 \end{cases} \|u_h\|.$$

While $d_G(\xi_{ih}, \Pi_h \xi_{ih}) \le c_1 h_*$, we obtain

$$|(T_h u_h)(\xi_{ih}) - (T_h u_h)(\Pi_h \xi_{ih})| \le \text{const}\, \varepsilon_{\nu h_*} \|u_h\|. \quad (17)$$

Note that, for $v_h = T_h u_h$,

$$(v_h - p_{hh_*} p_{h_*h} v_h)(\xi_{ih}) = v_h(\xi_{ih}) - v_h(\Pi_h \xi_{ih}),$$

and, together with (17),

$$\|(I_h - p_{hh_*} p_{h_*h}) T_h u_h\| \le \text{const}\, \varepsilon_{\nu h_*} \|u_h\|,$$

i.e. (14) holds.

Let us prove (15). Denoting again $v_h = T_h u_h$ we have

$$((p_{h_*h} T_h - T_{h_*} p_{h_*h}) T_h u_h)(\xi_{i'h_*}) = (T_h v_h)(\xi_{i'h_*}) - (T_{h_*} p_{h_*h} v_h)(\xi_{i'h_*}) =$$

$$= \sum_{j=1}^{l_h} \int_{G_{jh}} K(\xi_{i'h_*}, y)dy\, v_h(\xi_{jh}) - \sum_{j'=1}^{l_{h_*}} \int_{G_{j'h}} K(\xi_{i'h_*}, y)dy\, v_h(\xi_{j'h_*}) =$$

$$= \sum_{j=1}^{l_h} \int_{G_{jh}} K(\xi_{i'h_*}, y)dy \left[v_h(\xi_{jh}) - v_h(\Pi_h \xi_{jh}) \right].$$

Applying inequality (17) with $T_h u_h = v_h$ we obtain

$$\|(p_{h_*h} T_h - T_{h_*} p_{h_*h}) T_h u_h\| \le \sup_{x \in G} \int_{G_h} |K(x,y)|dy\, c\, \varepsilon_{\nu h_*} \|u_h\| \le$$

$$\le \text{const}\, \varepsilon_{\nu h_*} \|u_h\|,$$

and (15) is proved. Thus, estimate (12) holds.

It is easy to check that the solution of equation $u_h = T_h u_h + p_h f$ (solution of (4)) satisfies the equation

$$u_h = T_{h,h_*} u_h + f_{h,h_*},$$

too. For $c\varepsilon_{\nu h_*} < 1$, this equation is uniquely solvable and iterations (9) converge to its solution with speed (13). In other words, iterations (8) converge to the solution of (4) with speed (13). For method (4), the proof of the Theorem is completed.

For methods (4') and (4"), the assertions of the Theorem follow from (11) and (12) taking into account that

$$\|T'_h - T_h\| \le \text{const}\,\varepsilon'_{vh}\,, \quad \|T'_{h_*} - T_{h_*}\| \le \text{const}\,\varepsilon'_{vh_*}\,,$$

$$\|T''_h - T'_h\| \le \text{const}\,\varepsilon_{vh}\,, \quad \|T''_{h_*} - T'_{h_*}\| \le \text{const}\,\varepsilon_{vh_*}$$

(see [7,4]).

Remark. As we see, the assertions of Theorem 4 can be extended to any further method such that $\|\tilde{T}_h - T_h\| \le c\varepsilon_{vh}$ where $\tilde{T}_h \in L(E_h, E_h)$ is an approximation to T_h corresponding to this method.

There are special algorithms to evaluate l_h^2 integrals of system (4) in $O(l_h^2)$ arithmetical operations and such that

$$\|\tilde{T}_h - T_h\| \le c(h^2 + h^{2(n-\nu)}) \le c\varepsilon^2_{vh}.$$

These algorithms are based on composite versions of cubature formula (3). Details are presented in another paper.

7. **Initial guesses.** We propose two possible initial guesses for iteration method (8); in methods (8') and (8''), similar initial guesses can be used.

First possible initial guess is given by

$$u_h^0 = p_{hh_*}u_{h_*} = p_{hh_*}(I_{h_*} - T_{h_*})^{-1}p_{h_*}f. \tag{18}$$

Under the conditions of Theorem 1, its error is estimated by

$$\|u_h^0 - u_h\| \le \text{const}\,\varepsilon_{vh_*} \tag{19}$$

Indeed,

$$\|u_h^0 - u_h\| \le \|p_{hh_*}(u_{h_*} - p_{h_*}u)\| + \|p_{hh_*}p_{h_*}u - p_h u\| + \|p_h u - u_h\| \le$$

$$\le \text{const}\,(\varepsilon^2_{vh_*} + \varepsilon^2_{vh}) + \sup_{\xi_{ih}\in\Xi_h} |u(\xi_{ih}) - u(\Pi_h\xi_{ih})|$$

where u is the solution to integral equation (1) and we applied estimate (5). Using (A4) and (16) we obtain (19).

The second possible initial guess is given by

$$u_h^0 = T_h p_{hh_*}u_{h_*} + p_h f = T_h p_{hh_*}(I_{h_*} - T_{h_*})^{-1}p_{h_*}f + p_h f. \tag{20}$$

Under the conditions of Theorem 1, its error is estimated by

$$\|u_h^0 - u_h\| \le \text{const}\,\varepsilon^2_{vh_*}. \tag{21}$$

Indeed, introduce the function

$$v_{h_*}(x) = \sum_{j'=1}^{l_{h_*}} \int_{G_{j'h_*}} K(x,y)\,dy\,u_{h_*}(\xi_{j'h_*}) + f(x), \quad x\in G,$$

a prolongation of the grid function u_{h_*} to Ω. It is established in [7] that

$$\sup_{x\in G} |v_{h_*}(x) - u(x)| \le \text{const}\,\varepsilon^2_{vh_*}\,.$$

From (20) we can read that $v_{h_*}(\xi_{ih}) = u_h^0(\xi_{ih})$, $\xi_{ih}\in\Xi_{ih}$, and a consequence from the last inequality is that

$$\|u_h^0 - p_h u\| \le \text{const}\, \varepsilon_{vh_*}^2 .$$

On the other hand, due to (5),

$$\|u_h - p_h u\| \le \text{const}\, \varepsilon_{vh}^2 ,$$

and estimate (21) follows from these two inequalities.

8. The amount of arithmetical work. The number of iteration steps in method (8) depends on the relation between h and h_* and on the desired accuracy of u_h^k, compared with u_h, the solution to (4). Beginning with initial guess (18), we obtain an accuracy

$$\|u_h^k - u_h\| \le \varepsilon_{vh}^p , \quad p \ge 2, \tag{22}$$

if k is such that $(c\varepsilon_{vh_*})^{k+1} \le \varepsilon_{vh}^p$ (see (13) and (19)), i.e.

$$k \ge \frac{p|\log \varepsilon_{vh}|}{|\log \varepsilon_{vh_*} + \log c|} - 1$$

(we assume that $\varepsilon_{vh} < 1$, $c\varepsilon_{vh_*} < 1$). Note that

$$\log \varepsilon_{vh} \sim \left\{ \begin{matrix} \log h, & v \le n-1 \\ (n-v)\log h, & v > n-1 \end{matrix} \right\} \quad \text{as} \quad h \to 0.$$

Choosing $h_* \asymp h^\tau$ $(0 < \tau < 1)$ we see that accuracy (22) will be achieved in the following asymptotical number of iteration steps:

$$k = [\, p/\tau\,] - 1$$

where [q] denotes the smallest integer exceeding q. For initial guess (20) this number is

$$k = [\, p/\tau\,] - 2.$$

Most extensive computational work during an iteration step (8) is caused by the terms $T_h u_h^k$ and $(I_{h_*} - T_{h_*})^{-1} z_{h_*}$ – the cost is, respectively, l_h^2 and $l_{h_*}^3/3$ additions and multiplications. Putting $h_* \asymp h^\tau$, $0 < \tau < 2/3$, and taking into account (or assuming) that then $l_{h_*} \asymp l_h^\tau$, the term $T_h u_h^k$ with its l_h^2 additions and multiplications will be most expensive. Note that the terms $p_{h_*h} T_h v_h^k$ and $T_h p_{hh_*} u_{h_*}$ (the last one occurs in (20)) cost $l_h l_{h_*}$ additions and multiplications, and other operations in (8) are of $O(l_h)$ additions and multiplications.

As a corollary, we present the asymptotical number k of iteration steps (8) and the asymptotical amount of arithmetical work to obtain the accuracy $\|u_h^k - u_h\| \le \varepsilon_{vh}^2$ for some strategies $h_* = h^\tau$:

Strategy	$h_* = h^{1/2}$	$h_* = h^{1/3}$
Initial guess (18)	k=3, $3l_h^2$ additions and multiplications	k=5, $5l_h^2$ additions and multiplications
Initial guess (20)	k=2, $2l_h^2$ additions and multiplications	k=4, $4l_h^2$ additions and multiplications

In practice one ought to compute an iteration more. But usually one gives an $\varepsilon>0$ (e.g. $\varepsilon = \varepsilon^p_{vh}$, $p\geq 2$) and stops the iterations as $\|v_h^k\|\leq\varepsilon$.

For iteration methods (8') and (8"), the estimation of the amount of computational work is similar as above for method (8).

In case of convolution type kernel $K(x,y) = a(x)\varkappa(x-y)b(y)$, using a regular subdivision of G with

$$\Xi_h = \{\lambda h : \lambda = (\lambda_1, \ldots, \lambda_n) \in Z^n, \lambda h \in G\},$$

an evaluation of $T_h'v_h$ and $T_h''v_h$ can be performed in $O(l_h \log_2 l_h)$ arithmetical operations using multidimensional fast Fourier transformation, see e.g. [10]. Systems (4') and (4") can be solved with ε^p_{vh} – accuracy in $O(l_h \log_2 l_h)$ arithmetical operations using iteration methods (8') or (8") with $h_* \asymp h^{1/3}$.

References

1. Atkinson K. Iterative variants of the Nystrom method for the numerical solution of integral equations. Numer. Math., 1973, Bd.22, 17-31.
2. Brakhage H. Über die numerische Behandlung von Integralgleichungen nach der Quadraturformelmethode. Numer. Math., 1960, V2, 183-196.
3. Hackbusch W. Integralgleichungen. Teubner, Stuttgard, 1989.
4. Vainikko G. and Pedas A. Convergence rate of a modified cubature formula method for multidimensional weakly singular integral equations. Acta et comment. univ. Tartuensis, 1990, 913, 3-17.
5. Вайникко Г. Интегральные уравнения одной внутренней-внешней задачи и их приближенное решение. Изв. АН Эстонии, Физ.Матем., 1990, 39, №3, 185-195.

6 . Вайникко Г. Некоторые коллокационные методы решения многомерных слабо сингулярных интегральных уравнений.In:"Numerical Analysis and Mathematical Modelling",Warsaw, Banach center publ., 1990,V. 24, 91-105.
7 . Вайникко Г. М. Кусочно-постоянная аппроксимация решения многомерных слабо сингулярных интегральных уравнений. Ж. вычисл. матем. и матем. физики ,1991, т.31, №6,832-849.
8 . Вайникко Г., Педас А., Уба П. Методы решения слабо-сингулярных интегральных уравнений.Тарту,Тартуск. ун-т, 1984.
9. Даугавет И.К. Об итеративном решении уравнений, возникающих при компактной аппроксимации операторов. Ж. вычисл. матем. и матем. физ., 1980, т.20, №4, 1046-1049.
10. Иванов В.И. Методы вычислений на ЭВМ. Киев, Наукова Думка, 1986.
11. Уба П. Итеративное решение интегрального уравнения со слабо особенным ядром. Acta et comment. univ. Tartuensis, 1983, 633, 67-74.

РЕШЕНИЕ БОЛЬШИХ СИСТЕМ, ВОЗНИКАЮЩИХ ПРИ ДИСКРЕТИЗАЦИИ МНОГОМЕРНЫХ СЛАБО СИНГУЛЯРНЫХ ИНТЕГРАЛЬНЫХ УРАВНЕНИЙ

Г. Вайникко

Резюме

Для дискретизации линейного интегрального уравнения второго рода в области $G \subset R^n$ привлекаются метод кусочно-постоянной коллокации, метод кубатурных формул и его модификация Кантаровича-Крылова. Возникающие системы линейных алгебраических уравнений решаютця за $O(l_h^2)$, в некоторых случаях $O(l_h \log_2 l_h)$ арифметических действий, сохранив при этом точность исходных методов ($l_h \asymp h^{-n}$ – число неизвестных системы). Это достигается привлечением двусеточных итерационных методов.

Acta et commentationes universitatis Tartuensis, 937, 1992, 15-26.

ON THE NUMERICAL SOLUTION OF A WEAKLY SINGULAR INTEGRAL EQUATION

Arvet Pedas

A class of one-dimensional weakly singular integral equations of the second kind is observed. Five approximative methods for the numerical solution of the integral equation are considered: a quadrature formulas method (using a second order quadrature formula), a modified quadrature formulas method, introduced by Kantorovich and Krylov [14], a piecevise constant spline collocation method, a linear spline collocation method, a subregions method with the piecevise constant splines. The first four methods are investigated in other papers. The aim of this paper is to investigate the convergence rate of the last method and to compare all the five methods numerically on an weakly singular integral equation which arises in the theory of intrinsic viscosity of macromolecules [2].

1. Introduction. We consider the integral equation

$$u(t) - \int_a^b K(t,s)u(s)ds = f(t), \quad a \le t \le b, \tag{1}$$

where [a,b] is a finite interval and f(t) is continuous on [a,b]. The kernel K(t,s) will be assumed to satisfy the following conditions:

(i) $K(t,s) = g(t,s)\varkappa(t-s)$;

(ii) $g \in C^2([a,b]\times[a,b])$;

(iii) $\varkappa \in C^1([a-b,b-a]\setminus\{0\})$ and for every $\tau \neq 0$, $\tau \in [a-b,b-a]$,

$$|\varkappa'(\tau)| \le b_1|\tau|^{-\beta} \quad (b_1 = \text{const};\ 0<\beta<2). \tag{2}$$

Here $C^p(G)$ denotes the set of functions which are p times continuously differentiable on G. It follows from (2) that

$$|\varkappa(\tau)| \le b_0(|\ln|\tau|| + 1) \quad (b_0 = \text{const};\ \beta = 1); \tag{3}$$

$$|\varkappa(\tau)| \le b_0(|\tau|^{-\beta+1} + 1) \quad (b_0 = \text{const};\ \beta \neq 1). \tag{4}$$

We define a space of functions, the Banach space $E^{\beta}=E^{\beta,2}$, by

$$E^{\beta}=\left\{u\in C[a,b]\cap C^2(a,b):\ \sup_{a<t<b}\frac{|u''(t)|}{(t-a)^{-\beta}+(b-t)^{-\beta}}<\infty\right\},$$

$$\|u\|_{E^{\beta}}=\max_{a\le t\le b}|u(t)|+\sup_{a<t<b}\frac{|u''(t)|}{(t-a)^{-\beta}+(b-t)^{-\beta}}.$$

For $\beta=1$ an inclusion $u\in E^{\beta}$ denotes that

$$|u'(t)|\le c_1(|\ln(t-a)|+|\ln(b-t)|+1)\quad (a<t<b); \tag{5}$$

$$|u''(t)|\le c_2((t-a)^{-1}+(b-t)^{-1})\quad (a<t<b). \tag{6}$$

For $\beta\neq 1$ an inclusion $u\in E^{\beta}$ denotes that

$$|u^{(k)}(t)|\le c_k\left((t-a)^{-\beta+2-k}+(b-t)^{-\beta+2-k}\right)\quad (a<t<b,\ k=1,2). \tag{7}$$

The following result characterize the smoothness of the solution of equation (1).

Theorem 1. Let the conditions (i)-(iii) be fulfilled. If $f\in E^{\beta}$ then all the integrable solutions u of the integral equation (1) belong to the space E^{β}.

The Proof of Theorem 1 is given in [13], pp. 10-11 ; see also [6,15].

2. Discretization of the integral equation. Let

$$a=t_1<t_2<\ldots<t_{n+1}=b \tag{8}$$

be a mesh of [a,b] such that

$$h_n=\max_{1\le j\le n}(t_{j+1}-t_j)\to 0\quad \text{as}\quad n\to\infty. \tag{9}$$

Denote

$$s_i=(t_i+t_{i+1})/2. \tag{10}$$

We introduce the following five discretizations (Methods 1-5) of equation (1).

Method 1.

$$u_i-\sum_{\substack{j=1\\ j\neq i}}^{n+1}w_jK(t_i,t_j)u_j=f(t_i),\quad i=1,\ldots,n+1. \tag{11}$$

where

$$w_j=\begin{cases}(t_2-t_1)/2 & \text{for}\quad j=1,\\ (t_{j+1}-t_{j-1})/2 & \text{for}\quad 2\le j\le n,\\ (t_{n+1}-t_n)/2 & \text{for}\quad j=n+1.\end{cases} \tag{12}$$

Method 2.

$$u_i-\sum_{\substack{j=1\\ j\neq i}}^{n+1}w_jK(t_i,t_j)[u_j-u_i]-u_i\int_a^b K(t_i,s)ds=f(t_i),\quad i=1,\ldots,n+1. \tag{13}$$

Method 3.

$$u_i - \sum_{j=1}^{n} \Big(\int_{t_j}^{t_{j+1}} K(s_i,s)ds \Big) u_j = f(s_i), \quad i=1,\dots,n. \tag{14}$$

Method 4.

$$u_i - \sum_{j=1}^{n+1} \Big(\int_{t_{j-1}}^{t_{j+1}} K(t_i,s)\varphi_j(s)ds \Big) u_j = f(t_i), \quad i=1,\dots,n+1, \tag{15}$$

where $t_0=t_1$, $t_{n+2}=t_{n+1}$ and

$$\varphi_j(s) = \begin{cases} (s-t_{j-1})/(t_j-t_{j-1}), & t_{j-1} \le s \le t_j \\ (t_{j+1}-s)/(t_{j+1}-t_j), & t_j \le s \le t_{j+1} \\ 0 \quad , & \text{elsewhere} \end{cases} \tag{16}$$

Method 5.

$$u_i(t_{i+1}-t_i) - \sum_{j=1}^{n} a_{ij}u_j = f_i \ , \quad i=1,\dots,n, \tag{17}$$

where

$$a_{ij} = \int_{t_i}^{t_{i+1}} \int_{t_j}^{t_{j+1}} K(t,s)dtds, \quad f_i = \int_{t_i}^{t_{i+1}} f(t)dt. \tag{18}$$

System (11) corresponds to a quadrature formulas method which is based on the trapezoidal formula for numerical quadrature : approximating the integral in (1) by means of the trapezoidal formula and collocating in the points t_i (i=1, ... ,n+1) we obtain system (11) if we reject the term where i=j (the kernel K(t,s) is not defined for t=s).

System (13) corresponds to the Kantorowich-Krylov modification of the quadrature formulas method: rewriting equation (1) in the form (see [14])

$$u(t) - \int_a^b K(t,s)[u(s)-u(t)]ds - u(t)\int_a^b K(t,s)ds = f(t) \tag{19}$$

and using quadrature formulas method with the trapezoidal formula for the first integral in (19) we obtain system (13).

System (14) corresponds to a piecewise constant spline collocation method: denoting by ψ_j the characteristic function of (t_j,t_{j+1}), approximating the solution u of equation (1) by a piecewise constant function

$$\tilde{u}_n(s) = \sum_{j=1}^{n} u_j\psi_j(s) \tag{20}$$

and collocating the equation (1) in the points s_i (i=1, ... ,n) we obtain system (14).

System (15) corresponds to a linear spline collocation method: approximating the solution u of equation (1) by

$$\tilde{\tilde{u}}_n(s) = \sum_{j=1}^{n+1} u_j \varphi_j(s)$$

($\varphi_j(t)$ are functions (16)) and collocating the equation (1) in the points t_i (i=1, ...,n+1) we otain system (15).

System (17) corresponds to a subregions method with piecewise constant functions (20): approximating the solution u of equation (1) by (20) and requiring that

$$\int_{t_i}^{t_{i+1}} \left[\tilde{u}_n(t) - \int_a^b K(t,s)\tilde{u}_n(s)ds - f(t)\right]dt = 0 , \quad i=1,\dots,n,$$

we obtain system (17).

Methods 1-4 are investigated in other papers (see Section 3). We shall refer to the main results needed for the comparison with Method 5. The goal of this paper is to investigate the convergence rate of Method 5 and to compare all five methods numarically (see Sections 3 and 5).

3. Error estimates. Denote

$$\varepsilon_n = \begin{cases} h_n & \text{for } \beta<1, \\ h_n(1+|\ln h_n|) & \text{for } \beta=1, \\ h_n^{2-\beta} & \text{for } \beta>1. \end{cases}$$

Theorem 2. Let conditions (i)–(iii) be fulfilled and $f \in E^\beta$. Let 1 be a non-characteristic value of integral equation (1). Let (8) and (9) hold. Then there exists a integer $n_0>0$ such that, for $n \geq n_0$, system (11) has a unique solution (u_i), and

$$\max_{1\leq i\leq n+1} |u_i - u(t_i)| \leq \text{const}\cdot\varepsilon_n,$$

where u is the (unique) solution to (1).

For proof see [9,13]. An analogous result for the multidimensional case is proved in [11].

Theorem 3. Let the conditions of Theorem 2 be fulfilled. Then there exists an integer $n_0>0$ such that, for $n \geq n_0$, system (13) has a unique solution (u_i), and

$$\max_{1\leq i\leq n+1} |u_i - u(t_i)| \leq \text{const}\cdot\varepsilon_n^2.$$

A proof of Theorem 3 (for $g(t,s)=1, \varkappa \in C^2([a-b,b-a]\setminus\{0\})$ and $f \in C^2[a,b]$) is given in [13] pp.22-36 and, more generally in [17]. An analogous result for the multidimensional case is proved in [7].

Remark 1. Theorems 2 and 3 are easily modifiable to a case where the approximate schemes (11) and (13) are based on the rectangular formula for numerical quadrature (in this case $w_j=t_{j+1}-t_j$ $(j=1,\dots,n)$ and in part of collocation nodes are nodes (10)).

Theorem 4. Let the conditions of Theorem 2 be fulfilled. Then there exists a integer $n_0>0$ such that, for $n\ge n_0$, system (14) has a unique solution (u_i), and

$$\max_{1\le i\le n} |u_i-u(s_i)| \le \text{const}\cdot\varepsilon_n^2,$$

where (s_i) are nodes (10).

The proof is given in [12]. An analogous result for the multidimensional case is proved in [10,11].

Theorem 5. Let the condition of Theorem 2 be fulfilled. Then there exists a integer $n_0>0$ such that, for $n\ge n_0$, system (15) has a unique solution (u_i), and

$$\max_{1\le i\le n+1} |\tilde{u}_i-u(t_i)| \le \text{const}\begin{cases} n^{-2} & \text{for } \beta<1,\\ n^{-2}(\ln n)^2 & \text{for } \beta=1,\\ n^{-2(2-\beta)}(\ln n)^{\beta-1} & \text{for } \beta>1.\end{cases}$$

The proof is given in [13], p.47-48; see also [8,16].

Theorem 6. Let the conditions of Theorem 2 be fulfilled. Then there exists a integer $n_0>0$ such that, for $n\ge n_0$, system (17) has a unique solution (u_i), and

$$\max_{1\le i\le n} |u_i-\tilde{u}_i| \le \text{const}\cdot\varepsilon_n^2,$$

where

$$\tilde{u}_i = \frac{1}{t_{i+1}-t_i}\int_{t_i}^{t_{i+1}} u(s)ds. \qquad (21)$$

The proof is presented in Section 4. We see that Method 1 is a method of the first degree of accuracy and Methods 2-5 are methods of the second degree of accuracy.

Remark 2. Theorems 2-6 may be generalized to a case where instead of condition (ii) the following condition holds:

$$g\in C^2([a,b]\times([a,b]\setminus\{d\})),\ a<d<b,$$

and

$$\frac{\partial^k g(t,s)}{\partial s^k}\quad (0\le k<\beta)$$

may have discontinuity of the first kind at $s=d$ (compare [12]).

4. Proof of Theorem 6. Denote by $E=C[a,b]$ the Banach space of continuous functions on $[a,b]$ with the norm $\|u\|_E = \max_{t\in[a,b]}|u(t)|$

and by $E_n = m_n$ the Banach space of vectors $\tilde{u}_n = (u_1, \dots, u_n)$ with the norm $\|\tilde{u}_n\|_{E_n} = \max_{1 \le j \le n} |u_j|$. Consider (1) as an operator equation $u - Tu = f$ in E and (17) as an operator equation $\tilde{u}_n - T_n \tilde{u}_n = p_n f$ in E_n, defining

$$(Tu)(t) = \int_a^b K(t,s)u(s)ds , \quad u \in E; \tag{22}$$

$$(T_n \tilde{u}_n)_i = \sum_{j=1}^{n} \left(\frac{1}{t_{i+1} - t_i} \int_{t_i}^{t_{i+1}} \int_{t_j}^{t_{j+1}} K(t,s)dtds \right) u_j,$$

$$i = 1, \dots, n; \ \tilde{u}_n = (u_1, \dots, u_n) \in E_n; \tag{23}$$

$$(p_n u)_i = \frac{1}{t_{i+1} - t_i} \int_{t_i}^{t_{i+1}} u(s)ds , \quad i = 1, \dots, n; \ u \in E. \tag{24}$$

It follows from (i)-(iii) that the sequence (T_n) of operators $T_n \in L(E_n, E_n)$ converges compactly (see [9], p.32) to the operator $T \in L(E,E)$ in relation to connection operators $p_n \in L(E, E_n)$:

$$T_n \to T \text{ compactly.}$$

On the conditions of Theorem 2 now it follows from the convergence theorem for operator equations (see [9], p.49) that for sufficiently great n (say $n \ge n_0$) the equation $\tilde{u}_n - T_n \tilde{u}_n = p_n f$ has a unique solution $\tilde{u}_n = (u_1, \dots, u_n)$ and the following error estimate holds:

$$\|\tilde{u}_n - p_n u\|_{E_n} \le \text{const} \ \|p_n T_n u - T_n p_n u\|_{E_n},$$

where u is the (unique) solution to (1).

We must prove that

$$\|p_n T u - T_n p_n u\|_{E_n} \le \text{const} \cdot \varepsilon_n^2. \tag{25}$$

Following [12], we have (see (22)-(24), (8), (10), (21))

$$\|p_n T_n u - T_n p_n u\|_{E_n} = \max_{1 \le i \le n} \Big| \sum_{j=1}^{n} \frac{1}{t_{i+1} - t_i} \int_{t_i}^{t_{i+1}} \int_{t_j}^{t_{j+1}} K(t,s)u(s)dtds -$$

$$- \sum_{j=1}^{n} \frac{1}{t_{i+1} - t_i} \int_{t_i}^{t_{i+1}} \int_{t_j}^{t_{j+1}} K(t,s)dtds \cdot \tilde{u}_j \Big| =$$

$$= \max_{1 \le i \le n} \Big| \sum_{j=1}^{n} \frac{1}{t_{i+1} - t_i} \int_{t_i}^{t_{i+1}} \int_{t_j}^{t_{j+1}} K(t,s)|u(s) - \tilde{u}_j]dtds \Big| \le$$

$$\le \max_{1\le i\le n} (I_i^{(1)}+I_i^{(2)}+I_i^{(3)}+I_i^{(4)}),$$

where

$$I_i^{(1)}= \Big| \sum_{j\in J(i)} \frac{1}{t_{i+1}-t_i} \int_{t_i}^{t_{i+1}} \int_{t_j}^{t_{j+1}} K(t,s)[u(s)-\tilde{u}_j]dtds \Big|,$$

$$I_i^{(2)}= \Big| \sum_{j\in J} \frac{1}{t_{i+1}-t_i} \int_{t_i}^{t_{i+1}} \int_{t_j}^{t_{j+1}} K(t,s)[u(s)-\tilde{u}_j]dtds \Big|,$$

$$I_i^{(3)}= \Big| \sum_{j\notin J(i),j\in J} \frac{1}{t_{i+1}-t_i} \int_{t_i}^{t_{i+1}} \int_{t_j}^{t_{j+1}} [K(t,s)-K(t,s_j)][u(s)-\tilde{u}_j]dtds \Big|,$$

$$I_i^{(4)}= \Big| \sum_{j\notin J(i),j\in J} \frac{1}{t_{i+1}-t_i} \int_{t_i}^{t_{i+1}} \int_{t_j}^{t_{j+1}} K(t,s_j)[u(s)-\tilde{u}_j]dtds \Big|,$$

$$J(i) = \{ j: |t_j-s_i| \le h_n , |t_{j+1}-s_i| \le h_n \},$$

$$J = \{ j: |t_j-a| \le h_n , |b-t_{j+1}| \le h_n \}.$$

Denote

$$\gamma_i(h_n) = [a,b] \cap [s_i-h_n, s_i+h_n],$$

$$\eta_i(h_n) = \{ s\in[a,b]: a+h_n \le s \le b-h_n , |s-s_i| \ge h_n \}.$$

From conditions (i)-(iii) it follows that

$$\frac{1}{t_{i+1}-t_i} \int_{t_i}^{t_{i+1}} \int_{\gamma_i(h_n)} |K(t,s)|dtds \le \text{const}\cdot\varepsilon_n , i=1,\dots,n. \qquad (26)$$

From (5),(7) ($u\in E^\beta$, see Theorem 1) and

$$u(s)-\tilde{u}_j = u(s)-u(s_j)+u(s_j)-\tilde{u}_j = \int_{s_j}^{s} u'(t)dt + \frac{1}{t_{i+1}-t_i} \int_{t_j}^{t_{j+1}} \Big(\int_{t}^{s_j} u'(s)ds \Big) dt$$

we obtain that

$$\max_{t_j\le s\le t_{j+1}} |u(s)-\tilde{u}_j| \le \text{const}\cdot\varepsilon_n , j=1,\dots,n. \qquad (27)$$

Now we show that

$$I_i^{(k)} \le \text{const}\cdot\varepsilon_n^2 , i=1,\dots n; k=1,\dots,4.$$

Using (27) and (26) we obtain

$$I_i^{(1)} \le \text{const}\cdot\varepsilon_n \sum_{j\in J(i)} \frac{1}{t_{i+1}-t_i} \int_{t_i}^{t_{i+1}} \int_{t_j}^{t_{j+1}} |K(t,s)|dtds \le$$

$$\leq \text{const}\cdot\varepsilon_n \frac{1}{t_{i+1}-t_i}\int_{t_i}^{t_{i+1}}\int_{\gamma_i(2h_n)} |K(t,s)|dtds \leq$$

$$\leq \text{const}\cdot\varepsilon_n^2, \quad i=1,\dots,n;$$

$$I_i^{(2)} \leq \text{const}\cdot\varepsilon_n^2, \quad i=1,\dots,n;$$

$$I_i^{(3)} \leq \text{const}\cdot\varepsilon_n \sum_{j\neq J(i),j\in J} \frac{1}{t_{i+1}-t_i}\int_{t_i}^{t_{i+1}}\int_{t_j}^{t_{j+1}} |K(t,s)-K(t,s_j)|dtds \leq$$

$$\leq \text{const}\cdot\varepsilon_n h_n \sum_{j\neq J(i),j\in J} \frac{1}{t_{i+1}-t_i}\int_{t_i}^{t_{i+1}}\int_{t_j}^{t_{j+1}} \sup_{0<\vartheta<1}\left|\frac{\partial K(t,\vartheta s+(1-\vartheta)s_j)}{\partial s}\right| dtds.$$

For $j\neq J(i), t\in[t_i,t_{i+1}], s\in[t_j,t_{j+1}], \vartheta\in[0,1]$ we have

$$0<c_1\leq |t-[\vartheta s+(1-\vartheta)s_j]|\,|t-s|^{-1} \leq c_2 \quad (c_1,c_2=\text{const})$$

and therefore (see (i)-(iii))

$$I_i^{(3)} \leq \text{const}\cdot\varepsilon_n\cdot h_n \frac{1}{t_{i+1}-t_i}\int_{t_i}^{t_{i+1}}\int_{\eta_i(h_n)} |t-s|^{-\beta}dtds \leq$$

$$\leq \text{const}\cdot\varepsilon_n^2, \quad i=1,\dots,n.$$

Further, we have (see(10))

$$\int_{t_j}^{t_{j+1}} (s-s_j)ds = 0$$

and therefore, for $j\neq J(i), j\in J$

$$I_i^{(4)} \leq c \sum_{j\neq J(i),j\in J} \frac{1}{t_{i+1}-t_i}\int_{t_i}^{t_{i+1}}\int_{t_j}^{t_{j+1}} |K(t,s_j)|\sup_{0<\vartheta<1}|u''(\vartheta s+(1-\vartheta)s_j)|\,|s-s_j|^2 dtds.$$

Using (6),(7),(i)-(ii),(3) and (4) we finally obtain that

$$I_i^{(4)} \leq \text{const}\cdot\varepsilon_n^2, \quad i=1,\dots,n.$$

Therefore (25) holds and the proof of Theorem 6 is completed.

5. **Numerical results**. In this section we consider a weakly singular integral equation from the field of polymer physics [2]:

$$\lambda u(t) - \int_{-1}^{1} |t-s|^{-\alpha} u(s)ds = F(t), \quad -1\leq t\leq 1, \tag{28}$$

where $0<\alpha<1$. Numerical solutions for (28) have been obtained

in [4,3,1,5].Following [3], we consider (28) by

$$\alpha=1/2, \tag{29}$$

$$\lambda=3/2, \tag{30}$$

$$F(t)=1.5(1-t^2)^{3/4}-0.375\pi\sqrt{2}\,(2-t^2). \tag{31}$$

In this case conditions (i)-(iii) are fulfilled ($a=-1, b=1, g(t,s)=2/3$, $\varkappa(\tau)=|\tau|^{-1/2}$, $\beta=3/2$), $f=F/\lambda \in E^{3/2}$ (see Section 1), and (28) has an exact solution

$$u^*(t)=(1-t^2)^{3/4}.$$

Let $n\geq 1$ be a integer , $h=2/n$, $t_i=-1+(i-1)h$, $i=1,\ldots,n+1$, and $s_i=(t_i+t_{i+1})/2$. Using this selection of $[-1,1]$ we solved (28) numerically by Medhods 1-5 on the conditions (29)-(31). The integrals in the right hand side of (17) were calculated as follows:

$$f_i=\frac{2}{3}\int_{t_i}^{t_{i+1}}F(t)dt \approx (t_{i+1}-t_i)(1-s_i^2)^{3/4}-$$

$$-\ 0.25\sqrt{2}\pi\left[2(t_{i+1}-t_i)-\frac{t_{i+1}^3}{3}+\frac{t_i^3}{3}\right].$$

All the other integrals, which are needed for the constitution of systems (14),(15) and (17), were calculated analytically.

Systems (11),(13),(14),(15) and (17) for (28) were solved (by standard Gauss-techniques) in the Computing Centre of Tartu University on the Computer EC-1060. Some results for $n=4$, $n=8$, $n=16$, $n=32$, $n=64$ and $n=128$ are given in Table 1.

Table 1

Maximum error for Methods 1-5 and [3]

	Method 1	Method 2	Method 3	Method 4	Method 5	Method [3]
n=4	5.53	0.29	0.15	0.19	0.14	0.089
n=8	1.70	0.097	0.060	0.082	0.068	0.030
n=16	0.94	0.039	0.023	0.032	0.028	0.011
n=32	0.61	0.016	0.0080	0.011	0.0099	0.0038
n=64	0.43	0.0062	0.0029	0.0039	0.0036	
n=128	0.32	0.0028	0.0010	0.0019	0.0013	

In Table 1, the error

$$\max_{1\leq i\leq n+1}|u_i-u^*(t_i)|$$

for Methods 1,2,4 and, the error

$$\max_{1\le i\le n} |u_i - u^*(s_i)|$$

for Methods 3 and 5 are given. For supplementary comparision, we also give an error

$$\max_{1\le i\le n} |u_i - u^*(\tau_i)|$$

from [3], where (u_i) were obtained by means of a polynomial collocation method and

$$\tau_i = \cos(\frac{2i-1}{2n})\pi, \quad i=1,\dots,n.$$

According to Theorems 2-6 we have the following estimations for Methods 1-5 (h=2/n):

Method 1: $\max_{1\le i\le n+1} |u_i - u^*(t_i)| \le \text{const}\cdot\sqrt{h}$;

Method 2: $\max_{1\le i\le n+1} |u_i - u^*(t_i)| \le \text{const}\cdot h$;

Method 3: $\max_{1\le i\le n} |u_i - u^*(s_i)| \le \text{const}\cdot h$;

Method 4: $\max_{1\le i\le n+1} |u_i - u^*(t_i)| \le \text{const}\cdot h\,(1+\sqrt{|\ln h|})$;

Method 5: $\max_{1\le i\le n} \left|u_i - \frac{1}{h}\int_{t_i}^{t_{i+1}} u^*(s)ds\right| \le \text{const}\cdot h$.

From Table 1 we can see that the numerical results are consistent with these theoretical estimations (we actually see that the rate of convergence of Methods 2-5 is better than $O(h)$).

Reference

1. Cohen H., Ickovic J. Numerical treatment of singular integral equations. J. Comput. Phys., 1974, v.16, No.4, 371-382.
2. Kirkwood J.G., Riseman J. The intrinsic viscosities and diffusion constants of flexible macromolecules in solution. J. Chem. Phys., 1948, v.16, 565-573.
3. Phillips J.L. The use of collocation as a projection method for solving linear operator equations. SIAM J. Numer. Anal., 1972, v.9, No.1, 14-28.
4. Schlitt D.W. Numerical solution of a singular integral equation encountered in polymer physics. J. Math. Phys., 1968, v.9, 436-439.

5. Sloan I.H., Burn B.J. Collocation with polynomials for integral equations of the second kind: a new approach to the theory. J. Integr. Equations, 1979, v.1, No.1, 77-94.
6. Vainikko G., Pedas A. The properties of solutions of weakly singular integral equations. J. Austral. Math. Soc., Series B, 1981, v.22, 419-430.
7. Vainikko G., Pedas A. Convergence rate of a modified cubature formula method for multidimensional weakly singular integral equations. Acta et Comment. Universitatis Tartuensis 1990, 913, 3-17.
8. Vainikko G.,Uba P. A piecewise polynomial approximation to the solution of an integral equation with weakly singular kernel. J.Austral.Math.Soc..Series B,1981,v.22,431-438.
9. Вайникко Г.М. Анализ дискретизационных методов.Тарту,Тартуск. ун-т,1976.
10. Вайникко Г.М. Некоторые коллокационные методы решения многомерных слабо сингулярных интегральных уравнений. In : "Numerical Analysis and Mathematical Modelling", Warsaw, Banach center publ.,1990,v.24,91-95.
11. Вайникко Г.М. Кусочно-постоянная аппроксимация решения многомерных слабо сингулярных интегральных уравнений. Ж. вычисл. матем. и матем. физики. 1991, т.31, №6, 832-849.
12. Вайникко Г. , Педас А. Кусочно-постоянная аппроксимация решения слабо-особого интегрального уравнения. Уч. зап. Тартуск.ун-та, 1989,вып. 863, 31-39.
13. Вайникко Г., Педас А., Уба П. Методы решения слабо-сингулярных интегральных уравнений. Тарту, Тартуск. ун-т, 1984.
14. Канторович Л.В., Крылов В.И. Приближенные методы высшего анализа. Москва-Ленинград,Физматгиз.,4.изд.,1952.
15. Педас А. О гладгости решения интегрального уравнения со слабо сингулярным ядром. Уч. зап. Тартуск. ун-та. 1979. вып. 492, 56-68.
16. Педас А. Кусочно-линейная аппроксимация решения интегрального уравнения с логарифмической особенностью в ядре. Уч.зап. Тартуск. ун-та,1979. вып. 500. 33-42.
17. Педас А. О решении слабо-сингулярных уравнений методом механических квадратур с формулой трапеций. Уч.зап. Тартуск. ун-та,1987, вып. 762, 89-97.

О ЧИСЛЕННОМ РЕШЕНИИ ОДНОГО СЛАБО СИНГУЛЯРНОГО ИНТЕГРАЛЬНОГО УРАВНЕНИЯ

Арвет Педас

Резюме

Рассматривается интегральное уравнение (1), на ядро которого налагаются условия (i)-(iii).Строятся пять приближенных методов для решения уравнения (1)-методы (11),(13),(14),(15) и (17).Теоремы 2-5, касающиеся методов (11) и (13)-(15), доказаны в других работах. Целью настоящей работы является установление теоремы 6, описывающей быстроту сходимости метода подобластей с кусочно-постоянными функциями (метода (17)). Приводится его сравнение с методами (11) и (13)-(15) при численном решении одного конкретного слабо сингулярного интегрального уравнения .

Acta et commentationes universitatis Tartuensis, 937, 1992, 27-35.

APPROXIMATE COMPUTATION OF WEAKLY SINGULAR INTEGRALS

Peep Uba

A method for numerical integration of weakly singular functions based on the trapezoidal rule on a special non-uniform grid is investigated. Using the Runge method the fourth degree of convergence is obtained.

1. **Introduction**. Here we are concerned with a numerical computation of a weakly singular integral

$$I(g) = \int_0^b g(x)\,dx \tag{1}$$

where $g(x)$ satisfies

$$|g^{(k)}(x)| \le c|x|^{-\alpha-k}, \quad 0<x\le b, \quad \alpha<1, \quad k=0,1,2,3,4, \tag{2}$$

(algebraic singularity) or

$$|g(x)| \le c|\ln x|, \quad |g^{(k)}(x)| \le c|x|^{-k}, \quad 0<x\le b, \quad k=1,2,3,4$$

(logarithmic singularity).

Here $b<\infty$ and c is some positive constant.

It is clear that the integral (1) exists in the sense of Lebesque and that $g(x)$ is finite valued everywhere on interval $(0,b]$ except at point $x=0$ where it may be undefined.

The numerical treatment of such integrals has been studied in many papers, we refer only to [2,4,5]. G.Opfer in [2] investigates the replacing of the unbounded function $g(x)$ by a bounded function $f(x)$ and then applying an ordinary quadrature formula.

In [4] a special non-uniform grid is used with replacing the function $g(x)$ in the neighbourhood of point $x=0$ with a bounded function. It is shown that by using the trapezoidal formula the function $g(x)$ can be arbitrarily replaced there and by using the present grid the expected error is the smallest.

The application of Jacobi polynomials in constructing quadrature rules is presented, for example, in [5]. In the same place a method based on the weakening of the singularity is studied.

We refer also to Rice [3], who studied graded grids for approximation of functions with singularities.

The method presented here consists of three principal steps. Firstly, we determine two special non-uniform grids and replace the unbounded function g(x) by a bounded function in subintervals containing the point of singularity. Secondly, on determined grids by trapezoidal formulae we compute two approximate values of integral, and thirdly the Runge method for the correction of the received results is used.

This method is announced in [6]. We also refer to the term-paper of K.Kõppo [1] who appears to have been the first to study the above-mentioned method by numerical experiments.

2. Description of the method.

A. The first step. As knots of trapezoidal formulae we choose

$$x_i = b\left(\frac{i}{n}\right)^r, \quad i=0,1,\dots,n, \quad r = \frac{\beta}{1-\alpha}, \tag{3}$$

where n is some integer, $\beta \geq 1$, $\beta \in \mathbb{R}$ and r characterizes the degree of non-uniformity of the grid. These knots are located with a greater density toward point x=0. In case of logarithmic singularity $\alpha = 0$.

Similarly we determine another grid with knots

$$x_i^N = b\left(\frac{i}{N}\right)^r, \quad i=0,1,\dots,N, \quad N=2n.$$

It is easy to see that

$$x_i = x_{2i}^N, \quad i=0,1,\dots n,$$

and

$$x_{2i+1}^N \in (x_i, x_{i+1}), \quad i=0,1,\dots,n-1.$$

Without loss of generality as a finite valued function we choose the function

$$f(x) = \begin{cases} 0 & \text{for } x=0, \\ g(x) & \text{for } x \in (0,b]. \end{cases}$$

B. The second step. On each subinterval by using the trapezoidal rule we calculate the approximate value of integrals $I_i(f) = \int_{x_i}^{x_{i+1}} f(x)\,dx$:

$$I_i^n(f) := \frac{f(x_i)+f(x_{i+1})}{2}(x_{i+1}-x_i), \quad i=0,1,\dots,n-1,$$

and $I_i^N(f)$ for $i=0,1,\dots,2n-1$ analogously.

In this case here halves of the knots coincide which is useful in practice.

C. The third step. For each $i=1,2,\dots,n-1$ we determine the coefficient

$$k_i = \frac{(x^N_{2i+1} - x^N_{2i})^3 + (x^N_{2i+2} - x^N_{2i+1})^3}{(x^N_{2i+1} - x^N_{2i})^3 + (x^N_{2i+2} - x^N_{2i+1})^3 - (x^N_{2i+2} - x^N_{2i})^3} =$$

$$= \frac{((2i+1)^r - (2i)^r)^3 + ((2i+2)^r - (2i+1)^r)^3}{((2i+1)^r - (2i)^r)^3 + ((2i+2)^r - (2i+1)^r)^3 - ((2i+2)^r - (2i)^r)^3} \quad . \tag{3}$$

At this we notice that these coefficients depend only on r (on α) and i and not on n and can be calculated as follows

$$k_i = -(z_i + 1/z_i - 1)/3, \qquad z_i = ((2i+2)^r - (2i+1)^r)/((2i+1)^r - (2i)^r)$$

On subintervals $[x_i, x_{i+1}]$, $i=1,\dots,n-1$, we correct the received approximate values of integrals $I_i(f)$ as follows

$$\tilde{I}_i(f) := I^N_{2i}(f) + I^N_{2i+1}(f) + k_i \times [I^n_i(f) - I^N_{2i}(f) - I^N_{2i+1}(f)]$$

and finally we get the approximate value of integral (1):

$$\tilde{I}(g) = I^N_0(f) + I^N_1(f) + \sum_{i=1}^{n-1} \tilde{I}_i(f) \quad . \tag{4}$$

3. Theorem. *For the method described above the error estimation*

$$|I(g) - \tilde{I}(g)| \le c \begin{cases} n^{-\beta} & \text{for } \beta < 4, \\ n^{-4} \ln n & \text{for } \beta = 4, \\ n^{-4} & \text{for } \beta > 4 \end{cases} \tag{5}$$

is valid.

Proof. For the first summand in (4) we get the estimation

$$\left| \int_0^{x_1^N} g(x)\,dx - I^N_0(f) \right| = \left| \int_0^{x_1^N} g(x)\,dx - \frac{f(x_1^N) + f(0)}{2} \cdot x_1^N \right| \le$$

$$\le \int_0^{x_1^N} |g(x)|\,dx + |f(x_1^N)| \cdot x_1^N/2 \overset{(2)}{\le} c \left(\int_0^{x_1^N} x^{-\alpha}\,dx + (x_1^N)^{1-\alpha} \right) \le$$

$$\le c_1 (x_1^N)^{1-\alpha} = c_2 (1/n)^{r(1-\alpha)} = c_2 n^{-\beta}. \tag{6}$$

Using the estimation on the remainder term of the trapezoidal rule we get

$$\left| \int_{x_1^N}^{x_1} f(x)\,dx - I^N_1(f) \right| \le \max_{\xi \in [x_1^N, x_1]} |f''(\xi)| \frac{(x_1 - x_1^N)^3}{12} \overset{(2)}{\le}$$

$$\le c \left(\frac{1}{n}\right)^{r(-\alpha-2)} \cdot \left(\left(\frac{1}{n}\right)^r - \left(\frac{1}{2n}\right)^r \right)^3 \le c_1 \left(\frac{1}{n}\right)^{r(-\alpha-2)} \cdot \left(\frac{1}{n}\right)^{3r} =$$

$$= c_1 \left(\frac{1}{n}\right)^{r(1-\alpha)} = c_1 n^{-\beta} \tag{7}$$

which shows the needed error estimation on the second summand in (4).

In case of other summands in (4) we deduce the remainder term of the trapezoidal formula using the Taylor expansion of f(x), $f(x^N_{2i+1})$ and $f(x^N_{2i})$ at point $x = x^N_{2i+1}$. We can write

$$\int_{x^N_{2i}}^{x^N_{2i+2}} g(x)\,dx = I_i^n(f) + f''(x^N_{2i+1})\cdot K^S_{1,i} + f'''(x^N_{2i+1})\,K^S_{2,i} +$$

$$\frac{f^{IV}(\xi)}{120}\left[(x^N_{2i+2} - x^N_{2i+1})^5 + (x^N_{2i+1} - x^N_{2i})^5\right] - \left[f^{IV}(\xi')(x^N_{2i+1} - x^N_{2i})^4 +\right.$$

$$\left. + f^{IV}(\xi'')\cdot(x^N_{2i+2} - x^N_{2i+1})^4\right]\cdot\frac{(x^N_{2i+2} - x^N_{2i})}{48}$$

where

$$K^S_{1,i} = -(x^N_{2i+2} - x^N_{2i})^3/12,$$

$$K^S_{2,i} = \left\{(x^N_{2i+2} - x^N_{2i+1})^4 - (x^N_{2i+1} - x^N_{2i})^4 - 2(x^N_{2i+2} - x^N_{2i})\left[(x^N_{2i+2} - x^N_{2i+1})^3 -\right.\right.$$

$$\left.\left. - (x^N_{2i+1} - x^N_{2i})^3\right]\right\}/24$$

and $\xi\in\left[x^N_{2i}, x^N_{2i+2}\right]$, $\xi'\in\left[x^N_{2i}, x^N_{2i+1}\right]$, $\xi''\in\left[x^N_{2i+1}, x^N_{2i+2}\right]$.

Analogously for i=1,....,n-1 the equality

$$\int_{x^N_{2i}}^{x^N_{2i+2}} g(x)\,dx = I^N_{2i}(f) + I^N_{2i+1}(f) + f''(x^N_{2i+1})\cdot K^d_{1,i} + f'''(x^N_{2i+1})K^d_{2,i} +$$

$$+\left(\frac{f^{IV}(\xi)}{120} - \frac{f^{IV}(\xi')}{48}\right)(x^N_{2i+1} - x^N_{2i})^5 + \left(\frac{f^{IV}(\xi'')}{120} + \frac{f^{IV}(\xi''')}{48}\right)\cdot(x^N_{2i+2} - x^N_{2i+1})^5$$

is valid, where

$$K^d_{1,i} = \left[-(x^N_{2i+1} - x^N_{2i})^3 - (x^N_{2i+2} - x^N_{2i+1})^3\right]/12,$$

$$K^d_{2,i} = \left[(x^N_{2i} - x^N_{2i+1})^4 - (x^N_{2i+2} - x^N_{2i+1})^4\right]/24$$

and

$$\xi,\xi'\in\left[x^N_{2i}, x^N_{2i+1}\right];\quad \xi'',\xi'''\in\left[x^N_{2i+1}, x^N_{2i+2}\right].$$

Observing the Runge idea, we determine two coefficients k_i and k_i' such that

$$\begin{cases} k_i + k_i' = 1 \\ k_i\cdot K^S_{1,i} + k_i'\cdot K^d_{1,i} = 0 \end{cases}$$

From here we get the expression (3) and therefore also the following

inequality holds

$$R_i(f) := \Big| \int_{x_{2i}^N}^{x_{2i+2}^N} f(x)\,dx - k_1 \cdot I_i^n(f) - (1-k_1)\big(I_{2i}^N(f) + I_{2i+1}^N(f)\big)\Big| \le$$

$$\le |f'''(x_{2i+1}^N)| \,|k_1 \cdot K_{2,i}^S + (1-k_1) K_{2,i}^d| + c \cdot (x_{2i}^N)^{-\alpha-4} \cdot (x_{2i+2}^N - x_{2i}^N)^5.$$

Here the assumption (2) and inequalities

$$x_{2i+1}^N - x_{2i}^N \le x_{2i+2}^N - x_{2i+1}^N \le x_{2i+2}^N - x_{2i}^N$$

are used.

It is easy to see the bondedness of sequence $\{k_i\}$ which converges to -1/3 by $i \to \infty$ (or by $r \to 1$, which corresponds to the well-known Runge constant for the trapezoidal rule on the uniform grid).

With the help of equality $(2i+1)^r - (2i)^r = r \cdot \xi^{r-1}$, $\xi \in [2i, 2i+1]$, we can estimate the coefficients $K_{2,i}^d$ as follows

$$|K_{2,i}^d| \le c \cdot \Big|\Big(\Big(\frac{2i+1}{2n}\Big)^r - \Big(\frac{2i}{2n}\Big)^r\Big)^4 - \Big(\Big(\frac{2i+2}{2n}\Big)^r - \Big(\frac{2i+1}{2n}\Big)^r\Big)^4\Big| \le$$

$$\le c_1 \Big(\frac{1}{n}\Big)^{4r} \big((\xi^{r-1})^4 - (\xi_1^{r-1})^4\big) = c_2 \Big(\frac{1}{n}\Big)^{4r} \xi_2^{4r-5}$$

where $\xi \in [2i, 2i+1]$, $\xi_1 \in [2i+1, 2i+2]$, $\xi_2 \in [\xi, \xi_1]$ and c_i are some constants.

From this $|K_{2,i}^d| \le c \Big(\frac{1}{n}\Big)^{4r} (2i+2)^{4r-5}$ and analogously the inequalities

$$|K_{2,i}^S| \le c \Big(\frac{1}{n}\Big)^{4r} (2i+2)^{4r-5}$$

and

$$(x_{2i+2}^N - x_{2i}^N)^5 \le c \Big(\frac{1}{n}\Big)^{5r} (2i+2)^{5r-5}$$

hold. This permits us to conclude that

$$R_i(f) \overset{(2)}{\le} c \Big(\frac{2i+1}{2n}\Big)^{r(-\alpha-3)} \cdot \Big(\frac{1}{n}\Big)^{4r} (2i+2)^{4r-5} +$$

$$+ c_1 \Big(\frac{2i}{2n}\Big)^{r(-\alpha-4)} \cdot \Big(\frac{1}{n}\Big)^{5r} (2i+2)^{5r-5} \le$$

$$\le c_2 \Big(\frac{1}{n}\Big)^{r(1-\alpha)} \cdot i^{-\alpha r - 3r + 4r - 5} + c_3 \Big(\frac{1}{n}\Big)^{r(1-\alpha)} \cdot i^{-\alpha r - 4r + 5r - 5} =$$

$$= c_4 n^{-\beta} (i)^{\beta - 5}.$$

and therefore on each summand under sum-symbol in (4) the

inequality

$$\left| \int_{x_{2i}^N}^{x_{2i+2}^N} f(x) - \tilde{I}_i(f) \right| \le c\, n^{-\beta}\, (i)^{\beta - S}. \qquad (8)$$

holds. Using the equality

$$\sum_{i=1}^{n-1} i^{\beta - S} = c \begin{cases} 1 & \text{for } \beta < 4, \\ ln\, n & \text{for } \beta = 4, \\ n^{\beta - 4} & \text{for } \beta > 4 \end{cases}$$

from (8) and (6),(7) immediately the inequality (5) follows.

The Theorem is proved.

4. Numerical experiments.

The method is tested with the functions

$$g(x) = x^{-\alpha}, \quad 0 < \alpha \le 0{,}7$$

and

$$g(x) = \ln x .$$

In case of larger α the results obtained with the relatively similar program are not authentic because of the accumulation of round-off errors in the neigbourhood of the point of singularity.

In Figure 1 the behavior of real errors on subintervals by $\beta = 3$, $b = 0.1$, $\alpha = 0.5$ and $n = 64$ is presented. With curve (·····) the errors of the trapezoidal rule on the sparse grid ($n = 64$) are given, with curve (- - -) the errors of the trapezoidal rule on the dense grid are given and with continuous curve (——) the errors after correction are present. We remark that by other α ($0 \le \alpha < 0.7$), b ($0.0001 \le b \le 100$) and n ($n \le 256$) the errors behave fully analogously. By other β ($2 \le \beta \le 4.5$) the curves (·····) and (- - -) behave otherwise: the absolute values of errors are mainly decreasing by $\beta < 3$ and increasing by $\beta > 3$. In the last case in numerical experiments the convergence rate is lower than theoretically expected.

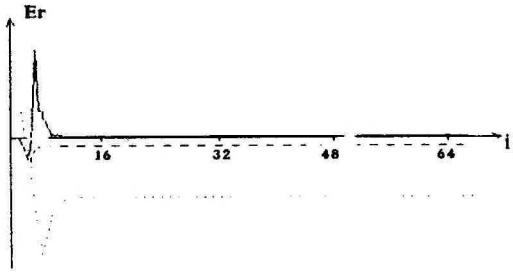


Figure 1. The errors on subintervals obtained with the present method.

The errors $I(g) - \tilde{I}(g)$ are condensed in the given Table 1.

N	8	16	32	64	128	256
trapezoidal rule	0.066	0.017	0.0042	0.0010	0.00026	0.000065
after correction	0.127	0.018	0.0024	0.00030	0.000038	0.0000048

Table 1. Real errors.

on first subintervals. This is caused by the fact that for small i the coefficients k_i are relatively large (see Table 2) and therefore

$$(K^{d}_{2,1}\cdot(1-k_1)+K^{s}_{2,1}\cdot k_1)\cdot f'''(x^{N}_{2i+1})$$

is also relatively large.

i=	1	2	3	4	7	15	31	63
α=0	-0.48	-0.39	-0.36	-0.35	-0.34	-0.334	-0.334	-0.3334
α=0.5	-1.42	-0.68	-0.51	-0.44	-0.37	-0.342	-0.335	-0.3338
α=0.7	-5.37	-1.66	-0.95	-0.69	-0.46	-0.361	-0.340	-0.3350

Table 2. Coefficients k_i by $\beta = 3$.

Recommendation. We advise to compute without correction on these subintervals where the inequality

$$|f'''(x^{N}_{2i+1})K^{d}_{1,1}| \le |K^{d}_{2,1}(1-k_1)+K^{s}_{2,1}\cdot k_1|\,|f'''(x^{N}_{2i+1})|$$

is valid. By using the assumption (2) it is possible to see that the number of such intervals is finite. For our functions this number depended only on α and β and is present in Table 3 by β = 3 (denoted with k).

α	0 - 0.05	0.06 - 0.34	0.35 - 0.5	0.51 - 0.59	0.6 - 0.66	0.67 - 0.7
k	2	4	6	8	10	12

Table 3. The number of subintervals without correction.

The numerical results received with this modified method are presented in Table 4 and in Figure 2.

N	4	8	16	32	64	128	256
trapezoidal rule	0.252	0.066	0.017	0.0042	0.0010	0.00026	0.000065
after modified correction	0.252	0.038	.0028	.00023	.000021	.0000022	.00000025

Tabel 4. Real errors obtained with modified method.

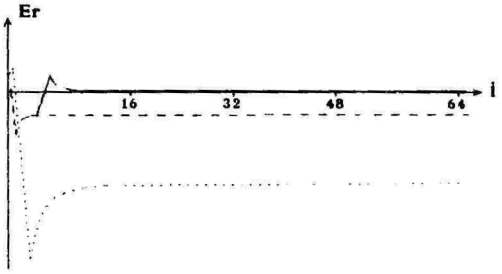


Figure 2. The errors on subintervals obtained with modified method.

References

1. Kõppo, K. Nõrgalt singulaarse integraali arvutamine. Kursusetöö. Tartu 1991. (Term-paper in Estonian).
2. Opfer, G. Evaluation of weakly singular integrals. In: Numerical treatment of integral equations. Basel. 1980, p.191-202.
3. Rice J.R. On the degree of convergence of nonlinear spline approximation. In: Approximations with special emphasis on spline functions. Academic Press, New York, 1969.
4. Бахвалов Н.С. Численные методы. Том 1. М.: Наука, 1973.
5. Крылов В.И. Приближенное вычисление интегралов. М.: Наука, 1967.
6. Уба, П. Численное интегрирование функций со слабой особенностью. В сб.: V Всесоюзный симпозиум "Метод дискретных особенностей в задачах математической физики." Тезисы докладов. Часть II Одесса 1991. с.59.

ПРИБЛИЖЕННОЕ ВЫЧИСЛЕНИЕ СЛАБО СИНГУЛЯРНЫХ ИНТЕГРАЛОВ

П.Уба

Резюме

Проблемы приближенного вычисления интеграла

$$\int_0^b g(x)dx, \qquad |b| < \infty,$$

где функция $g(x)$ имеет в точке $x=0$ интегрируемую особенность, изучались многими авторами (см., например [2,4,5]). При "явно" заданных особенностях ($g(x)=g_1(x)\log(x)$ или $g(x)= g_1(x)x^{\alpha}$, $-1 < \alpha < 0$, $g_1(x)$ - достаточно гладкая функция) авторы предлагают или воспользоватся неравномерными сетками (см. [4]) или построить соответствующие квадратурные формулы Гаусса ([5]).

В данной статье исследуется случай, когда о функции $g(x)$ известно лишь то, что

$$|g^{(k)}(x)| \le c|x|^{-\alpha-k}, \quad 0<x\le b, \quad \alpha<1, \quad k=0,1,2,3,4,$$

или

$$|g(x)| \le c|\ln x|, \quad |g^{(k)}(x)| \le c|x|^{-k}, \quad 0<x\le b, \quad k=1,2,3,4.$$

Для приближенного нахождения таких интегралов предлагается применить формулы трапеции на сильно неравномерных сетках с последовательным уточнением по Рунге . Доказано, что при правильном выборе параметра неравномерности сетки достижима сходимость метода четвертого порядка. Применены численные эксперименты, на базе которых выяснилась необходимость некоторой модификации предложенного алгоритма.

Acta et commentationes universitatis Tartuensis, 937, 1992 , 36-46.

COLLOCATION METHOD FOR FINDING PERIODIC SOLUTIONS OF QUADRATIC AUTONOMOUS SYSTEMS

Peep Miidla

One reason which complicates the numerical study of solutions of differential equations and systems is the need to verify the theoretical and qualitative conditions of convergence theorems. In this paper we give the conditions of convergence of collocation and Galerkin methods for finding periodical solutions of autonomous systems. Those conditions are "less qualitative" and so more easily controllable with a computer in the concrete problems case. The main results touch upon the systems with quadratic right-hand terms.

1. Preliminary assumptions and results. We consider the problem of finding non-trivial periodic solutions of autonomous system

$$Z' = F(Z), \tag{1}$$

where $Z=(z_1,\dots,z_m)$. Let $F(Z)=(f_1(Z),\dots,f_m(Z))$ be a given vector-function of class C^1 (i.e. functions $f_i(z_1,\dots,z_m)$, $i=1,\dots,m$ are continuous and continuously differentiable by $z_1,\dots,z_m$) at least in the regions described below. Assume that system (1) has a periodic solution $Z^*=Z^*(t)=(z_1^*(t),\dots,z_m^*(t))$ with some less period $\omega^*>0$, i.e. $Z^*(t)=Z^*(t+\omega^*)$ $\forall t$ and there does not exist ω, $0<\omega<\omega^*$, with the same property. We need some more precise information about Z^*, namely, assume that the number α from the range of first component of Z^* is known such that

$$\exists t_\alpha : z_1^*(t_\alpha)=\alpha \quad \text{and} \quad z_1^{*\prime}(t_\alpha)\neq 0. \tag{2}$$

Without restricting generality we may take $t_\alpha = 0$. Further we introduce other presumptions about system (1).

Now for the convenience of reading we recall some known results in the form corresponding to our aims. At first, we need the convergence theorem for the equations with regularly converging

operators from [3].

Theorem 1. Let two Banach spaces E and F be given. Consider operators $A: \Omega \subset E \to F$ and $A_n: \Omega_n \subset E \to F$ (Ω and Ω_n are open sets, $n \in \mathbb{N} = \{1,2,3,4,\ldots\}$) for which the following conditions be fulfilled.

(i) The equation $Au = 0$ has a solution $u^* \in \Omega$ and the operator A is Frechet differentiable at u^*.

(ii) There is a positive δ such that the operators A_n ($n \in \mathbb{N}$) are Frechet differentiable in the ball $\|u - u^*\|_E \le \delta$, and for any $\varepsilon > 0$ there is a δ_ε ($0 < \delta_\varepsilon < \delta$) such that, for every $n \in \mathbb{N}$,

$$\|A_n'(u) - A_n'(u^*)\| \le \varepsilon \quad \text{whenever} \quad \|u - u^*\|_E \le \delta_\varepsilon.$$

(iii) $\|A_n u^*\|_F \to 0$ ($n \in \mathbb{N}$).

(iv) (The condition of regular convergence of Frechet derivatives.)

$$\|A_n'(u^*)\| \le \text{const} \quad (n \in \mathbb{N});$$

$$\|[A_n'(u^*) - A'(u^*)]u\| \to 0 \quad \forall u \in E \quad (n \in \mathbb{N});$$

$\|u_n\|_E \le \text{const}$, sequence $([A_n'(u^*)]u_n)$ is compact in F $\Rightarrow$

$\Rightarrow$ sequence (u_n) is compact in E.

(v) $A_n'(u^*)$ ($n \in \mathbb{N}$) are Fredholm operators of index zero, i.e. $\dim \mathcal{N}(A_n'(u^*)) = \operatorname{codim} \mathcal{R}(A_n'(u^*))$ and $\mathcal{R}(A_n'(u^*))$ is closed. Here $\mathcal{N}(\cdot)$ and $\mathcal{R}(\cdot)$ denote the zero space and the range of corresponding operators.

(vi) $\mathcal{N}(A'(u^*)) = \{0\}$.

Then there exist $n_0 \in \mathbb{N}$ and δ_0 ($0 < \delta_0 \le \delta$) such that the equation $A_n u = 0$ has for $n \ge n_0$ a unique solution u_n^* in the ball $\|u - u^*\|_E \le \delta_0$. Besides $u_n^* \to u^*$ with an error estimate

$$c_1 \|A_n u^*\|_F \le \|u_n^* - u^*\|_E \le c_2 \|A_n u^*\|_F , \qquad (3)$$

where c_1 and c_2 are positive constants.

Let us consider system (1) again. Condition (2) makes it possible to define a Poincaré mapping in some open neighbourhood U of point $Z^*(0) \in \Sigma$ on the hyperplane $\Sigma = \{Z \in \mathbb{R}^m : z_1 = \alpha\}$. We shall use the following notations for the neigbourhoods of point $Z^0 = (z_1^0, z_2^0, \ldots, z_m^0) = (\alpha, z_2^0, \ldots, z_m^0) \in \Sigma$:

$$W_\varrho(Z^0) = \left\{ Z \in \mathbb{R}^m : \sum_{i=2}^{m} (z_i - z_i^0)^2 < \varrho^2,\ |z_1 - \alpha| < \varrho \right\},$$

$$U_\varrho(Z^0) = W_\varrho(Z^0) \cap \Sigma.$$

Now U can be considered as $U_\varrho(Z^*(0))$ with some $\varrho > 0$. We also use the notations

$$\|Z\|_{\mathbb{R}^m} = \left\{ \sum_{i=1}^{m} z_i^2 \right\}^{1/2}$$

and $P: U_\varrho(Z^*(0)) \to \Sigma$ for (m-1-dimensional) Poincaré mapping.

As it is known, the Poincaré mapping is usually introduced by the one-parameter group of shifts $\{S_t\}$ along the trajectories of system (1). So, by the definition, $P(Z^*(0)) = S_{\omega^*}(Z^*(0))$. Moreover, for every point $Z \in U_\varrho(Z^*(0))$ there exists the least "receiving time" to Σ which we denote by $\omega = \omega(Z)$. So, $P(Z) = S_\omega(Z)$ for $Z \in U_\varrho(Z^*(0))$. Further, we can put

$$Y = Z - Z^*(0), \quad Q(Y) = P(Z) - Z^*(0), \quad Z \in U_\varrho(Z^*(0)).$$

As F is a C^1-function, the expanding

$$Q(Y) = Y^0 + LY + Q_1(Y) \tag{4}$$

is possible in the neighbourhood of origin,

$$Y^0 = P(Z) - Z^*(0), \quad Z \in U_\varrho(Z^*(0)).$$

Here L denotes the linear part of Poincaré mapping. We now can demand that there exist positive constants ϱ_0 and k_0 such that for arbitrary Y', Y'', $\|Y'\|_{\mathbb{R}^m} \le \varrho$, $\|Y''\|_{\mathbb{R}^m} \le \varrho$, $\varrho \le \varrho_0$,

$$\|Q_1(Y') - Q_1(Y'')\|_{\mathbb{R}^m} \le k_0 \cdot \varrho \|Y' - Y''\|_{\mathbb{R}^m}. \tag{5}$$

The treatment above is completely nonconstructive , still we do not know $Z^*(0) \in \Sigma$. But, due to [2], the same treatment is obtained if instead of $Z^*(0)$ in expansion (4) and inequality (5) we use arbitrary (in the sense of the next theorem below) point $Z^0 \in \Sigma$.

Theorem 2. Let $S_\omega(Z^0) \in \Sigma$ $\|S_\omega(Z^0) - Z^0\|_{\mathbb{R}^m} = \varepsilon$ and for some $\bar{\varrho}_0 \le \varrho_0$ the next inequality holds:

$$\|(L - I)^{-1}\| \left(\frac{\varepsilon}{\bar{\varrho}_0} + k_0 \cdot \bar{\varrho}_0 \right) \le 1, \tag{6}$$

where L is the linear part of expanding (4) for the point Z^0, I is unity transformation and ϱ_0, k_0 are the constants from (5). Then in

the $\bar{\varrho}_0$ - neighbourhood of Z^0 there exists one and only one fixed point of Poincaré mapping P.

For proof see [2]. This theorem, naturally, has sense only whenever expanding (4) is correct, i.e. if the Poincaré mapping is defined in $U_{\varrho_0}(Z^0)$.

2. Discovery of the fixed point of Poincaré mapping. There is another interesting result in [2]. Namely, the estimation of constant k_0 in the case where the terms on the right-hand side of system (1) are polynomials of power not higher than two. The following estimate is obtained:

$$k_0 \le \tilde{k}_0 = m\cdot\Big[b_1\cdot\Big(c_1^2+\frac{1}{2}\Big) + c_1\cdot c_5(c_1+b_1\cdot\varrho) + c_7 + \\ + \frac{c_4}{c_3}\left(c_1\cdot c_5\cdot c_9\cdot\sqrt{m}+c_8\cdot c_9+b_2\right) + c_1\cdot c_5\cdot c_9\cdot\sqrt{m}+c_8\cdot c_9\Big], \tag{7}$$

where

$$b_1 = 2c_1^3 c_2\cdot\omega_1, \quad b_2 = 2c_1\cdot c_2\Big(c_1^2+\frac{1}{2}\Big)\cdot\omega_1,$$

$$c_7 = \frac{c_1\cdot c_8}{c_3}\left(c_4+c_6+\frac{c_4\cdot c_6}{c_3}\right)(c_1+b_1\cdot\varrho),$$

$$c_8 = c_5\cdot b_1\cdot\varrho + \frac{1}{2}c_1\cdot c_2\cdot\varrho + c_1\cdot c_2\cdot b_1\cdot\varrho^2 + \frac{1}{2}b_1 c_2\varrho^3,$$

$$c_9 = \frac{c_1+2b_1(2c_1^2+1)}{c_3-\varrho(c_1\cdot c_5+c_8\varrho)}.$$

The constants $c_1, c_2, \ldots, c_6$ depend upon the system in the neigbourhood of point $Z^0\in\Sigma$ for which $S_\omega Z^0\in\Sigma$, m is the dimension of the system and ϱ is the parameter of neighbourhood. Denote by $\Phi(t,t_0)$ the fundamental matrix of solutions of the linearized system

$$Y' = F'(S_t Z^0)Y \tag{8}$$

such that $\Phi(t_0,t_0)=1$.
Then

$$c_1 = \max_{0\le t_0\le t_1\le\omega_1}\|\Phi(t_1,t_0)\|,$$

where spectral norm is considered. In [2] the algorithm of effective computing of c_1 is given. For quadratic right-hand terms other constants are straightly computable:

$$c_2 : \sum_{i,j,k} \left| \frac{\partial^2 f_i}{\partial z_j \partial z_k} y_j z_k \right| \le c_2 \cdot \|Y\|_{R^m} \cdot \|Z\|_{R^m} ;$$

$$c_3 = \inf_{Z \in W_Q(Z^0)} |f_1(Z)| ; \qquad c_4 = \max_{Z \in W_Q(Z^0)} \max_i |f_i(Z)| ;$$

$$c_5 = \max_{Z \in W_Q(Z^0)} \|F'(Z)\| ; \qquad c_6 = \max_{Z \in W_Q(Z^0)} \|F(Z)\|_{R^m}.$$

Note that the method of the estimation of ε and ω_1 in (6) or in (7) is also shown in [2]. This method is based on the construction of pseudo-trajectory with the finite-difference method. The same result might be obtained via collocation method or the Galerkin method but this is not a matter under discussion in this paper. We summarize this referred result in the following way.

Let the right-hand side terms of system (1) be the polynomials of power not higher than two and let the assumptions about (1) made in Part 1 of this paper be fulfilled. Then to verify the existence of the unique fixed point of Poincaré mapping in $U_Q(Z^0)$ for some point $Z^0 \in R^m$ it is enough to compute the value of $\tilde{k}_0$ in (7) and then check up inequalities (5) and (6). For details of the numerical treament see [2].

The uniqueness of the fixed point of Poincaré mapping in some open set implies isolatedness of the closed trajectory passing through this point. Denote by Z_Q^0 the unique fixed point of Poincaré mapping in $U_Q(Z^0)$, if only such exists; $Z_Q^0 \in \Sigma$. Since for distinct points Z^0 the fixed point Z_Q^0 may be the same, we leave out upper index zero; so, Z_Q would be a fixed point of Poincaré mapping, for which there exists $Z^0 \in \Sigma$ such that Z_Q is the unique fixed point of Poincaré mapping in $U_Q(Z^0)$. Fix now an opened region $\Pi \subset \Sigma$ and denote by $V(Q,\Pi)$ the set of all points $Z_Q \in \Pi$ (for arbitrarly fixed $Q>0$). It does not follow from our assumptions about system (1) that sets $V(Q,\Pi)$ are not empty for every or some $Q>0$ even if $Z^*(0) \in \Pi$. The reason is that we do not know whether $Z^*(0)$ is an isolated fixed point of (corresponding) Poincaré mapping or not. But if we find Z^0 for which Theorem 2 is fulfilled with some $Q=Q_0$ and $U_{Q_0}(Z^0) \subset \Pi$ then sets $V(Q,\Pi)$ are not empty for every $Q \le Q_0$.

The exact description of sets $V(Q,\Pi)$ would be an interesting problem to solve further. It is clear that the numerical treatment mentioned above is of the character of trial and error and so it is not enough to satisfy those purposes, at least theoretically.

We now assume that $f_1(Z) \neq 0$ for $Z \in \Pi$. It makes us sure that among the points of $V(\varrho,\Pi)$ there are no fixed points corresponding to trivial periodic solutions of system (1) or to the so-called stable points of system (1). This assumption is not a real restriction because we are free in the choice of Π. If for some $\varrho = \varrho_0$ the set $V(\varrho_0,\Pi)$ is not empty then for every $Z \in V(\varrho,\Pi)$ we may consider the least returning time of the trajectory $S_t Z$ to Σ, i.e. the period $\omega(Z)$ of the corresponding solution. The fundamental matrix $\Phi(t,0)$ of linearized system $Y' = F'(S_t Z)Y$ determines the multiplicators of system (1) corresponding to periodical solution $S_t Z$; we recall that multiplicators are defined as eigenvalues of $\Phi(\omega(Z),0)$. It is known that among those there is the simple multiplicator equal to 1, or, the equivalent statement, the corresponding linearized system has exactly one-dimensional space of $\omega(Z)$-periodical solutions spanned on $\frac{d}{dt}(S_t Z)$.

3. Convergence of the collocation method. Consider (1) in assumptions made in the beginning of Section 1. It is convenient to make the change of variable $t \to \lambda t$, where $\lambda = 2\pi/\omega$ and ω is the new variable for unknown period of Z^*. The system (1) takes form

$$\lambda X' = F(X), \tag{9}$$

where 2π-periodic function and the value of parameter λ are to be determined . Note that to ω^*-periodic solution Z^* of system (1) corresponds to the pair $\{X^*,\lambda^*\}$ of solutions of (9), where X^* is 2π-periodic and $\lambda^* = 2\pi/\omega^*$; $X^* = (x_1^*,\dots,x_m^*)$.

Introduce the following Banach spaces.

H^1 - space of absolutely continuous 2π-periodic vector-functions with equivalent norms:

$$\|X\|_{H^1} = \left(\|X\|^2_{L^2(0,2\pi;\mathbb{R}^m)} + \|X'\|^2_{L^2(0,2\pi;\mathbb{R}^m)}\right)^{1/2},$$

$$\|X\|'_{H^1} = \|X\|_{C(0,2\pi;\mathbb{R}^m)} + \|X'\|_{L^2(0,2\pi;\mathbb{R}^m)}.$$

$H^0 = L^2(0,2\pi;\mathbb{R}^m)$ where the functions of $L^2(0,2\pi;\mathbb{R}^m)$ we take as continued in 2π-periodic,

$$\|Y\|_{H^0} = \left(\sum_{i=1}^{m}\left[\int_0^{2\pi} y_i(s)\,ds\right]^2\right)^{1/2}.$$

$E = H^1 \times \mathbb{R}$ and $F = H^0 \times \mathbb{R}$ with norms

$$\|(X,\lambda)\|_E = \|X\|_{H^1} + |\lambda|,$$

$$\|(Y,\mu)\|_F = \|Y\|_{H^0} + |\mu|.$$

Below the simple notations C and L^2 instead of $C[0,2\pi;\mathbb{R}^m]$ and $L^2[0,2\pi;\mathbb{R}^m]$ respectively will be used.

We intend to find the approximation to X^* in the form $X_n(t) = (x_{1n}(t), \ldots, x_{mn}(t))$, where the components are trigonometrical polynomials,

$$x_{in}(t) = \frac{c_{0i}}{2} + \sum_{k=0}^{n} (c_{ki}\cos kt + d_{ki}\sin kt),\ i=1,\ldots,m.$$

For this we use the method of collocations

$$\begin{cases} \lambda X_n'(t_i) = F(X_n(t_i)),\ i=0,1,\ldots,2n,\ t_i = ih,\ h = 2\pi/(2n+1), \\ x_{in}(0) = \alpha; \end{cases} \tag{10}$$

or the Galerkin method

$$\begin{cases} \lambda \int_0^{2\pi} e_i(t) X_n(t)dt = \int_0^{2\pi} e_i(t) F(X_n(t))dt,\ i=0,\ldots,2n, \\ x_{in}(0) = \alpha, \end{cases} \tag{11}$$

where $e_{2k}(t) = \cos kt$, $e_{2k+1}(t) = \sin kt$.

The equivalent presentation of problems (9) and (10) we give in the form of operator equations

$$A(X,\lambda) = 0 \tag{12}$$

and

$$A_n(X_n,\lambda_n) = 0, \tag{13}$$

where operators $A: E \to F$ and $A_n: E \to F$ are in the forms

$$A(X,\lambda) = \begin{pmatrix} \lambda X' - F(X) \\ x_1(0) - \alpha \end{pmatrix} \tag{14}$$

and

$$A_n(X,\lambda) = \begin{pmatrix} \lambda P_n X' - P_n F(X) \\ x_1(0) - \alpha \end{pmatrix}. \tag{15}$$

Here P_n is the Lagrange projector, which corresponds to the interpolation of functions with trigonometric polynomials on the uniform set. Analogous presentation is obtained for the Galerkin method (11). The corresponding operator equation is (13) where in the definition (15) of operator A_n instead of P_n stays orthoprojector $\mathcal{O}_n$ which corresponds to the approximation of functions with finite Fourier sums. In [4] the properties of P_n and $\mathcal{O}_n$ are completely investigated; for us the following conditions are important:

$$\|P_n\|_{C\to L^2} \le \text{const};$$

$$\|P_n X - X\|_{L^2} \to 0 \quad \forall X \in C, \quad n \to \infty; \tag{16}$$

$$\|\mathcal{O}_n\|_{L^2\to L^2} \le \text{const};$$

$$\|\mathcal{O}_n X - X\|_{L^2} \to 0 \quad \forall X \in L^2, \quad n \to \infty.$$

Further we consider only the case of the collocation method because $\mathcal{O}_n$ satisfies more stronger conditions than we shall use for P_n.

Prove now that if the mentioned assumptions about system (1) hold, then for operators A and A_n ($n \in \mathbb{N}$) defined by (14) and (15) correspondingly the presumptions (i)-(v) of Theorem 1 are true.

(i) The pair (X^*, λ^*) is the demanded solution of (12) in $E = H^1 \times \mathbb{R}$ and A' is in the form:

$$\left[A'(X^*,\lambda^*)\right](Y,\mu) = \begin{pmatrix} \lambda Y' - F'(X^*(t))Y + \mu \cdot X^{*\prime}(t) \\ y_1(0) \end{pmatrix}.$$

(ii) The derivative of A_n is the following:

$$\left[A_n'(X,\lambda)\right](Y,\mu) = \begin{pmatrix} \lambda Y' - P_n F'(X(t))Y + \mu \cdot P_n X' \\ y_1(0) \end{pmatrix}.$$

This is a continuous operator because F is differentiable. Establish the second demanded condition:

$$\left\|\left[A_n'(X,\lambda) - A_n'(X^*,\lambda^*)\right](Y,\mu)\right\|_F =$$

$$= \left\|(\lambda - \lambda^*)Y' - P_n\left[F'(X(t)) - F'(X^*(t))\right]Y + \mu P_n(X'(t) - X^{*\prime}(t))\right\|_{H^0} \le$$

$$\le |\lambda - \lambda^*| \cdot \|Y'\|_{L^2} + \|P_n\|_{C\to L^2} \cdot \left\|\left[F'(X(t)) - F'(X^*(t))\right]Y\right\|_C +$$

$$+ |\mu| \cdot \|P_n\|_{C\to L^2} \cdot \|X' - X^{*\prime}\|_C \le \varepsilon \cdot \left(\|Y\|_{H^1}' + |\mu|\right).$$

The last inequality holds whenever $\|X - X^*\|_C + |\lambda - \lambda^*| \le \delta$. Equivivalence of norms $\|\cdot\|_1$ and $\|\cdot\|_1'$ gives us

$$\varepsilon = \text{const} \cdot \max\{|\lambda - \lambda^*|, \|P_n\|_{C\to L^2} \cdot \max_k \|f_k'(X) - f_k'(X^*)\|_C,$$

$$\|P_n\|_{C\to L^2} \cdot \|X' - X^{*\prime}\|_C\}.$$

(iii) We have

$$\|A_n(X^*,\lambda^*)\|_F = \|A_n(X^*,\lambda^*) - A(X^*,\lambda^*)\|_F =$$

$$= \|\lambda^* \cdot (P_n X^{*\prime} - X^{*\prime}) - (P_n F(X^*) - F(X^*))\|_{L^2} \le$$

$$\leq |\lambda^*| \cdot \|P_n X^{*\prime} - X^{*\prime}\|_{L^2} + \|P_n F(X^*) - F(X^*)\|_{L^2} \to 0$$

when $n \to \infty$ because of (16).

(iv) Compute

$$\|[A_n'(X^*,\lambda^*)](Y,\mu)\|_F = \|\lambda^* Y' - P_n F(X^*)Y + \mu P_n X^{*\prime}\|_{L^2} + |y_1(0)| \leq$$

$$\leq |\lambda^*| \cdot \|Y'\|_{L^2} + \|P_n\|_{C \to L^2} \cdot \|F(X^*)\|_C \cdot \|Y\|_C + |\mu| \cdot \|P_n\|_{C \to L^2} \cdot \|X^{*\prime}\|_C +$$

$$+ \|Y\|_C \leq M \cdot (\|Y\|_{H^1}' + |\mu|).$$

$$\|[A_n'(X^*,\lambda^*) - A'(X^*,\lambda^*)](Y,\mu)\|_F = \|P_n F'(X^*)Y - F'(X^*)Y +$$

$$+ \mu \cdot (P_n X^{*\prime} - X^{*\prime})\|_{L^2} \leq \|P_n F'(X^*)Y - F'(X^*)Y\|_{L^2} +$$

$$+ |\mu| \cdot \|P_n X^{*\prime} - X^{*\prime}\|_{L^2} \to 0 \quad \text{because of (16).}$$

Let $\|Y_n\|_{H^1} + |\mu_n| \leq \text{const}$ and the sequence $([A_n'(X^*,\lambda^*)](Y_n,\mu_n))$ be compact in F. The last means, indeed, compactness in L^2 of the sequence $(\lambda^* Y_n' - P_n F'(X^*)Y_n)$. The boundedness of (Y_n) in H^1 implies compactness in C of $(P_n F'(X^*)Y_n)$ because of (16) and our main assumptions about (1). Finally, the assumption about compactness of $([A_n'(X^*,\lambda^*)](Y_n,\mu_n))$ implies necessary result: (Y_n) is compact in H^1 and so $\{Y_n,\mu_n\}$ is compact in E.

(v) The operator $B = \lambda^* \cdot \frac{d}{dt} : H^1 \to H^0$ is a Fredholm operator with index zero. By the way, $\dim \mathcal{N}(B) = \operatorname{codim} \mathcal{R}(B) = m$. Other components of $A_n'(X^*,\lambda^*)$ are finite-dimensional and so the operator $A_n'(X^*,\lambda^*)$ is a Fredholm operator with index zero as well (see also [1,3]).

We saw that prepositions (i)-(v) of Theorem 1 are fulfilled for operator equations (12) and (13). Only the (vi) is "absent" for the application of the whole result. In [1] analogous treatment was realized for the equations of order m but not for systems and there condition (vi) was taken as a previous assumption. Now we formulate the main result.

Theorem 3. Consider system (1) in the following assumptions.

1° Terms of the right-hand side of (1) are polynomials of power not higher than two.

2° For some $Z^0 \in \mathbb{R}^m$ Theorem 2 is true with some $\varrho = \varrho_0$ and k_0

given in (7).

3^0 $f_1(Z) \neq 0$ for $Z \in U_{\varrho}(Z^0)$.

Then there exist $n_0 \in \mathbb{N}$ and δ_0 such that the collocation method (10) with $\alpha = z_1^0$ determines for $n \geq n_0$ a unique solution (X_n^*, λ_n^*) such that the trajectory $X_n^*(t)$ passes $U_{\varrho}(Z^0)$ on the plane $\Sigma = \{Z \in \mathbb{R}^m : z_1 = z_1^0\}$. Besides $\|X_n^* - X^*\|_{H^1} \to 0$, $|\lambda_n^* - \lambda^*| \to 0$ $(n \in \mathbb{N})$ where (X^*, λ^*) will be the solution of problem (9). The convergence speed is

$$\|X_n^* - X^*\|_{H^1} + |\lambda_n^* - \lambda^*| \leq c \cdot \|P_n X^{*\prime} - X^{*\prime}\|_{L^2}. \qquad (17)$$

Proof. Consider the problem of finding the periodical solution of (1) and collocation method (10) in the form of operator equations (12) and (13) correspondingly and show that the presumptions of Theorem 1 hold.

From 1^0 it follows that F is differentiable and 2^0 implies the existence of solution $(X^*, \lambda^*) = (S_{t/\lambda^*} Z^0_{\varrho_0}, 2\pi/\omega(Z^0_{\varrho_0}))$ (cf. Section 2) of equation (12).

As proved above, the conditions (i)-(v) of Theorem 1 are fulfilled, at that also in open set

$$\Omega = \{(X, \lambda) \in E : X(0) \in U_{\varrho_1}(Z^0_{\varrho_0}), \lambda_1 < \lambda < \lambda_2\}$$

with some real positive ϱ_1, λ_1 and λ_2 such that

$$U_{\varrho_1}(Z^0_{\varrho_0}) \subset U_{\varrho_0}(Z^0) \subset \Sigma, \quad \lambda_1 < \frac{2\pi}{\omega(X(0))} < \lambda_2.$$

According to the discussion of Section 2 condition (vi) of Theorem 1 follows from 2^0 and 3^0. The application of the whole Theorem 1 with $\Omega_n = \Omega$ $(n \in \mathbb{N})$ gives us our statements. Prove the estimation (17). From (3) we get

$$\|X_n^* - X^*\|_{H^1} + |\lambda_n^* - \lambda^*| \leq c_2 \|A_n(X^*, \lambda^*)\|_F =$$

$$= c_2 \|A_n(X^*, \lambda^*) - A(X^*, \lambda^*)\|_F = c_2 \|\lambda^* P_n X^{*\prime} - \lambda^* X^{*\prime} + P_n F(X^*) - F(X^*)\|_{L^2} =$$

$$= c_2 \|2\lambda^* (P_n X^{*\prime} - X^{*\prime})\|_{L^2} = c \cdot \|P_n X^{*\prime} - X^{*\prime}\|_{L^2},$$

Q.E.D.

Analogical result can be formulated for the Galerkin method (11). Then orthoprojectors O_n must be taken instead of Lagrange projectors P_n in proofs.

References

1. Miidla P. About Finding the Periodic Solutions of Autonomous Differential Equations and Systems. Conference del Seminario di Matematica dell'Universita' di Bari. N0.199, 1984, pp.1-23.
2. Sinai J.G., Vul E.B. Discovery of Closed Orbits of Dynamical Systems with the Use of Computers. Journal of Statistical Physics, Vol.23, No.1, 1980, pp.27-47.
3. Vainikko G. Approximative Methods for Nonlinear Equations (Two approaches to the Convergence Problem). Nonlinear Analysis, Theory, Methods & Applications. Vol.2, No.6, 1978, pp.647-687.
4. A.Zygmund. Trigonometric Series, Vol.I - II. Cambridge, Univ.-Press, 1959.

МЕТОД КОЛЛОКАЦИИ ДЛЯ НАХОЖДЕНИЯ ПЕРИОДИЧЕСКОГО РЕШЕНИЯ КВАДРАТИЧНОЙ АВТОНОМНОЙ СИСТЕМЫ

Пээп Мийдла

Резюме.

В статье исследуются условия сходимости методов коллокации и Галеркина нахождения периодического решения автономной системы (1). Доказана теорема 3 сходимости методов, в которой требование о простоте мултипликатора равной единице системы (1) вытекает как следствие. При этом предполагается,что правые части системы (1) - суть многочлены степени не высше двух. Новым требованием является выполненность теоремы 2 с оценочной константой (7). Численный алгоритм проверки неравенства (6) приведен в [2]. Доказательство опирается на абстрактной теореме сходимости из [3].

Acta et commentationes universitatis Tartuensis, 937, 1992 , 47-62.

ON THE CONCEPT OF DISCRETE CONVERGENCE IN THE CASE OF NORMED SPACES

Otto Karma

In this paper one possible concretization of F.Stummel's scheme of discrete approximation for normed spaces is considered . The concepts of consistency, closedness, stability, $\mathcal{BR}$-stability, regularity and $\mathcal{B}$-regularity of the sequences of operators at some point v are discussed . The behaviour of the sequences of the solutions and approximate solutions of approximating equations is examined.

1. Introduction. Let us consider the equation $Au = v$ with the operator $A: U \to V$. In many cases, to get the approximate solution of this equation, the following approach is applied. The operator A is replaced by the other operator B_i, which is in some sense close to A (approximating A), so that the new problem is easier to investigate. The solution of the new problem is then considered to be an approximation of the solution of the original problem. At that it is often reasonable to choose the operator B_i acting from X_i to Y_i with X_i and Y_i different from U and V, respectively. In addition it is, mostly, natural to look at the operator B_i as at a member of the family (B_i) of operators which are approaching, in some sense, to A.

Various schemes for connecting the spaces X_i and U (Y_i and V) can be used. A very general scheme of the discrete convergence was developed in [1, 2]. For normed spaces U, X_i, $i \in \mathbb{N}$ one example of discrete convergence is as follows: the sequence of operators $r_i: U \to X_i$, $i \in \mathbb{N}$ is fixed and the convergence $x_i \to u$ is defined as the convergence $\|x_i - r_i u\|_{X_i} \to 0$ ($i \in \mathbb{N}$). Moreover, for normed spaces any reasonable discrete convergence can be presented this way.

Nevertheless, in some cases it may appear to be more convenient for concrete applications, as well as for developing abstract theory, to use also other schemes of discrete approximation.

Here we propose the following scheme. Let for every X_i, $i \in \mathbb{N}$ the normed space U_i and "connecting" operators $p_i^U : U \to U_i$, $p_i^X : X_i \to U_i$ (prolongations, projectors, embeddings, restrictions, ...) be fixed. We say that $x_i \to u$ if $\|p_i^X x_i - p_i^U u\|_{U_i} \to 0$ $(i \in \mathbb{N})$.

In this scheme the abstract theory of approximation methods can be developed analogically to the case when the convergence is defined by the operators $r_i : U \to X_i$ (look e.g. [3]). The advantage of the scheme, proposed here, is that that it is more natural for some practical applications and that, in some cases, naturally defined connecting operators p_i^U, p_i^X are linear and bounded whereas naturally defined operators $r_i : U \to X_i$ are not such. For example, in the case of projection methods we can take $U_i = U$, $p_i^U = I$, $X_i \subset U$, p_i^X embedding operators, $V_i = Y_i = p_i^V V$ with $p_i^Y = I$ and with p_i^V discretization operators defined by the concrete method $(B_i = p_i^V A p_i^X)$.

In the following we shall denote by $\mathbb{N}$ the sequence of natural numbers and by $N, N', \ldots$ its infinite subsequences. The notation $\alpha_i \to \alpha$ $(i \in \mathbb{N})$ with α, α_i real numbers means that the sequence $(\alpha_i)_{i \in \mathbb{N}}$ converges to α. The notation $\alpha_i \le c$ $(i \in \mathbb{N})$ means that $\alpha_i \le c$ for all $i \in \mathbb{N}$ with constant c, which is independent on i. As a rule, spaces are marked by capital letters (with subindices) and their elements by corresponding small letters (with upper indices), i.e. $u, u', \ldots \in U$, $x_i, x_i', \ldots \in X_i$ etc. We shall often omit the record $(i \in \mathbb{N})$ if it is clear which sequence of indices is considered, especially in the proofs.

2. **Convergence of sequences (x_i) with $x_i \in X_i$** . Let U, U_i, X_i $(i \in \mathbb{N})$ be normed spaces over the same field of constants $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ and let $P^U = \{p_i^U : U \to U_i\}_{i \in \mathbb{N}}$, $P^X = \{p_i^X : X_i \to U_i\}_{i \in \mathbb{N}}$ be two systems of "connecting" operators . We call a sequence $(x_i)_{i \in \mathbb{N}}$ of elements $x_i \in X_i$, $i \in \mathbb{N}$ converging (or , more exactly, discretely $P^U P^X$-converging) to a limit element $u \in U$ if $\|p_i^X x_i - p_i^U u\|_{U_i} \to 0$ $(i \in \mathbb{N})$; we write it $x_i \to u$ $(i \in \mathbb{N})$.

By this definition, it is clear that

$$x_i \to u \quad (i \in \mathbb{N}) \iff x_i \to u \quad (i \in N') \quad \forall N' \subset \mathbb{N}.$$

We call a sequence $(x_i)_{i \in \mathbb{N}}$ of elements $x_i \in X_i$, $i \in \mathbb{N}$ (discretely) compact if its every subsequence contains a convergent subsequence:

$$\forall N' \subset \mathbb{N} \quad \exists N'' \subset N', \quad u \in U : x_i \to u \quad (i \in N'').$$

It is clear that every convergent sequence is compact and that every subsequence of a compact sequence is also compact. It is easy to verify by reductio ad absurdum proof that if a compact sequence has only one cluster point u then this sequence is converging to u.

Further we assume that the systems P^U and P^X are asymptotically linear, i.e.:

$$\|p_i^U(\alpha u+\alpha' u')-(\alpha p_i^U u+\alpha' p_i^U u')\|_{U_i}\to 0 \quad (i\in\mathbb{N})\ \forall u,u'\in U,\ \alpha,\alpha'\in\mathbb{K}, \tag{2.1}$$

$$\|p_i^X(\alpha x_i+\alpha' x_i')-(\alpha p_i^X x_i+\alpha' p_i^X x_i')\|_{U_i}\to 0\ (i\in\mathbb{N})\ \forall x_i\to u, x_i'\to u'(i\in\mathbb{N}),\alpha,\alpha'\in\mathbb{K}. \tag{2.2}$$

In this article we say that we have a normed discrete approximation scheme $\mathcal{A}(U,U_i,X_i,P^U,P^X)$, shortly normed DAS $\mathcal{A}(U,U_i,X_i)$, if the following assumptions (N0), (N1), (N2) are fulfilled:

$$\text{(N0)}\quad \forall u\ \exists(x_i)_{i\in\mathbb{N}}: x_i\to u\ (i\in\mathbb{N}), \tag{2.3}$$

$$\text{(N1)}\quad \|p_i^U u\|_{U_i}\to 0\ (i\in N\subseteq\mathbb{N})\ \Rightarrow\ u=0, \tag{2.4}$$

$$\text{(N2)}\quad \|p_i^X x_i\|_{U_i}\to 0\ (i\in\mathbb{N})\ \Leftrightarrow\ \|x_i\|_{X_i}\to 0\ (i\in\mathbb{N}). \tag{2.5}$$

We notice that, by assumption (N0), every convergent sequence $x_i\to u$ $(i\in N\subseteq\mathbb{N})$ can be "filled up" to be convergent sequence $x_i\to u$ $(i\in\mathbb{N})$:

$$\exists(x_i')_{i\in\mathbb{N}}: x_i'\to u\ (i\in\mathbb{N}),\ x_i'=x_i\ (i=\mathbb{N}).$$

It is clear that every normed DAS is a metric discrete limit space (with $Ru=\{(x_i)_{i\in\mathbb{N}}\mid \|p_i^X x_i-p_i^U u\|_{U_i}\to 0\ (i\in\mathbb{N})\}$) and a metric discrete approximation by [2] and that every discrete approximation $\mathcal{A}(E,\prod E_i,R)$ of normed spaces in [1] can be considered as a normed DAS $\mathcal{A}(E,E_i,U_i)$ with $U_i=E_i$, $p_i^U=I$, $i\in\mathbb{N}$ (See e.g. [1]). If for a normed DAS $\mathcal{A}(U,U_i,X_i)$ the implication $x_i\to u\ \Rightarrow\ \|x_i\|_{X_i}\to\|u\|_U$ holds then $\mathcal{A}(U,U_i,X_i)$ can be considered as a discrete approximation $\mathcal{A}(U,\prod X_i,R)$ of normed spaces.

Further we shall use the same notation $\|\cdot\|$ for norms in all normed spaces.

From (2.1)-(2.5) it follows easily that :

1) $\|p_i^U 0_U\|\to 0\ (i\in\mathbb{N}),\ \|p_i^X 0_{X_i}\|_{U_i}\to 0\ (i\in\mathbb{N}),$ (2.6)

2) $0_{X_i}\to 0_U\ (i\in\mathbb{N}),$ (2.7)

3) $x_i\to 0_U\ (i\in N\subseteq\mathbb{N})\Leftrightarrow\|p_i^X x_i\|\to 0\ (i\in N\subseteq\mathbb{N})\Leftrightarrow\|x_i\|\to 0\ (i\in N\subseteq\mathbb{N})$ (2.8)

4) $x_i\to u,\ x_i\to u'\ \ (i\in N\subseteq\mathbb{N})\ \Rightarrow\ u'=u$ (2.9)
(every convergent sequence has exactly one limit element),

5) $x_i\to u,\ x_i'\to u'\ \ (i\in N\subseteq\mathbb{N})\ \Rightarrow$
$\Rightarrow\ (\alpha x_i+\alpha' x_i')\to(\alpha u+\alpha' u')\ \ (i\in\mathbb{N}),\ \forall\alpha,\alpha'\in\mathbb{K}$ (2.10)
(the limit process is linear),

6) $x_i \to u,\ x_i' \to u\ \ (i \in N) \iff x_i \to u,\ \|x_i - x_i'\| \to 0\ \ (i \in N) \iff$
$\iff x_i \to u,\ x_i - x_i' \to 0_U\ \ (i \in N),$ (2.11)

7) $(x_i)_{i \in N}, (x_i')_{i \in N}$ are compact $\Rightarrow (\alpha x_i + \alpha' x_i')_{i \in N}$ is compact $\forall \alpha, \alpha' \in K$.

We give here short proofs of assertions (2.6)-(2.11):

1) (2.6) follows from (2.1) and (2.5) (in (2.1) we take $\alpha = \alpha' = 0$),

2) (2.6) $\Rightarrow \|p_i^X 0_{X_i} - p_i^U 0_U\| \to 0 \iff 0_{X_i} \to 0_U$,

3) $x_i \to 0 \iff \|p_i^X x_i - p_i^U 0_U\| \to 0 \iff \|p_i^X x_i\| \to 0 \iff \|x_i\| \to 0$,

4) $x_i \to u,\ x_i \to u' \iff \|p_i^X x_i - p_i^U u\|,\ \|p_i^X x_i - p_i^U u'\| \to 0 \Rightarrow$

$\Rightarrow \|p_i^U u - p_i^U u'\| \to 0 \Rightarrow \|p_i^U (u - u')\| \to 0 \Rightarrow u - u' = 0$,

5) $\|p_i^X (\alpha x_i + \alpha' x_i') - p_i^U (\alpha u + \alpha' u')\| =$

$= \|\alpha p_i^X x_i + \alpha' p_i^X x_i' - \alpha p_i^U u - \alpha' p_i^U u'\| + o(1) \to 0$,

6) $x_i \to u,\ x_i' \to u \iff \|p_i^X x_i - p_i^U u\| \to 0,\ \|p_i^X x_i' - p_i^U u\| \to 0 \iff$

$\iff x_i \to u,\ \|p_i^X x_i - p_i^X x_i'\| \to 0 \iff x_i \to u,\ \|p_i^X (x_i - x_i')\| \to 0 \iff$

$\iff x_i \to u,\ \|x_i - x_i'\| \to 0 \iff x_i \to u,\ x_i - x_i' \to 0_U$.

For normed spaces E,F we denote by $\mathcal{L}(E,F)$ the normed space of bounded linear operators $A: E \to F$ with the norm $\|A\| = \sup\{\|Ae\| \mid e \in E,\ \|e\| = 1\}$.

The important special case, when (2.1) and (2.2) are fulfilled, is the case of linear operators p_i^U, p_i^X, $i \in N$. In this case the assumption (N2) is equivalent to the following assumption (2.12):

$$\exists i_0, c, c' > 0:\ p_i^X \in \mathcal{L}(X_i, U_i),\ \|p_i^X\| \le c,\ \|p_i^X x_i\| \ge c' \|x_i\|\ \ \forall i \ge i_0,\ x_i \in X_i. \quad (2.12)$$

Let us prove it.

a) It is clear that if (2.12) holds then (2.5) holds, too.

b) Let (N2) hold but for some subsequence N' of indices there exist elements x_i so that $\|x_i\| = 1$, $\|p_i^X x_i\| \to \infty$ $(i \in N')$. Then for $x_i' := x_i / \|p_i^X x_i\|$ we have $\|x_i'\| \to 0$, $\|p_i^X x_i'\| = 1$ $(i \in N')$, contrary to (N2).

c) Let (N2) hold but for some subsequence N' of indices there exist elements x_i so that $\|p_i^X x_i\| / \|x_i\| \to 0$ $(i \in N')$. Then for $x_i' := x_i / \|x_i\|$ we have $\|x_i'\| = 1$, $\|p_i^X x_i'\| \to 0$ $(i \in N')$, contrary to (N2).

Particularly, if connecting operators p_i^X, $i \in N$ are linear then from the asumption (N2) it follows that the normed spaces X_i are isomorphic with the subspaces $p_i^X X_i$ of U_i for $i \ge i_0$.

3. Consistency, closedness and $\mathcal{B}\mathcal{R}$-stability at $y \in V$ of sequences of operators. Let us consider the equation

$$Au = v^0, \quad A: \mathcal{D}(A) \subset U \to V \tag{3.1}$$

and the sequence of "approximate" equations

$$B_i x_i = y_i^0, \quad B_i: \mathcal{D}(B_i) \subset X_i \to Y_i, \quad i \in \mathbb{N}. \tag{3.2}$$

For numerical methods it is of interest to prove the convergence of the solutions and approximate solutions of (3.2) to the solution of (3.1). With this purpose assumptions suitable for applications and investigations are to be made.

We propose that U, V, X_i, Y_i $(i \in \mathbb{N})$ are normed spaces over the same field of constants $\mathbb{K} \in \{\mathbb{C}, \mathbb{R}\}$ and that some normed discrete approximation schemes $\mathcal{A}(U, U_1, X_1, P^U, P^{X_1})$, $\mathcal{A}(V, V_1, Y_1, P^V, P^{V_1})$ are fixed. By $A^{-1}v$ we denote the set $\{u \in U \mid Au = v\}$.

The sequence of operators $A, (B_i)_{i \in \mathbb{N}}$ is called:

1) consistent at $v \in V$ if

$$\forall u \in A^{-1}v \quad \exists (x_i^u)_{i \in \mathbb{N}}: \; x_i^u \in \mathcal{D}(B_i), \; x_i^u \to u, \; B_i x_i^u \to Av = v \quad (i \in \mathbb{N}), \tag{3.3}$$

2) closed at $v \in V$ if, for every $N \subset \mathbb{N}$,

$$x_i \in \mathcal{D}(B_i), \; x_i \to u, \; B_i x_i \to v \quad (i \in N) \;\Rightarrow\; u \in \mathcal{D}(A), \; Au = v. \tag{3.4}$$

Let further $\mathcal{B}$ be some fixed nonvoid set of sequences $(x_i)_{i \in N}$, $N \subset \mathbb{N}$ with $x_i \in X_i$ so that

$$(x_i)_{i \in N} \subset \mathcal{B} \;\Rightarrow\; (x_i)_{i \in N'} \subset \mathcal{B} \quad \forall N' \subset N. \tag{3.5}$$

Note that as a special case $\mathcal{B}$ may coincide with the set $\mathcal{A}$ of all sequences $(x_i)_{i \in N}$, $N \subset \mathbb{N}$. If $(x_i)_{i \in N} \in \mathcal{B}$, we say that $(x_i)_{i \in N}$ is a $\mathcal{B}$-sequence.

In this article we say that the sequence of operators $(B_i)_{i \in \mathbb{N}}$ is:

1) $\mathcal{R}$-stable at $v \in V$ if, for every $N \subset \mathbb{N}$,

$$x_i, x_i' \in \mathcal{D}(B_i), \; x_i \to u, \; B_i x_i \to v, \; B_i x_i' \to v \; (i \in N) \;\Rightarrow\; x_i' \to u \; (i \in N), \tag{3.6}$$

2) $\mathcal{B}\mathcal{R}$-stable at $v \in V$ if, for every $N \subset \mathbb{N}$,

$$(x_i') \in \mathcal{B}, \; x_i, x_i' \in \mathcal{D}(B_i), \; x_i \to u, \; B_i x_i \to v, \; B_i x_i' \to v \quad (i \in N) \;\Rightarrow\; x_i' \to u \quad (i \in N). \tag{3.7}$$

It is clear that if $(B_i)_{i \in \mathbb{N}}$ is $\mathcal{R}$-stable at v then $(B_i)_{i \in \mathbb{N}}$ is $\mathcal{B}\mathcal{R}$-stable at v for every $\mathcal{B} \subset \mathcal{A}$.

Let us explain these concepts in the terms of approximations of the solution of (3.1).

We call a sequence $(x_i)_{i\in N}$ with $x_i \in \mathcal{D}(B_i)$ a sequence of ε_i solutions for (3.2) if $\|B_i x_i - y_i^0\| \le \varepsilon_i$ $(i\in N)$. If $y_i^0 \to v^0$ $(i\in N)$ then $(x_i)_{i\in N}$ is a sequence of ε_i-solutions for (3.2) with $\varepsilon_i \to 0$ $(i\in N)$ if and only if $x_i \in \mathcal{D}(B_i)$, $B_i x_i \to v^0$ $(i\in N)$. (Indeed, by (2.11), $B_i x_i \to v^0$, $y_i^0 \to v^0 \Leftrightarrow y_i^0 \to v^0$, $\|B_i x_i - y_i^0\| \to 0$.) If $v^0 \in \mathcal{R}(A)$ and $A,(B_i)_{i\in N}$ is consistent at v^0 then there always exist sequences $(x_i)_{i\in N}$ so that $x_i \in \mathcal{D}(B_i)$, $B_i x_i \to v^0$ $(i\in N)$.

Let us assume that $y_i^0 \to v^0$ $(i\in N)$. Then :

1) $A,(B_i)_{i\in N}$ is consistent at $v^0 \Leftrightarrow$ for every solution u^0 of (3.1) there exists a converging to u^0 sequence $(x_i)_{i\in N}$ of ε_i-solutions for (3.2) with $\varepsilon_i \to 0$ $(i\in N)$,

2) $A,(B_i)_{i\in N}$ is closed at $v^0 \Leftrightarrow$ if there exists converging to u sequence $(x_i)_{i\in N}$, $N \subset N$ of ε_i-solutions for (3.2) with $\varepsilon_i \to 0$ then u is a solution of (3.1),

3) $(B_i)_{i\in N}$ is $\mathcal{BR}$-stable at $v^0 \Leftrightarrow$ if some sequence $(x_i)_{i\in N}$, $N \subset N$ of ε_i-solutions for (3.2) with $\varepsilon_i \to 0$ $(i\in N)$ is converging to u then every $\mathcal{B}$-sequence $(x_i')_{i\in N}$ of ε'_i-solutions for (3.2) with $\varepsilon_i' \to 0$ $(i\in N)$ is also converging to u.

These assertions follow immediately from definitions.

For numerical methods it is typical that the family (B_i) of approximating operators is constructed so that $A,(B_i)_{i\in N}$ is consistent and closed (at every $v\in V$). Then it is known that stability of (B_i) is important for converging of the approximate solutions of (3.2) to the solution of (3.1). We illustrate this by the following Theorem 3.1.

Theorem 3.1. Let $v^0 \in \mathcal{R}(A)$. Let us consider the following assertions (A) and (S) :

(A) equation (3.1) has unique solution u^0 and for any $\mathcal{B}$-sequence $(x_i')_{i\in N}$ from $x_i' \in \mathcal{D}(B_i)$, $B_i x_i' \to v^0$ $(i\in N)$ follows that $x_i' \to u^0$ $(i\in N)$,

(S) $(B_i)_{i\in N}$ is $\mathcal{BR}$-stable at v^0.

Then :

1) if $A,(B_i)_{i\in N}$ is closed at v^0 and (A) holds then (S) holds,

2) if there exists $\mathcal{B}$-sequence $(x_i^0)_{i\in N}$ so that $B_i x_i^0 \to v^0$ $(i\in N)$, if $A,(B_i)_{i\in N}$ is consistent at v^0 and if (S) holds then (A) holds .

Proof. 1) If for some $(x_i)_{i\in N}$, $N \subset N$ we have $x_i \to u$, $B_i x_i \to v^0$ then, by closedness condition and unique solvability of (3.1), $u = A^{-1}v^0 = u^0$. Therefore, by (A), every $\mathcal{B}$-sequence $(x_i')_{i\in N}$ such that $B_i x_i' \to v^0$ will converge to u, as it is required for $\mathcal{BR}$-stability at v^0.

2) For every $u \in A^{-1}v^0$ by consistency assumption there exists a consistency sequence $(x_i^u)_{i\in N}$ so that $x_i^u \to u$, $B_i x_i^u \to v^0$ $(i\in N)$. Thereby from $\mathcal{BR}$-stability follows that $x_i^0 \to u$ for every $u \in A^{-1}v^0$. Unique

solvability of (3.1) is now the consequence of the uniqueness of the limit element for $(x_i^0)_{i\in N}$.

Let further $u^0 = A^{-1}v^0$. Then $x_i^0 \to u^0$, $B_i x_i^0 \to v^0$ and thereby, by (S), every $\mathcal{B}$-sequence $(x_i')_{i\in N}$ such that $B_i x_i' \to v^0$ is also converging to u^0. ■

4. **$\mathcal{B}$-regularity at $v\in V$ of sequences of operators**. It appears that the main difficulty in applying Theorem 3.1 to concrete numerical methods is to prove $\mathcal{BR}$-stability of (B_i). Therefore various sufficient conditions for stability are used.

We say that the sequence $(B_i)_{i\in N}$ of operators $B_i: \mathcal{D}(B_i) \subset X_i \to Y_i$ is :

1) regular at $v^0 \in V$ if, for every $N \subset N$,

$$B_i x_i \to v^0 \ (i\in N) \ \Rightarrow \ (x_i)_{i\in N} \text{ is compact.} \qquad (4.1)$$

2) $\mathcal{B}$-regular at $v^0 \in V$ if, for every $N \subset N$,

$$(x_i')_{i\in N} \in \mathcal{B},\ B_i x_i' \to v^0 \ (i\in N) \ \Rightarrow \ (x_i')_{i\in N} \text{ is compact.} \qquad (4.2)$$

It is clear that if $(B_i)_{i\in N}$ is regular at v^0 then $(B_i)_{i\in N}$ is $\mathcal{B}$-regular at v^0 for every $\mathcal{B}$.

If $y_i^0 \to v^0$ in (3.1), (3.2) then $\mathcal{B}$-regularity of (B_i) at v^0 means that every $\mathcal{B}$-sequence $(x_i')_{i\in N}$ of ε_i'-solutions for (3.2) with $\varepsilon_i' \to 0$ $(i\in N)$ is compact.

It appears that $\mathcal{B}$-regularity and $\mathcal{BR}$-stability conditions are often consequences of each other in practical applications. At the same time $\mathcal{B}$-regularity is sometimes easier to verify than $\mathcal{BR}$-stability.

We say that A is injective to v^0 if the equation $Au = v^0$ has at most one solution, i.e.: $Au = v^0$, $Au' = v^0 \Rightarrow u = u'$.

Theorem 4.1. Let us consider the following assertions (S) and (R):

(S) $(B_i)_{i\in N}$ is $\mathcal{BR}$-stable at v^0,

(R) $(B_i)_{i\in N}$ is $\mathcal{B}$-regular at v^0.

Then :

1) if $v^0 \in \mathcal{R}(A)$, if $A,(B_i)_{i\in N}$ is consistent at v^0 and if (S) holds then (R) holds,

2) if A is injective to v^0, if $A,(B_i)_{i\in N}$ is closed at v^0 and if (R) holds then (S) holds.

If there exists a $\mathcal{B}$-sequence $(x_i^0)_{i\in N}$ so that $B_i x_i^0 \to v^0 (i\in N)$ then :

3) if $A,(B_i)_{i\in N}$ is consistent at v^0 and (S) holds then A is injective to v^0,

4) if $A,(B_i)_{i\in N}$ is closed at v^0 and (R) holds then $v^0 \in \mathcal{R}(A)$,

5) if $A,(B_i)_{i\in N}$ is consistent and closed at v^0 then :
$v^0 \in \mathcal{R}(A)$, (S) $\Leftrightarrow$ A is injective to v^0, (R) $\Leftrightarrow$ (S),(R).

Proof. 1) Let $(x_i') \in \mathcal{B}$, $x_i' \in \mathcal{D}(B_i)$, $B_i x_i' \to v^0$ and $v^0 \in \mathcal{R}(A)$. Then by Theorem 3.1 it follows from (S) that (x_i') is a converging sequence and therefore (x_i') is compact as well.

2) Let $(x_i') \in \mathcal{B}$, $x_i, x_i' \in \mathcal{D}(B_i)$, $x_i \to u$, $B_i x_i \to v^0$, $B_i x_i' \to v^0$ $(i \in \mathbb{N})$. By $\mathcal{B}$-regularity of (B_i) at v^0, $(x_i')_{i \in \mathbb{N}}$ is compact. Let $\mathbb{N}'$ be an arbitrary subsequence of $\mathbb{N}$. Then there exist $\mathbb{N}'' \subseteq \mathbb{N}'$ and $u' \in U$ so that $x_i' \to u'$ $(i \in \mathbb{N}'')$. At that, by closedness of $A,(B_i)$ at v^0, $Au' = v^0$. Whereas A is injective to v^0, we conclude that every subsequence $(x_i')_{i \in \mathbb{N}'}$ of the sequence $(x_i')_{i \in \mathbb{N}}$ has subsequence, converging to the same limit element u'. Thus the sequence $(x_i')_{i \in \mathbb{N}}$ itself must also converge to u', as it is required for $\mathcal{BR}$-stability of (B_i) at v^0.

3) Let (B_i) be $\mathcal{BR}$-stable at v^0. If $Au = v^0$ then $v^0 \in \mathcal{R}(A)$ and equation (3.1) is uniquely solvable by Theorem 3.1. Thus $Au = v^0$, $Au' = v^0 \Rightarrow u = u'$. Therefore from $\mathcal{BR}$-stability of (B_i) at v^0 follows injectivity of A to v^0.

4) Let (B_i) be $\mathcal{B}$-regular at v^0. Then the $\mathcal{B}$-sequence $(x_i^0)_{i \in \mathbb{N}}$ is compact. Let $x_i^0 \to u$ $(i \in \mathbb{N} \subseteq \mathbb{N})$. Then, by closedness of $A,(B_i)$ at v^0, $u \in \mathcal{D}(A)$ and $Au = v^0$. Therefore from $\mathcal{B}$- regularity of (B_i) at v^0 follows that $v^0 \in \mathcal{R}(A)$.

5) Assertion 5) follows from assertions 1)-4) of this Theorem. ■

We note that, by assumption (N0) for $\mathcal{A}(V, V_i, Y_i)$, at least in the case when $\mathcal{B} = \mathcal{A}$ and $\overline{\mathcal{R}(B_i)} = Y_i$, $i \in \mathbb{N}$ there always exist sequences $(x_i)_{i \in \mathbb{N}} \in \mathcal{B}$ so that $B_i x_i \to v^0$. As a consequence of Theorems 3.1 and 4.1 we get the following Theorem 4.2.

Theorem 4.2. Consider the following equations (4.3) and (4.4):

$$Au = v^0, \quad A: \mathcal{D}(A) \in U \to V \quad , \tag{4.3}$$

$$B_i x_i = y_i^0 , \quad B_i: \mathcal{D}(B_i) \in X_i \to Y_i, \quad i \in \mathbb{N}. \tag{4.4}$$

Let $y_i^0 \to v^0$ $(i \in \mathbb{N})$ and let $A,(B_i)_{i \in \mathbb{N}}$ be consistent and closed at v^0. Let there exist $\mathcal{B}$-sequence $(x_i')_{i \in \mathbb{N}}$ of ε_i-solutions for (4.4) with $\varepsilon_i \to 0$ $(n \in \mathbb{N})$. Then the following assertions (I)-(IV) are equivalent:

(I) equation (4.3) has unique solution u^0 and every $\mathcal{B}$- sequence $(x_i)_{i \in \mathbb{N}}$ of ε_i-solutions for (4.4) with $\varepsilon_i \to 0$ $(n \in \mathbb{N})$ is converging to u^0,

(II) equation (4.3) has at least one solution (i.e. $v^0 \in \mathcal{R}(A)$) and $(B_i)_{i \in \mathbb{N}}$ is $\mathcal{BR}$-stable at v^0,

(III) equation (4.3) has at most one solution (i.e. A is injective to v^0) and $(B_i)_{i \in \mathbb{N}}$ is $\mathcal{B}$-regular at v^0,

(IV) $(B_i)_{i \in \mathbb{N}}$ is $\mathcal{BR}$-stable and $\mathcal{B}$-regular at v^0. ■

To examine the convergence of (approximate) solutions of equations (4.4) to the solution of equation (4.3), we must have $v^0 \in \mathcal{R}(A)$ but we need not have A injective to v^0. Thus in this case $\mathcal{B}$-regularity at v^0 is a weaker and more natural assumption than $\mathcal{BR}$-stability at v^0.

From definitions of $\mathcal{B}$-regularity and closedness at v^0 we get the following Theorem 4.3.

Theorem 4.3. Let $A,(B_i)_{i\in\mathbb{N}}$ be closed at v^0 and $(B_i)_{i\in\mathbb{N}}$ be $\mathcal{B}$-regular at v^0. If $B_i x_i' \to v^0 (i\in\mathbb{N})$ for some $\mathcal{B}$-sequence $(x_i')_{i\in\mathbb{N}}$ then $(x_i')_{i\in\mathbb{N}}$ is compact and all its cluster points are solutions of the equation $Au = v^0$. ■

In other words, under the assumptions of Theorem 4.3, if $y_i^0 \to v^0$ and if $(x_i')_{i\in\mathbb{N}}$ is $\mathcal{B}$-sequence of ε_i-solutions for (4.4) with $\varepsilon_i \to 0$ then (x_i') is compact and from $x_i' \to u$ $(i\in\mathbb{N}' \subset \mathbb{N})$ follows that u is the solution of (4.3).

The disadvantage of Theorems 3.1, 4.2 and 4.3 is that in these theorems there is no assertion about the existence of the solutions of approximate equations (4.4) and no error estimations. To get them, we need some complementary assumptions. We shall consider only the case of linear operators here.

5. Sequence of linear operators $B_i : \mathcal{D}(B_i) \subset X_i \to Y_i$. For linear operators it appears that if a sequence of operators is consistent at v^0 then it is closed ($\mathcal{R}$-stable, regular) at v^0 and at 0_V at the same time.

Proposition 5.1. Let $A : \mathcal{D}(A) \subset U \to V$, $B_i : \mathcal{D}(B_i) \subset X_i \to Y_i$ $(i\in\mathbb{N})$ be linear operators and let $v^0 \in \mathcal{R}(A)$. Then the following two assertions 1) and 2) are equivalent:

1) $A,(B_i)_{i\in\mathbb{N}}$ is consistent at v^0,

2) $A,(B_i)_{i\in\mathbb{N}}$ is consistent at 0_V and for v^0 there exist an element $u^0 \in A^{-1}v^0$ and a sequence $(x_i^0)_{i\in\mathbb{N}}$ so that $x_i^0 \in \mathcal{D}(B_i)$, $x_i^0 \to u^0$, $B_i x_i^0 \to Au^0 = v^0$ $(i\in\mathbb{N})$.

Proof. 1)→2). Let $\bar{u}$ be an arbitrary element of $A^{-1}0_V$ and let $(x_i^u)_{i\in\mathbb{N}}$ be consistency sequence at v^0 so that $x_i^u \to u$, $B_i x_i^u \to Au = v^0$. Then $u^+ := u + \bar{u} \in \mathcal{D}(A)$, $Au^+ = v^0$. Let, further, $(x_i^+)_{i\in\mathbb{N}}$ be a consistency sequence at v^0 so that $x_i^+ \to u^+$, $B_i x_i^+ \to Au^+ = v^0$. Then for $\bar{x}_i = x_i^+ - x_i^u$ we have $\bar{x}_i \to u^+ - u = \bar{u}$, $B_i\bar{x}_i = B_i x_i^+ - B_i x_i^u \to v^0 - v^0 = 0$, i.e. $(\bar{x}_i)_{i\in\mathbb{N}}$ is a consistency sequence at 0_V so that $\bar{x}_i \to \bar{u}$, $B_i\bar{x}_i \to 0_V$.

2)→1). Let u be an arbitrary element of $A^{-1}v^0$ and let $(\bar{x}_i)_{i\in\mathbb{N}}$ be a consistency sequence at 0_V so that $\bar{x}_i \to u - u^0 \in A^{-1}0_V$, $B_i\bar{x}_i \to 0_V$. Then $x_i^+ = \bar{x}_i + x_i^0 \to (u - u^0) + u^0 = u$, $B_i x_i^+ = B_i\bar{x}_i + B_i x_i^0 \to 0 + v^0 = v^0$, i.e. $(x_i^+)_{i\in\mathbb{N}}$ is consistency sequence at v^0 so that $x_i^+ \to u$, $B_i x_i^+ \to v^0$. ■

Proposition 5.2. Let $A: \mathcal{D}(A) \subset U \to V$, $B_i: \mathcal{D}(B_i) \subset X_i \to Y_i$ $(i \in \mathbb{N})$ be linear operators. Let $v^0 \in \mathcal{R}(A)$ and let there exist an element $u^0 \in A^{-1}v^0$ and a sequence $(x_i^0)_{i\in\mathbb{N}}$ so that $x_i^0 \in \mathcal{D}(B_i)$, $x_i^0 \to u^0$, $B_i x_i^0 \to v^0 = Au^0$ $(i \in \mathbb{N})$. Then $A,(B_i)_{i\in\mathbb{N}}$ is closed at v^0 if and only if $A,(B_i)_{i\in\mathbb{N}}$ is closed at 0_V.

Proof. i) Let $A,(B_i)$ be closed at v^0 and let $\bar{x}_i \to \bar{u}$, $B_i\bar{x}_i \to 0_V$. Then for $x_i^+ := x_i^0 + \bar{x}_i$ we have $x_i^+ \to u^0 + \bar{u}$, $B_i x_i^+ = B_i x_i^0 + B_i\bar{x}_i \to v^0 + 0_V = v^0$. Therefore $(u^0 + \bar{u}) \in \mathcal{D}(A)$, $A(u^0 + \bar{u}) = v^0$ by closedness of $A,(B_i)$ at v^0. Consequently $\bar{u} \in \mathcal{D}(A)$ and $A\bar{u} = 0_V$ as it is required for closedness of $A,(B_i)$ at 0_V.

ii) Let $A,(B_i)$ be closed at 0_V and let $x_i \to u$, $B_i x_i \to v^0$. Then for $\bar{x}_i := x_i^0 - x_i$ we have $\bar{x}_i \to u^0 - u$, $B_i\bar{x}_i = B_i x_i^0 - B_i x_i \to v^0 - v^0 = 0_V$. Therefore $(u^0 - u) \in \mathcal{D}(A)$, $A(u^0 - u) = 0$ by closedness of $A,(B_i)$ at 0_V. Consequently $u \in \mathcal{D}(A)$, $Au = v^0$ as it is required for closedness of $A,(B_i)$ at v^0. ■

Proposition 5.3. Let $A: \mathcal{D}(A) \subset U \to V$, $B_i: \mathcal{D}(B_i) \subset X_i \to Y_i$ $(i \in \mathbb{N})$ be linear operators. Let there exist a sequence $(x_i^0)_{i\in\mathbb{N}}$ so that $x_i^0 \in \mathcal{D}(B_i)$, $x_i^0 \to u^0$, $B_i x_i^0 \to v^0$ $(i \in \mathbb{N})$. Then:

1) $(B_i)_{i\in\mathbb{N}}$ is $\mathcal{R}$-stable at $v^0 \Leftrightarrow$ $(B_i)_{i\in\mathbb{N}}$ is $\mathcal{R}$-stable at 0_V,

2) $(B_i)_{i\in\mathbb{N}}$ is regular at $v^0 \Leftrightarrow$ $(B_i)_{i\in\mathbb{N}}$ is regular at 0_V.

Proof. i) (Assertion 1, part $\Rightarrow$) For the sequence $(0_{X_i})_{i\in\mathbb{N}}$ of the elements $0_{X_i} \in X_i$ we have $0_{X_i} \to 0_U$, $B_i 0_{X_i} = 0_{Y_i} \to 0_V$. Therefore the following necessary and sufficient condition for $\mathcal{R}$-stability of $(B_i)_{i\in\mathbb{N}}$ at $0 \in V$ holds:

$$(B_i)_{i\in\mathbb{N}} \text{ is } \mathcal{R}\text{-stable at } 0_V \Leftrightarrow$$
$$\Leftrightarrow \{\forall N \subset \mathbb{N}: x_i \in \mathcal{D}(B_i),\ B_i x_i \to 0_V\ (i \in N) \Rightarrow x_i \to 0_U\ (i \in N)\}. \tag{5.1}$$

Let $x_i \in \mathcal{D}(B_i)$, $B_i x_i \to 0_V$. Then $B_i(x_i^0 + x_i) = B_i x_i^0 + B_i x_i \to v^0 + 0 = v^0$ and $x_i^0 + x_i \to u^0$ by $\mathcal{R}$-stability of (B_i) at v^0. Therefore $x_i = (x_i^0 + x_i) - x_i^0 \to u^0 - u^0 = 0$ as it is required for $\mathcal{R}$-stability of (B_i) at 0_V by (5.1).

ii) (Assertion 1, part $\Leftarrow$) Let $x, x_i' \in \mathcal{D}(B_i)$, $x_i \to u$, $B_i x_i \to v^0$, $B_i x_i' \to v^0$ $(i \in N)$. Then $B_i(x_i' - x_i) = B_i x_i' - B_i x_i \to v^0 - v^0 = 0_V$ and, by $\mathcal{R}$-stability of (B_i) at 0_V and (5.1), therefore, $x_i' - x_i \to 0_U$. Thus, by (2.11), $x_i' \to u$ as it is required for $\mathcal{R}$-stability of (B_i) at v^0.

iii) (Assertion 2, part $\Rightarrow$) Let $B_i\bar{x}_i \to 0_V$ $(i \in N)$. Then $B_i(x_i^0 + \bar{x}_i) \to v^0$ and $(x_i^0 + \bar{x}_i)$ is compact by regularity of (B_i) at v^0. Therefore $(\bar{x}_i) = ((x_i^0 + \bar{x}_i) - x_i^0)$ is also compact, as it is required for regularity of (B_i) at 0.

iv) (Assertion 2, part $\Leftarrow$) Let $B_i x_i \to v^0$. Then $B_i(x_i - x_i^0) = B_i x_i - B_i x_i^0 \to$ $\to 0_V$ and $(x_i - x_i^0)$ is compact by regularity of (B_i) at 0_V. Therefore $(x_i) = ((x_i - x_i^0) + x_i^0)$ is also compact as it is required for regularity of (B_i) at v^0. ■

Note that for the sequence $(B_i)_{i\in N}$ of linear operators B_i:

1) if $(B_i)_{i\in N}$ is $\mathcal{BR}$-stable at 0_V then, for every $N\subset\mathbb{N}$,

$$(x_i')_{i\in N}\in\mathcal{B},\ x_i'\in\mathcal{D}(B_i),\ B_ix_i'\to 0_V\ (i\in N)\ \Rightarrow\ x_i'\to 0\ (i\in N). \qquad (5.2)$$

2) if $(0_{X_i})_{i\in N}\in\mathcal{B}$ and $(B_i)_{i\in N}$ is $\mathcal{BR}$-stable at 0_V then, for every $N\subset\mathbb{N}$,

$$x_i\in\mathcal{D}(B_i),\ x_i\to u,\ B_ix_i\to 0_V\ (i\in N)\ \Rightarrow\ u=0. \qquad (5.3)$$

3) if $(0_{X_i})_{i\in N}\in\mathcal{B}$ and if (5.2), (5.3) hold then $(B_i)_{i\in N}$ is $\mathcal{BR}$-stable at 0_V.

Proposition 5.4. Let $A:\mathcal{D}(A)\subset U\to V$, $B_i:\mathcal{D}(B_i)\subset X_i\to Y_i$ $(i\in\mathbb{N})$ be linear operators. Let the set $\mathcal{B}$ be such that

$$(x_i)_{i\in N},\ (x_i')_{i\in N}\in\mathcal{B}\ \Rightarrow\ (x_i+x_i')_{i\in N},\ (x_i-x_i')_{i\in N}\in\mathcal{B} \qquad (5.4)$$

and let there exist a sequence $(x_i^0)_{i\in N}\in\mathcal{B}$ so that $x_i^0\in\mathcal{D}(B_i)$, $x_i^0\to u^0$, $B_ix_i^0\to v^0$ $(i\in\mathbb{N})$. Then:

1) $(B_i)_{i\in N}$ is $\mathcal{BR}$-stable at $v^0\ \Leftrightarrow$ $(B_i)_{i\in N}$ is $\mathcal{BR}$-stable at 0_V.

2) $(B_i)_{i\in N}$ is $\mathcal{B}$-regular at $v^0\ \Leftrightarrow$ $(B_i)_{i\in N}$ is $\mathcal{B}$-regular at 0_V.

Proof of Proposition 5.4 almost fully repeats the proof of Proposition 5.3.

i) (Assertion 1), part $\Rightarrow$) We note at first that under our assumptions $(0_{X_i})_{i\in N}\in\mathcal{B}$ and if $(B_i)_{i\in N}$ is $\mathcal{BR}$-stable at v^0 then for every $N\in\mathbb{N}$:

$$x_i\in\mathcal{D}(B_i),\ x_i\to u^+,\ B_ix_i\to v^0\ (i\in N)\ \Rightarrow\ u^+=u^0. \qquad (5.5)$$

Let $(x_i')\in\mathcal{B}$, x_i', $x_i\in\mathcal{D}(B_i)$, $x_i\to u'$, $B_ix_i'\to 0_V$, $B_ix_i\to 0_V$. Then $(x_i^0+x_i)\in\mathcal{D}(B_i)$, $(x_i^0+x_i)\to u^0+u'$, $B_i(x_i^0+x_i)\to v^0$. By $\mathcal{BR}$-stability of (B_i) at v^0 and (5.5) then $u^0+u'=u^0$ and $u'=0$. Further, $(x_i^0+x_i')\in\mathcal{B}$, $(x_i^0+x_i')\in\mathcal{D}(B_i)$, $B_i(x_i^0+x_i')\to v^0$ and thus, by $\mathcal{BR}$-stability of (B_i) at v^0, $x_i^0+x_i'\to u^0$. By (2.11) then $x_i'\to 0_U$. Therefore $x_i'\to u'\ (=0_U)$ as it is required for $\mathcal{BR}$-stability of (B_i) at 0_V.

ii) (Assertion 1), part $\Leftarrow$) Let $(x_i')\in\mathcal{B}$, x_i', $x_i\in\mathcal{D}(B_i)$, $x_i\to u'$, $B_ix_i'\to v^0$, $B_ix_i\to v^0$ $(i\in\mathbb{N})$. Then $(x_i'-x_i^0)\in\mathcal{B}$, $(x_i-x_i^0)$, $(x_i'-x_i^0)\in\mathcal{D}(B_i)$, $(x_i-x_i^0)\to u'-u^0$, $B_i(x_i-x_i^0)\to 0_V$, $B_i(x_i'-x_i^0)\to 0_V$. By $\mathcal{BR}$-stability of (B_i) at 0_V and (5.3), (5.2) therefore $u'-u^0=0_U$, $x_i'-x_i^0\to 0_U$. By (2.11) thus $x_i'\to u^0$. Therefore $x'\to u'\ (=u^0)$ as it is required for $\mathcal{BR}$-stability of (B_i) at v^0.

iii) (Assertion 2), part $\Rightarrow$) Let $\bar{x}_i\in\mathcal{B}$, $B_i\bar{x}_i\to 0_V$ $(i\in\mathbb{N})$. Then $(x_i^0+\bar{x}_i)\in\mathcal{B}$, $B_i(x_i^0+\bar{x}_i)\to v^0$ and $(x_i^0+\bar{x}_i)$ is compact by $\mathcal{BR}$-regularity of (B_i) at v^0. Therefore $(\bar{x}_i)$ is compact as it is required for $\mathcal{B}$-regularity of (B_i) at 0.

iv) (Assertion 2), part $\Leftarrow$) Let $(x_i)\in\mathcal{B}$, $B_ix_i\to v^0$. Then $(x_i-x_i^0)\in\mathcal{B}$, $B_i(x_i-x_i^0)\to 0_V$ and $(x_i-x_i^0)$ is compact by $\mathcal{B}$-regularity of (B_i) at 0_V. Therefore (x_i) is compact as it is required for $\mathcal{B}$-regularity of (B_i) at v^0. ■

We note here that for linear operators from $\mathcal{BR}$-stability at 0_V follows consistency, closedness and regularity at 0_V, as it is stated in the following Proposition 5.5. We recall that for linear operator A

the following holds:

$$A \text{ is injective} \leftrightarrow A \text{ is injective to some } v \in \mathcal{R}(A) \leftrightarrow N(A)=\{0_U\}$$
$$\leftrightarrow N(A)=\{u\in U \mid Au=0\}=\{0_U\}. \tag{5.6}$$

Proposition 5.5 . Let $A: \mathcal{D}(A)\subset U\to V$, $B_i:\mathcal{D}(B_i)\subset X_i\to Y_i$ $(i\in\mathbb{N})$ be linear operators. Let $(0_{X_i})_{i\in\mathbb{N}}\in\mathcal{B}$ and let $A,(B_i)_{i\in\mathbb{N}}$ be $\mathcal{BR}$-stable at 0_V. Then:

1) A is injective $\leftrightarrow$ $A,(B_i)_{i\in\mathbb{N}}$ is consistent at 0_V ,

2) $A,(B_i)_{i\in\mathbb{N}}$ is closed at 0_V ,

3) $(B_i)_{i\in\mathbb{N}}$ is $\mathcal{B}$-regular at 0_V.

Proof. 1) If A is injective then $A^{-1}0_V=\{0_U\}$ and $(0_{X_i})_{i\in\mathbb{N}}$ is a consistency sequence for every $u\in A^{-1}0_V$. If $A,(B_i)_{i\in\mathbb{N}}$ is consistent at 0_V then for every $u\in A^{-1}0_V$ there exists a consistency sequence $x_i^u\to u$, $B_ix_i^u\to 0$ $(i\in\mathbb{N})$. By (5.3) then $u=0$, i.e. $N(A)=A^{-1}0_V=\{0_U\}$.

2) If $x_i\in\mathcal{D}(B_i)$, $x_i\to u$, $B_ix_i\to 0_V$ $(i\in N)$ then, by $\mathcal{BR}$ -stability of (B_i) at 0_V and (5.3), $u=0$. Therefore $u\in\mathcal{D}(A)$, $Au=0$.

3) If $(x_i)\in\mathcal{B}$, $B_ix_i\to 0_V$ $(i\in N)$ then $x_i\to 0_U$ $(i\in N)$ and thus $(x_i)_{i\in N}$ is compact. ■

6. Stability of sequences of linear operators. We say that the sequence $(B_i)_{i\in\mathbb{N}}$ of linear operators $B_i:\mathcal{D}(B_i)\subset X_i\to Y_i$ is:

1) stable if the following condition (6.1) holds :

$$\exists i_0\in\mathbb{N},\ c_0>0:\ i\ge i_0\Rightarrow\exists B_i^{-1}\in\mathcal{L}(Y_i,X_i),\ \|B_i^{-1}\|\le c_0\ , \tag{6.1}$$

2) stable on range if the following condition (6.2) holds :

$$\exists i_1\in\mathbb{N},\ c_1>0:\ i\ge i_1\Rightarrow\|B_ix_i\|\ge c_1\|x_i\|\quad\forall x_i\in\mathcal{D}(B_i). \tag{6.2}$$

It is clear that from (6.2) it follows that operators B_i, $i\ge i_1$ are injective.

If $(B_i)_{i\in\mathbb{N}}$ is stable on range then, for $i\ge i_1$, the equations $B_ix_i=y_i$ are uniquely solvable for every $y_i\in\mathcal{R}(B_i)$ and their solutions x_i^0 are uniformly stable with respect to such pertuberations δ_i of y_i that $y_i+\delta_i$ remains in $\mathcal{R}(B_i)$:

$$B_ix_i^0=y_i,\ \delta_i\in\mathcal{R}(B_i),\ B_i\tilde{x}_i=y_i+\delta_i\ \Rightarrow\ \|\tilde{x}_i-x_i^0\|\le\frac{1}{c_i}\|\delta_i\|\quad(i\ge i_1).$$

It appears that for the sequences of linear operators the concepts of stability, stability on range, $\mathcal{R}$-stability at v and $\mathcal{BR}$-stability at 0_V coincide for many cases, interesting for applications.

Proposition 6.1. $(B_i)_{i\in\mathbb{N}}$ is stable if and only if $(B_i)_{i\in\mathbb{N}}$ is stable on range and for some i_2 all B_i, $i\ge i_2$ are surjective.

Proof. 1) (6.1)$\Rightarrow$(6.2), whereas $\|x_i\| = \|B_i^{-1}B_ix_i\| \le \|B_i^{-1}\|\|B_ix_i\| \le$ $\le c_0\|B_ix_i\|$ $(i \ge i_0,\ c_0 > 0)$.

2) For $i \ge i_0 = \max\{i_1, i_2\}$ there exist operators $B_i^{-1}: Y_i \to X_i$, whereas $\mathcal{R}(B_i) = Y_i$ and B_i are injective. These operators are linear (as inverses of linear operators) and for them

$$\|y_i\| = \|B_iB_i^{-1}y_i\| \ge c_1\|B_i^{-1}y_i\|, \quad \forall y_i \in Y_i,$$

i.e. $\|B_i^{-1}y_i\| \le \frac{1}{c_1}\|y_i\|$, $\forall y_i \in Y_i$. Thus (6.1) holds with $c_0 = \frac{1}{c_1}$. ■

Let us denote by $\mathcal{F}_+$ a class of linear operators B_i such that

$$\mathcal{N}(B_i) = \{0\} \ (B_i \text{ is injective}) \Rightarrow B_i \text{ is surjective}. \qquad (6.3)$$

(For (6.3) it is sufficient that $\dim X_i = \dim Y_i < \infty$, which is usual for numerical methods.) It is clear by Proposition 6.1 that if $B_i \in \mathcal{F}_+$, $i \in \mathbb{N}$ then for stability of $(B_i)_{i\in\mathbb{N}}$ and stability on range of $(B_i)_{i\in\mathbb{N}}$ are equivalent conditions.

Proposition 6.2. If $(B_i)_{i\in\mathbb{N}}$ is stable on range then $(B_i)_{i\in\mathbb{N}}$ is $\mathcal{R}$-stable at every $v \in V$ ($\mathcal{BR}$-stable at every $v \in V$ for every $\mathcal{B} \subset \mathcal{A}$).

Proof. Let $x_i, x_i' \in \mathcal{D}(B_i)$, $B_ix_i \to v$, $B_ix_i' \to v$, $x_i \to u$ $(i \in N)$. Then $\bar{x}_i := (x_i' - x_i) \in \mathcal{D}(B_i)$ and $B_i\bar{x}_i = B_ix_i' - B_ix_i \to 0$. Therefore, by (6.2), $\bar{x}_i = x_i' - x_i \to 0$ and thus $x_i' = x_i + \bar{x}_i \to u$, as it it required for $\mathcal{R}$-stability at v. ■

Proposition 6.3. Let the set $\mathcal{B}$ be such that for some $\alpha > 0$

$$\mathcal{B} \supseteq \{(x_i)_{i\in\mathbb{N}} \mid x_i \in \mathcal{D}(B_i),\ \|x_i\| = \alpha\}. \qquad (6.4)$$

Then $(B_i)_{i\in\mathbb{N}}$ is stable on range if and only if $(B_i)_{i\in\mathbb{N}}$ is $\mathcal{BR}$-stable at 0_V.

Proof. Taking into account Proposition 6.2 and implication (5.2), it is sufficient here to prove that if

$$(x_i) \in \mathcal{B},\ x_i \in \mathcal{D}(B_i),\ B_ix_i \to 0_V \ (i \in N) \Rightarrow x_i \to 0 \ (i \in N) \qquad (6.5)$$

then (6.2) must hold.

If (6.2) does not hold then for every $n \in N$ there must exist an index $i(n) \in N$ and an element $x_{i(n)} \in \mathcal{D}(B_{i(n)})$ so that $i(n+1) > i(n)$ and $\|B_{i(n)}x_{i(n)}\| \le \|x_{i(n)}\|/n$. Then for subsequence $N' = \{i(1), i(2), \ldots\} \subseteq N$ and elements $x_i' := \alpha x_i/\|x_i\|$, $i \in N'$ we have $(x_i')_{i\in N'} \in \mathcal{B}$, $x_i' \in \mathcal{D}(B_i)$, $\|B_ix_i'\| \to 0$ $(i \in N')$. Therefore, by (2.8) for $\mathcal{A}(V, V_i, Y_i)$, $B_ix_i' \to 0$ $(i \in N')$ and if (6.5) holds then $x_i' \to 0$ $(i \in N')$. At the same time $x_i' \not\to 0$ $(i \in N')$ by (2.8), whereas $\|x_i'\| = \alpha \not\to 0$ $(i \in N')$. Thus, if (6.2) does not hold then (6.5) does not hold, too. Therefore (6.5)$\Rightarrow$(6.2). ■

As a consequence of Propositions 6.1, 6.2 and 6.3 we have the following Theorem 6.1.

Theorem 6.1. Let $A:\mathcal{D}(A)\subset U\to V$, $B_i:\mathcal{D}(B_i)\subset X_i\to Y_i$ $(i\in\mathbb{N})$ be linear operators. Let $B_i\in\mathcal{F}_+$ and the set $\mathcal{B}$ be such that (6.4) holds for some $\alpha>0$. Then the following assertions 1),2),3) and 4) are equivalent:

1) $(B_i)_{i\in\mathbb{N}}$ is stable,

2) $(B_i)_{i\in\mathbb{N}}$ is stable on range,

3) $(B_i)_{i\in\mathbb{N}}$ is $\mathcal{R}$-stable at every $v\in V$,

4) $(B_i)_{i\in\mathbb{N}}$ is $\mathcal{BR}$-stable at 0_V. ■

7. **Some results for linear operators**. Consider the following equations

$$Au=v^0, \tag{7.1}$$

$$B_ix_i=y_i^0\ ,\quad i\in\mathbb{N} \tag{7.2}$$

with linear operators $A:\mathcal{D}(A)\subset U\to V$, $B_i:\mathcal{D}(B_i)\subset X_i\to Y_i$, $i\in\mathbb{N}$.

Theorem 7.1. Let $v^0\in\mathcal{R}(A)$ and $A,(B_i)_{i\in\mathbb{N}}$ be consistent and closed at v^0. Let $B_i\in\mathcal{F}_+$, $i\in\mathbb{N}$ and let the set $\mathcal{B}$ be such that, for some $\alpha>0$,

$$\mathcal{B}\supset\{(x_i)_{i\in\mathbb{N}}\mid x_i\in\mathcal{D}(B_i),\ \|x_i\|\le\alpha,\ i\in\mathbb{N}\}. \tag{7.3}$$

Then the following assertions (I),(II) and (III) are equivalent:

(I) equation (7.1) has unique solution u^0 and the sequence $(B_i)_{i\in\mathbb{N}}$ is stable, i.e.:

there exist an index i_0 and a constant $c_0>0$, so that for every $i\ge i_0$ equation (7.2) is uniquely solvable for every $y_i^0\in Y_i$ and $\|x_i^0\|\le c_0\|y_i^0\|$ for the solutions x_i^0 of (7.2).

(II) the sequence $(B_i)_{i\in\mathbb{N}}$ is $\mathcal{BR}$-stable at 0_V.

(III) $\mathcal{N}(A)=\{0\}$ and the sequence $(B_i)_{i\in\mathbb{N}}$ is $\mathcal{B}$-regular at 0_V.

If any of the assertions (I), (II) or (III) holds and $y_i^0\to v^0$ $(i\in\mathbb{N})$ then:

1) the sequence $(x_i^0)_{i\in\mathbb{N}}$ of solutions x_i^0 of (7.2) for $i\ge i_0$ is converging to the solution u^0 of (7.1) and the following estimation (7.4) holds:

$$\|x_i^0-x_i\|\le c\|y_i^0-B_ix_i\|,\quad\forall x_i\in X_i\quad(i\ge i_0) \tag{7.4}$$

(the distance from the solution x_i^0 of (7.2) to any representative x_i of the solution u^0 of (7.1) is estimated by the defect of (7.2) on x_i),

2) every sequence $(x_i^\varepsilon)_{i\in\mathbb{N}}$ of ε_i-solutions for (7.2) with $\varepsilon_i\to0$ $(i\in\mathbb{N})$ is converging to the solution u^0 of (7.1) and the following estimation (7.5) holds:

$$\|x_i^\varepsilon-x_i\|\le c(\|y_i^0-B_ix_i\|+\varepsilon_i)\quad\forall x_i\in X_i\quad(i\ge i_0). \tag{7.5}$$

Proof. i) (I) ⇒ (II) by Theorem 6.1.

ii) (II) ⇔ (III) by Theorem 4.1, whereas $(0_{X_i})_{i\in\mathbb{N}} \in \mathcal{B}$ by (7.3) and the sequence $A,(B_i)_{i\in\mathbb{N}}$ is consistent and closed at 0_Y by Propositions 5.1 and 5.2.

iii) (II) ⇒ (I): stability of (B_i) is the consequence from Theorem 6.1, and the unique solvability of (7.1) follows from Theorem 3.1 with $\mathcal{B}=\mathcal{A}$ (by Theorem 6.1, (B_i) is $\mathcal{R}$-stable at v^0).

iv) By Theorem 3.1 with $\mathcal{B}=\mathcal{A}$ every sequence of ε_i-solutions for (7.2) with $\varepsilon_i \to 0$ is converging to $u^0 = A^{-1}v^0$, particularly $x_i^0 \to u^0$ as sequence of ε_i-solutions for (7.2) with $\varepsilon_i = 0$. Estimations (7.5) and (7.4) follow from the definition of the stability of (B_i) in (6.1):

$$\|x_i^\varepsilon - x_i\| = \| B_i^{-1}B_i x_i^\varepsilon - B_i^{-1}y_i^0 + B_i^{-1}y_i^0 - B_i^{-1}B_i x_i\| \le$$

$$\le c_0(\|y_i^0 - B_i x_i\| + \varepsilon_i), \quad i \ge i_0. \quad \blacksquare$$

Theorem 7.2. Let $A,(B_i)_{i\in\mathbb{N}}$ be consistent and closed at 0_Y. Let $A \in \mathcal{T}_+$, $B_i \in \mathcal{T}_+$, $i \in \mathbb{N}$ and let the set $\mathcal{B}$ be such that (7.3) holds for some $\alpha > 0$. Then the following assertions (I), (II) and (III) are equivivalent:

(I) equation (7.1) is uniquely solvable for every $v^0 \in V$ and the sequence $(B_i)_{i\in\mathbb{N}}$ is stable,

(II) the sequence $(B_i)_{i\in\mathbb{N}}$ is $\mathcal{BR}$-stable at 0_Y,

(III) $\mathcal{N}(A) = \{0\}$ and the sequence $(B_i)_{i\in\mathbb{N}}$ is $\mathcal{B}$-regular at 0_Y.

Proof. (I) ⇒ (II) by Theorem 6.1, (II) ⇔ (III) by Theorem 4.1, (II),(III) ⇒ (I) by Theorem 6.1 and the assumption that $A \in \mathcal{T}_+$. ■

Remark 7.1. If connecting operators p_i^X, p_i^Y, $i \in \mathbb{N}$ are linear then there exist, by assumption (N2), an index $i_0 \in \mathbb{N}$ and constants c_1, $c_1' > 0$, c_2 and $c_2' > 0$ so that (2.12) holds, i.e.

$$c_1'\|x_i\| \le \|p_i^X x_i\| \le c_1\|x_i\| \quad \forall x_i \in X_i, \ i \ge i_0,$$

$$c_2'\|y_i\| \le \|p_i^Y y_i\| \le c_2\|y_i\| \quad \forall y_i \in Y_i, \ i \ge i_0.$$

In this case we can get from estimation (7.4) the following estimations (7.6) and (7.7):

$$\|x_i^0 - x_i\| \le c_0\|y_i^0 - B_i x_i\| \le \frac{c_0}{c_2'}\|p_i^Y y_i^0 - p_i^V v^0 + p_i^V v^0 - p_i^Y B_i x_i\| \le$$

$$\le \frac{c_0}{c_2'}\left(\|p_i^Y y_i^0 - p_i^V v^0\| + \|p_i^V Au^0 - p_i^Y B_i x_i\|\right), \quad \forall x_i \in X_i \tag{7.6}$$

(the distance from the solution x_i^0 of (7.2) to any representative x_i of the solution u^0 of (7.1) is estimated by the sum of two errors: the discretization error of data and the consistency error of (A,B_i) on u^0, x_i),

$$\|p_i^U u^0 - p_i^X x_i^0\| \le \|p_i^U u^0 - p_i^X x_i\| + \|p_i^X x_i - p_i^X x_i^0\| \le$$

$$\le \|p_i^U u^0 - p_i^X x_i\| + \frac{c \cdot c_1}{c_2^1} \left(\|p_i^V y_i^0 - p_i^V v^0\| + \|p_i^V A u^0 - p_i^Y B_i x_i\|\right), \quad \forall x_i \in X_i \tag{7.7}$$

(the error of x_i^0 is estimated by the sum of three errors: the error of approximation of solution u^0 by some x_i, the error of approximation of data v^0 and the consistency error of (A,B_i) on solution u^0 and its approximation x_i).

Estimations for x_i^ε can be written down analogously.

References

1. Stummel F. Diskrete Konvergenz linearer Operatoren. I. Math. Ann., 1970, 190, 45-92
2. Stummel F. Discrete convergence of mappings. Proc. Conf. on Num. Analysis, Dublin, August 1972. Academic Press, New-York-London, 1973.
3. Vainikko G. Funktionalanalysis der Diskretisierungsmethoden. Teubner Verlagsges, Leipzig, 1976.

О ПОНЯТИИ ДИСКРЕТНОЙ СХОДИМОСТИ В СЛУЧАЕ НОРМИРОВАННЫХ ПРОСТРАНСТВ

Отто Карма

Резюме

В статье рассматривается одна возможная конкретизация введенной Ф.Стуммелем схемы дискретной сходимости. В этой схеме исследуются взаимноотношения следующих трех требований:

1) сходимость всех приближенных решений (ε_i-решений) аппроксимирующих уравнений к решению точного уравнения при сходимости невязки ε_i к нулю,

2) устойчивость ($\mathcal{BR}$-устойчивость в точке v) последовательности приближенных операторов,

3) регулярность ($\mathcal{B}$-регулярность в точке v) последовательности приближенных операторов.

Основные результаты сформулированы в виде теорем. Они подтерждают что и в этой схеме перечисленные выше три требования тесно связаны между собой а в определенных условиях они просто эквивалентны.

Acta et commentationes universitatis Tartuensis, 937, 1992, 63-76.

QUASIOPTIMAL ERROR ESTIMATE FOR THE REGULARIZED RITZ-GALERKIN METHOD WITH THE A-POSTERIORI CHOICE OF THE PARAMETER

Uno Hämarik

The regularized Ritz-Galerkin method for ill-posed problems with a-posteriori choice of the regularization parameter is considered. The error estimate which has the optimal order both in respect of data errors and the discretization error is given.

1. **Introduction**. Let H be a Hilbert space. Consider the equation

$$Au=f\ ,\quad f\in\mathcal{R}(A), \tag{1.1}$$

where $A\in\mathcal{L}(H,H)$, $A=A^*\geq 0$. We do not suppose that $\mathcal{R}(A)$ is closed and so in general problem (1.1) is ill-posed. Assume that instead of f only element $f_\delta\in H$ is available with $\|f_\delta - f\|\leq\delta$. To get an approximation to a solution of (1.1), we have to discretize the problem. For $h>0$, let P_h be an orthogonal projection in H. The Ritz-Galerkin method for (1.1) has form

$$A_h u_h = P_h f_\delta\ ,\quad A_h = P_h A P_h\ ,\ u_h\in\mathcal{R}(P_h).$$

We regularize this equation in a standard way (see Section 2) using initial approximation u_0 and regularization parameter r. The proper choice of $r=r(\delta,h)$ quarantees that approximation $u_{h,r}\in\mathcal{R}(P_h)$ we get converges to u_* as $\delta\to 0$, $h\to 0$, where u_* is the solution of (1.1) nearest to u_0. To estimate the error, we suppose

$$u_*=A^p v\ ,\ \|v\|\leq\varrho\ ,\quad u_0-u_*=A^p w\ ,\ \|w\|\leq\varrho,\quad p>0. \tag{1.2}$$

By this assumption the best error estimate which can be hoped is

$$\|u_{h,r}-u_*\| = O((\varrho\delta^p)^{1/(p+1)} + \varrho\|(I-P_h)A^p\|). \tag{1.3}$$

Namely, the error due to δ in any approximation method is at least $O((\varrho\delta^p)^{1/(p+1)})$ (see [15]) and so as $u_{h,r}\in\mathcal{R}(P_h)$, here the discretization error is at least $O(\|(I-P_h)u_*\|)$, which by assumption (1.2) takes form

$O(\|(I-P_h)A^p\|\varrho)$. For a class of equations (1.1) with self-regularization assumption estimate (1.3) follows from error estimate in [4], taking $r^{-p} \asymp (\delta/\varrho)^{p/(p+1)} + \|(I-P_h)A^p\|$. This choice of r gives estimate (1.3) for general equation (1.1) too, as shown in [5]. In [10] error estimate

$$\|u_{h,r} - u_*\| = O((\varrho\delta^p)^{1/(p+1)} + \varrho\|(I-P_h)A\|^{\min(p,1)}) \qquad (1.4)$$

was given, choosing r in the a-priori or in the a-posteriori way. In this paper we give estimate (1.3), choosing r in an a-posteriori way by similar rules as in [10].

Note that in [1-4, 6,7,9,11] regularized projection methods for equation (1.1) with nonselfadjoint operator were considered. The estimates to the discretization error in [4,6] are better than estimates in other works.

In Section 2 the regularized Ritz-Galerkin method is described. In Section 3 we give some auxiliary results , needed in Sections 4,5, where the a-priori and the a-posteriori choice of r are considered. In Section 6 the choice of h is discussed. The last section,Section 7, is devoted to the applications to integral equations. The examples are given where (see (1.3),(1.4))

$$\|(I-P_h)A^p\| = O(h^{\max(p,1)}), \quad \|(I-P_h)A\|^{\min(p,1)} = O(h^{\min(p,1)}).$$

2. Regularization of the Ritz-Galerkin method. Let $g_r: [0,a] \to \mathbb{R}$ ($r \geq 0$, $\|A\| \leq a$) be a Borel measurable function, satisfying conditions

1) $$\sup_{0\leq\lambda\leq a} |g_r(\lambda)| \leq \gamma r , \quad r \geq 0, \qquad (2.1)$$

2) $$\sup_{0\leq\lambda\leq a} \lambda^p |1-\lambda g_r(\lambda)| \leq \gamma_p r^{-p}, \quad r>0,\ 0\leq p\leq p_0, \qquad (2.2)$$

where p_0 , γ and γ_p are positive constants. Moreover , in the a-posteriori choice of r we assume that

3) the function $r \to |1-\lambda g_r(\lambda)|$ is decreasing for every $\lambda \geq 0$,

4) for $\lambda \in [0,a]$, $r>0$ the function $r \to g_r(\lambda)$ is continuously differentiable and it holds

4') $$\frac{\partial(g_s(\lambda))}{\partial s}\Big|_{s=r} \leq \gamma' \beta_r(\lambda)(1-\lambda g_r(\lambda)) , \quad \gamma' = \text{const} \qquad (2.3)$$

or

4'') $$\frac{\partial(g_s(\lambda))}{\partial s}\Big|_{s=r} \leq \gamma'(1-\lambda g_r(\lambda)) , \quad \gamma' = \text{const}. \qquad (2.4)$$

Here

$$\beta_r(\lambda) = \begin{cases} 1, \text{ if } p_0 = \infty, \\ (1-\lambda g_r(\lambda))^{1/p_0}, \text{ if } p_0 < \infty. \end{cases} \qquad (2.5)$$

We find the approximation to u_s in form

$$u_{h,r} = (I - g_r(A_h)A_h)P_h u_0 + g_r(A_h)P_h f_\delta. \qquad (2.6)$$

The following regularization algorithms have form (2.6) and corresponding functions $g_r(\lambda)$ satisfy conditions 1)-4), 4'), 4'').

1. The method of Lavrentiev

$$u_{h,r} = (r^{-1}I + A_h)^{-1}P_h f_\delta.$$

Here $u_0 = 0$, $g_r(\lambda) = (\lambda + r^{-1})^{-1}$, $p_0 = 1$.

2. The iterated method of Lavrentiev. Natural number $m \geq 1$ is given and $u_{h,r,0} = P_h u_0$. We find iteratively

$$u_{h,r,n} = (r^{-1}I + A_h)^{-1}(r^{-1}u_{h,r,n-1} + P_h f_\delta), \quad n = 1, \ldots, m. \qquad (2.7)$$

As the approximation we take $u_{h,r} = u_{h,r,m}$. Here $g_r(\lambda) = \lambda^{-1}[1-(1+r\lambda)^{-m}]$, $p_0 = m$.

3. The method of iteration (explicit scheme). Let $\mu \in (0, 1/a)$, $u_{h,0} = P_h u_0$,

$$u_{h,r} = (I - \mu A_h)u_{h,r-1} + \mu P_h f_\delta, \quad r = 1, 2, \ldots .$$

Here $g_r(\lambda) = \lambda^{-1}[1-(1-\mu\lambda)^r]$, $p_0 = \infty$.

4. The method of iteration (implicit scheme). Let $\mu > 0$ and

$$u_{h,r} = (A_h + \mu I)^{-1}(\mu u_{h,r-1} + P_h f_\delta), \quad r = 1, 2, \ldots .$$

Here $g_r(\lambda) = \lambda^{-1}[1-(\mu/(\mu+\lambda))^r]$, $p_0 = \infty$.

More information about these and the other regularization methods can be found in [4,10-13,15]. Computational schemes and their realization for integral equations of the first kind are given in [10].

3. Some auxiliary results. Here we generalize some results of [5], needed in Sections 4,5. Denote

$$\xi_{h,p} := \|(I - P_h)A^p\|, \qquad \xi_h := \inf_{\alpha > 0} (2^{1/2}\xi_{h,\alpha})^{1/\alpha},$$

$$S_{h,r} := I - A_h g_r(A_h), \qquad B_{h,r} := \begin{cases} I, & \text{if } p_0 = \infty, \\ S_{h,r}^{1/p_0}, & \text{if } p_0 < \infty. \end{cases} \qquad (3.1)$$

We use the inequality of moments (see[8]): if $D_1 \in \mathcal{L}(H,H)$, $D_1 = D_1^* \geq 0$, $v \in H$ and $0 \leq \alpha \leq 1$, then $\|D_1^\alpha v\| \leq \|D_1 v\|^\alpha \|v\|^{1-\alpha}$. From here with $v = D_2 w$, $D_2 \in \mathcal{L}(H,H)$, $w \in H$ follows

$$\|D_1^\alpha D_2\| \leq \|D_1 D_2\|^\alpha \|D_2\|^{1-\alpha} \qquad (3.2)$$

Taking here $D_1 = A^p$, $D_2 = I - P_h$, $\alpha = q/p$, we get

$$\xi_{h,q} \le \xi_{h,p}^{q/p} \quad (\forall q \le p). \tag{3.3}$$

In the following c is a general positive constant, not depending on r, h, δ and ϱ and taking different values in different places.

Lemma 3.1. [5]. For every $\alpha \in (0,1]$

$$\|(P_h A P_h)^\alpha - P_h A^\alpha\| \le c\xi_{h,\alpha}. \tag{3.4}$$

Lemma 3.2. Let $s \ge 0$, $0 \le t \le p_0 + s$. Then

$$r^t \|B_{h,r}^s A_h^t S_{h,r} P_h A^p\| \le c(r^{-p'} + \xi_{h,p}), \quad p' = \min\{p, p_0+s-t\}. \tag{3.5}$$

Proof. Denote $D_{h,r}^{s,t} := B_{h,r}^s A_h^t S_{h,r}$. From (2.2) follows that for every $\tau \ge 0$ (see (2.5))

$$\|D_{h,r}^{s,\tau}\| = \|B_{h,r}^s S_{h,r} A_h^\tau\| \le \sup_{0\le\lambda\le a} |\beta_r^s(\lambda)(1-\lambda g_r(\lambda))| \lambda^\tau =$$

$$= \sup_{0\le\lambda\le a} \left[|1-\lambda g_r(\lambda)| \lambda^{\tau p_0/(p_0+s)}\right]^{(p_0+s)/p_0} \le cr^{-\min\{\tau, p_0+s\}}. \tag{3.6}$$

If $p \le 1$, then (3.5) follows from (3.4) and (3.6):

$$\|D_{h,r}^{s,t} P_h A^p\| \le \|D_{h,r}^{s,t}\| \|P_h A^p - A_h^p\| + \|D_{h,r}^{s,t+p}\| \le cr^{-t}\xi_{h,p} + cr^{-p'-t}.$$

If $p > 1$, we use representation

$$P_h A^p = \sum_{i=0}^{k-1} A_h^i P_h A (I - P_h) A^{p-i-1} + A_h^k P_h A^{p-k} =: \Sigma + \sigma,$$

where k is the integer part of $\min\{p, p_0+s-t+1\}$. In case $p-k \le 1$ we have from (3.3), (3.4), (3.6)

$$\|D_{h,r}^{s,t} P_h A^p\| \le \sum_{i=0}^{k-1} \|D_{h,r}^{s,t+i}\| \|A(I-P_h)\| \|(I-P_h)A^{p-i-1}\| +$$

$$+ \|D_{h,r}^{s,t+k}(P_h A^{p-k} - A_h^{p-k})\| + \|D_{h,r}^{s,t+p}\| \le c \sum_{i=0}^{k-1} r^{-t-i}\xi_{h,1}\,\xi_{h,p-i-1} +$$

$$+ cr^{-\min\{t+k, p_0+s\}}\xi_{h,p-k} + cr^{-t-p'} \le cr^{-t}\left[\sum_{i=0}^{k} r^{-i}\xi_{h,p}^{(p-i)/p} + r^{-p'}\right] \le$$

$$\le cr^{-t}(\xi_{h,p} + r^{-p'}). \tag{3.7}$$

Let $p-k > 1$. Then $k \ge p_0+s-t$ and (3.6) gives

$$\|D_{h,r}^{s,t}\sigma\| = \|D_{h,r}^{s,t} A_h^k P_h A^{p-k}\| \le \|D_{h,r}^{s,t+k}\| \|A^{p-k}\| \le cr^{-t-p'}\|A^{p-k}\|.$$

Term $\|D_{h,r}^{s,t}\Sigma\|$ may be estimated as in (3.7), hence (3.7) holds in case $p-k > 1$, too. Lemma 3.2 is proved.

Lemma 3.3 . Let $k \in \mathbb{R}$ and function $G_{r,k}:[0,\|A\|]\to\mathbb{R}$, $r>0$ satisfies for every $\lambda\in[0,\|A\|]$ conditions

$$\lambda|G_{r,k}(\lambda)| \le \varkappa_1 r^k \tag{3.8}$$

$$\lambda^{1/2}|G_{r,k}(\lambda)| \le \varkappa_* r^{k+1/2}, \tag{3.9}$$

where $\varkappa_1, \varkappa_*$ are positive constants. Then for every $\alpha\in(0,1/2]$

$$\|G_{r,k}(P_hAP_h)P_hA(I-P_h)\| \le c_\alpha r^k \cdot \begin{cases} r^\alpha \xi_{h,\alpha}, & \text{if } r^\alpha\xi_{h,\alpha}\le 1, \\ r^{1/2}\xi_{h,\alpha}^{1/(2\alpha)}, & \text{if } r^\alpha\xi_{h,\alpha}>1, \end{cases} \tag{3.10}$$

where

$$c_\alpha = \varkappa_*^{2\alpha}\left[\varkappa_*^{4\alpha}+\varkappa_1^{4\alpha}\right]^{(1-2\alpha)/(4\alpha)} \le 2^{(1-2\alpha)/(4\alpha)} \max(\varkappa_*,\varkappa_1). \tag{3.11}$$

Proof. Denote $D_{r,k}:=P_hG_{r,k}(P_hAP_h)$, $x:=\|(I-P_h)AD_{r,k}\|$. Then for every $\alpha\in(0,1/2]$

$$x \le \|(I-P_h)A^\alpha\|\|A^{1-\alpha}D_{r,k}\| = \xi_{h,\alpha}\|A^{1-\alpha}D_{r,k}\|. \tag{3.12}$$

From (3.8) follows

$$\|AD_{r,k}\|^2 = \|P_hAD_{r,k}\|^2+\|(I-P_h)AD_{r,k}\|^2 \le (\varkappa_1 r^k)^2+x^2. \tag{3.13}$$

Using inequality of moments (3.2) with $D_1=A^{1/2}$, $D_2=A^{1/2}D_{r,k}$, $1-2\alpha$ instead of α and inequalities (3.13), (3.9), we have

$$\|A^{1-\alpha}D_{r,k}\|^2 \le \|AD_{r,k}\|^{2(1-2\alpha)}\|A^{1/2}D_{r,k}\|^{4\alpha} \le \left[(\varkappa_1 r^k)^2+x^2\right]^{1-2\alpha}(\varkappa_* r^{k+1/2})^{2\alpha}$$

The last inequality with (3.12) implies that

$$x^2 \le d^2\left[(\varkappa_1 r^k)^2+x^2\right]^{1-2\alpha}, \quad d=\varkappa_*^{2\alpha}\xi_{h,\alpha}r^{(2k+1)\alpha} \tag{3.14}$$

We show now that from (3.14) follows

$$x^2 \le d^2(\varkappa_1 r^k)^{2(1-2\alpha)}\left[1+d^2(\varkappa_1 r^k)^{-4\alpha}\right]^{(1-2\alpha)/(2\alpha)} \tag{3.15}$$

Namely noting $y:=x^2\cdot d^{-2}(\varkappa_1 r^k)^{4\alpha-2}$, $s:=d^2(\varkappa_1 r^k)^{-4\alpha}$, we see that the implication (3.14)$\Rightarrow$(3.15) is equivalent to the implication

$$y \ge \bar y := (1+s)^{(1-2\alpha)/(2\alpha)} \Rightarrow \varphi(y) := y^{1/(1-2\alpha)} - ys - 1 > 0 \text{ for } y \ge \bar y.$$

The last implication holds since $\varphi(\bar y)=\bar y-1>0$ and $\varphi'(y)>0$ for $y\ge\bar y$. Inequality (3.15) implies that

$$x \le \varkappa_*^{2\alpha}\varkappa_1^{1-2\alpha}r^{k+\alpha}\xi_{h,\alpha}\left[1+\varkappa_*^{4\alpha}\varkappa_1^{-4\alpha}\xi_{h,\alpha}^2 r^{2\alpha}\right]^{(1-2\alpha)/(4\alpha)}$$

From here follows (3.10). Lemma 3.3 is proved.

Corollary 3.4. For every $\alpha\in(0,1/2]$

$$\|g_r(A_h)P_hA(I-P_h)\| \le c_\alpha(r^\alpha\xi_{h,\alpha}+r^{1/2}\xi_{h,\alpha}^{1/(2\alpha)}), \tag{3.16}$$

$$\|B^s_{h,r} S_{h,r} P_h A(I-P_h)\| \le c_\alpha (r^{\alpha-t}\xi_{h,\alpha} + r^{-1/2}\xi^{1/(2\alpha)}_{h,\alpha}), \quad \text{if } s\ge 0,\ p_0+s\ge 1, \tag{3.17}$$

where c_α is given by (3.11) with $\varkappa_1 = \gamma_0 + 1$, $\varkappa_* = [\gamma(\gamma_0+1)]^{1/2}$ for (3.16), with

$$\varkappa_1 = \gamma^{(p_0+s)/p_0}_{p_0/(p_0+s)}, \quad \varkappa_* = \gamma^{(p_0+s)/p_0}_{p_0/(2(p_0+s))} \tag{3.18}$$

for (3.17).

Proof. We get these estimates from (3.10), so as $G_{r,0}(\lambda) = g_r(\lambda)$ and $G_{r,-1}(\lambda) = \beta^s_r(\lambda)(1-\lambda g_r(\lambda))$ (see (2.5)) with $s \ge 1-p_0$ satisfy (3.8), (3.9) on the base of (2.1), (2.2), (3.6).

Note that estimates (3.16), (3.5) with $s=t=0$ were given in [5].

4. A-priori choice of r. In [5] the case, where not only f but also A were given approximately, was considered : we know $A_\delta \in \mathcal{L}(H,H)$, $A_\delta = A^*_\delta \ge 0$ with $\|A_\delta - A\| \le \delta$ and use in (2.6) operator $A_h = P_h A_\delta P_h$. The following theorem was proved.

Theorem 4.1. 1. Suppose that $P_h \to I$ pointwise, $\xi_{h,t} \to 0$ $(h\to 0)$. If $r = r(\delta,h)$ is chosen so that

$$r\delta \to 0, \quad r \to \infty, \quad r\,\xi_{h,*} \le c, \qquad (\delta\to 0, h\to 0), \tag{4.1}$$

where $\xi_{h,*} := \inf\limits_{\alpha>0} 2^{1/(4\alpha^2)}\,\xi^{1/\alpha}_{h,\alpha}$, then $u_{h,r} \to u_*$ $(\delta\to 0, h\to 0)$.

2. a) If (1.2) holds with $p \le p_0$ and r is chosen by rule

$$c_1[(\delta/\varrho)^{1/(p+1)} + \xi_h] \le r^{-1} \le c_2[(\delta/\varrho)^{1/(p+1)} + \xi^{1/p}_{h,p}],$$

then (1.3) holds.

b) If (1.2) holds with $p > p_0$, then choice of r by rule

$$c_1[(\delta/\varrho)^{1/(p_0+1)} + (\xi_h^{1/2}\,\xi_{h,p})^{2/(2p_0+1)}] \le r^{-1} \le$$

$$\le c_2[(\delta/\varrho)^{1/(p_0+1)} + (\xi_h^{1/2}\,\xi_{h,p})^{2/(2p_0+1)}] + \xi^{1/p_0}_{h,p}]$$

gives

$$\|u_{h,r} - u_*\| \le c[(\varrho\delta^{p_0})^{1/(p_0+1)} + \varrho(\xi_h^{1/2}\,\xi_{h,p})^{2p_0/(2p_0+1)} + \varrho\,\xi_{h,p}]. \tag{4.2}$$

We also outline the proof of Theorem 4.1 in case $A_\delta = A$. We have

$$u_{h,r} - u_* = S_{h,r}P_h(u_0-u_*) - [I - g_r(A_h)P_hA(I-P_h)](I-P_h)u_* + \\ + g_r(A_h)P_h(f_\delta - f) \tag{4.3}$$

and estimating by (2.1), (3.16) yields

$$\|u_{h,r} - u_*\| \le \inf_{\alpha\in(0,1/2]}[\,\|S_{h,r}P_h(u_0-u_*)\| + \\ + (1 + c_\alpha(r^\alpha\xi_{h,\alpha} + r^{1/2}\xi^{1/(2\alpha)}_{h,\alpha}))\|(I-P_h)u_*\| + \gamma r\delta\,] =: \psi(r) \tag{4.4}$$

The Banach-Steinhaus theorem gives $\|S_{h,r}P_h(u_0-u_*)\| \to 0$ $(r\to\infty,\ h\to 0)$, since by (2.2) $\|S_{h,r}P_h\| \le \gamma_0$, $u_0-u_* \in \overline{\mathcal{R}(A)}$ and (3.5) gives $\|S_{h,r}P_hA\| \to 0$ $(r\to\infty,\ h\to 0)$. Hence choice of r by (4.1) guarantees $u_{h,r}\to u_*$ $(\delta\to 0,\ h\to 0)$. If (1.2) holds, then $\|(I-P_h)u_*\| \le \varrho\xi_{h,p}$ and the estimation of $\|S_{h,r}P_h(u_0-u_*)\| \le \varrho\|S_{h,r}P_hA^p\|$ by (3.5) gives to (4.4) the form

$$\|u_{h,r}-u_*\| \le c\left\{\varrho\left[r^{-\min(p,p_0)}+\left(1+r^{1/2}\xi_h^{1/2}\right)\xi_{h,p}\right]+r\delta\right\}.$$

Now the second part of Theorem 4.1 is an easy consequence of the choice of r.

Note that in [10] convergence $u_{h,r}\to u_*$ $(\delta\to 0,\ h\to 0)$ was proved by the choice of r by rule $r\to\infty$, $r\delta\to 0$, $r\xi_{h,1}\le c$, which is more restrictive than (4.1) (see (3.3), Section 7).

5. <u>A-posteriori choice of r</u>. Applying A_h to (4.3) leads to

$$A_hu_{h,r}-P_hf_\delta = S_{h,r}[A_h(u_0-u_*)-P_hA(I-P_h)u_*]+S_{h,r}P_h(f_\delta-f).$$

Let condition 3) for $g_r(\lambda)$ hold. Applying yet $B_{h,r}^s$ with $s\ge 0$ and taking into account the fact, that due to (2.1)

$$\|B_{h,r}^sS_{h,r}\| \le \sup_{0\le\lambda\le a}|1-\lambda g_r(\lambda)|^{1+s/p_0} \le \sup_{0\le\lambda\le a}|1-\lambda g_0(\lambda)|^{1+s/p_0} = 1,$$

we have

$$\begin{aligned}\|B_{h,r}^s(A_hu_{h,r}-P_hf_\delta)\|-\delta \le \|B_{h,r}^sS_{h,r}[A_h(u_0-u_*)-P_hA(I-P_h)u_*]\| \le \\ \le \|B_{h,r}^s(A_hu_{h,r}-P_hf_\delta)\|+\delta \quad (\forall s\ge 0).\end{aligned} \tag{5.1}$$

From polar representation $P_hA^{1/2}=A_h^{1/2}U$, $\|U\|\le 1$ and (3.6) follows that the middle term here is $\mathcal{O}(\|B_{h,r}^sS_{h,r}\,A_h^{1/2}\|)\to 0$ $(r\to\infty)$, hence $\lim\limits_{r\to\infty}\|B_{h,r}^s(A_hu_{h,r}-P_hf_\delta)\| \le \delta$ $(\forall s\ge 0)$. Let us now formulate the rules for choice of r.

<u>Rule</u> 1. Let $1<b_0\le b$. If $\|B_{h,0}(A_hu_0-P_hf_\delta)\| \le b_0\delta$, choose $r=0$. Otherwise choose $0<r\le\bar r := \xi_{h,*}^{-1}$ (see (4.1)) such that

$$b_0\delta \le \|B_{h,r}(A_hu_{h,r}-P_hf_\delta)\| \le b\delta. \tag{5.2}$$

If there is no $r\le\bar r$, such that (5.2) holds, choose $r=\bar r$.

For iteration methods it is convenient to apply the following

<u>Rule</u> 2. Let $1<b$, $0<\Theta<1$. If $\|B_{h,0}(A_hu_0-P_hf_\delta)\| \le b\delta$, choose $r=0$. Otherwise choose $0<r\le\bar r$ such that there is a $r'\in[\Theta r,r]$ with

$$\|B_{h,r}(A_hu_{h,r}-P_hf_\delta)\| \le b\delta \le \|B_{h,r'}(A_hu_{h,r'}-P_hf_\delta)\|. \tag{5.3}$$

If there is no $r \le \bar{r}$, such that (5.3) holds, choose $r = \bar{r}$ or $r = [\bar{r}]$, where $[\bar{r}]$ is the largest integer, not greater than $\bar{r}$.

In case $p_0 > 1$ we also consider Rules 1',2', which we get from Rules 1,2, respectively, substituting $B_{h,r}$ by I.

To prove the convergence rates for these parameter selection rules, we need the following lemma the assertion of which is based on an idea of T.Raus (Lemma 4.4 in [13]).

Lemma 5.1. Let one of the following conditions a),b) hold:

a) $g_r(\lambda)$ satisfies conditions 1) - 4),4') and it holds

$$\|B_{h,r_\delta}(A_h u_{h,r_\delta} - P_h f_\delta)\| \le b\delta , \tag{5.4}$$

b) $g_r(\lambda)$ satisfies conditions 1)- 4),4''), it holds $p_0 \ge 1$ and $\|A_h u_{h,r_\delta} - P_h f_\delta\| \le b\delta$.

Then for every $\alpha \in (0,1/2]$ it holds $\|S_{h,r_\delta} P_h (u_0 - u_*)\| \le \inf_{r \ge 0} \psi_\alpha(r)$,

$$\psi_\alpha(r) := \|S_{h,r} P_h (u_0 - u_*)\| + \bar{\gamma} c_\alpha (\alpha^{-1} r^\alpha \xi_{h,\alpha} + 2 r^{1/2} \xi_{h,\alpha}^{1/(2\alpha)}) \|(I - P_h) u_*\| +$$

$$+ \bar{\gamma}(b+1) r\delta ,$$

where $\bar{\gamma} > \gamma'$ (see (2.3) in case a),(2.4) in case b)) and c_α is given by (3.11),(3.18) with $s=1$ in case a), with $s=0$ in case b).

Proof. Let a) hold. Then by (5.1),(5.4) we have

$$\|B_{h,r_\delta} S_{h,r_\delta} [A_h (u_0 - u_*) - P_h A (I - P_h) u_*]\| \le (b+1)\delta . \tag{5.5}$$

Let $\alpha \in (0,1/2]$ be fixed and r_0 be the minimum point for $\psi_\alpha(r)$. We show that $r_\delta \ge r_0$. If $r_0 = 0$, it is obvious. If $r_0 \ne 0$, then noting $v := P_h (u_0 - u_*)$ we have

$$\psi_\alpha'(r_0) = 0.5 \|S_{h,r_0} v\|^{-1} \frac{\partial}{\partial r} (\|S_{h,r} v\|^2)\Big|_{r=r_0} +$$

$$+ \bar{\gamma} c_\alpha (r_0^{\alpha-1} \xi_{h,\alpha} + r_0^{-1/2} \xi_{h,\alpha}^{1/(2\alpha)}) \|(I - P_h) u_*\| + \bar{\gamma}(b+1)\delta = 0 . \tag{5.6}$$

Let $Q(\lambda)$ be the spectral family of the projectors of operator A_h. Using condition 4') and equality

$$\frac{\partial (1 - \lambda g_r(\lambda))^2}{\partial r} = -2\lambda (1 - \lambda g_r(\lambda)) \frac{\partial (g_r(\lambda))}{\partial r} ,$$

we get

$$- \frac{\partial}{\partial r} (\|S_{h,r} v\|^2) = - \int_0^{\|A_h\|} \frac{\partial}{\partial r} (1 - \lambda g_r(\lambda))^2 \, d\langle Q(\lambda) v, v \rangle \le$$

$$\le 2\gamma' \int_0^{\|A_h\|} \lambda \beta_r(\lambda) (1 - \lambda g_r(\lambda))^2 \cdot d\langle Q(\lambda) v, v \rangle = 2\gamma' (A_h B_{h,r} S_{h,r} v, S_{h,r} v) \le$$

$$\le 2\gamma' \|B_{h,r} S_{h,r} A_h v\| \|S_{h,r} v\| .$$

From the last inequality, (3.17) and (5.6), we have

$$\|B_{h,r_0}S_{h,r_0}[A_h(u_0-u_*)-P_hA(I-P_h)u_*]\| \ge \|B_{h,r_0}S_{h,r_0}A_h(u_0-u_*)\| -$$

$$- \|B_{h,r_0}S_{h,r_0}P_hA(I-P_h)\|\,\|(I-P_h)u_*\| \ge -\frac{1}{2\gamma}\,\|S_{h,r_0}v\|\ \frac{\partial}{\partial r}\,(\|S_{h,r}v\|)\Big|_{r=r_0} -$$

$$- c_\alpha(r_0^{\alpha-1}\xi_{h,\alpha}+r_0^{-1/2}\xi_{h,\alpha}^{1/(2\alpha)})\,\|(I-P_h)u_*\| > (b+1)\delta.$$

The last inequality with (5.5) and condition 3) implies that $r_\delta > r_0$. Using once more condition 3) we get the assertion of Lemma 5.1:

$$\|S_{h,r_\delta}P_h(u_0-u_*)\| \le \|S_{h,r_0}P_h(u_0-u_*)\| \le \psi_\alpha(r_0) = \inf_{r\ge 0}\psi_\alpha(r).$$

If condition b) holds instead of a), in the proof $B_{h,r}$ and $\beta_r(\lambda)$ should to be replaced by I and 1 respectively. Lemma 5.1 is proved.

Theorem 5.2. Let $r=r(\delta,h)$ be chosen by one of Rules 1, 2, 1', 2'. By use of Rules 1, 2 assume that conditions 1) – 4), 4') for $g_r(\lambda)$ are satisfied and define $\bar p := p_0$; for Rules 1', 2' assume that conditions 1)- 4),4") are satisfied with $p_0>1$ and define $\bar p:=p_0-1$. If $P_h\to I$ pointwise and $\xi_{h,t}\to 0$ $(h\to 0)$, then $u_{h,r}\to u_*$ $(\delta\to 0,\ h\to 0)$. If (1.2) holds , then

$$r(\delta,h) \le c((\delta/\varrho)^{-1/(p'+1)}+(\varrho/\delta)\xi_{h,p}),\quad p'=\min(p,\bar p) \tag{5.7}$$

and it holds estimate (1.3) in case $p\le\bar p$, estimate (5.8) in case $p>\bar p$:

$$\|u_{h,r}-u_*\| \le c[(\varrho\delta^{\bar p})^{1/(\bar p+1)} + \varrho\,\xi_{h,p} + \varrho\,\xi_{h,*}^{\min(p,p_0)}]. \tag{5.8}$$

Proof. For $r\le\bar r$ from (4.4) we have

$$\|u_{h,r}-u_*\| \le \|S_{h,r}P_h(u_0-u_*)\| + c\|(I-P_h)u_*\| + \gamma r\delta. \tag{5.9}$$

Term $\|(I-P_h)u_*\|\to 0$ $(h\to 0)$ and has estimate $\varrho\,\xi_{h,p}$, if (1.2) holds.

Consider the first term in estimate (5.9). Let our choice of r gives $r(\delta,h) < \bar r$. Then for $r_\delta = r(\delta,h)$ assumptions of Lemma 5.1 hold (assumption a), if Rule 1 or 2 is used, assumption b), if Rule 1' or 2' is used) and this lemma gives $\|S_{h,r(\delta,h)}P_h(u_0-u_*)\| \le \inf_{r\ge 0}\psi_\alpha(r)$ $(\forall\,\alpha\in(0,1/2])$. Term $\psi_\alpha(r)$ may be considered in a similar way as $\psi(r)$ in (4.4). From proof of Theorem 4.1 we get for $\|S_{h,r(\delta,h)}P_h(u_0-u_*)\|$ convergence and under condition (1.2) estimate in the right hand side of (1.3) (if $p\le p_0$) or (4.2) (if $p>p_0$). Estimate (4.2) is not worse than (5.8), since $(\xi_h^{1/2}\,\xi_{h,p})^{2p_0/(2p_0+1)} \le \xi_h^{p_0} + \xi_{h,p}$. In case $r(\delta,h)=\bar r$ the proof of Theorem 4.1 gives $\|S_{h,r(\delta,h)}P_h(u_0-u_*)\|\to 0$ $(\delta\to 0,\ h\to 0)$ and by assumption (1.2) estimate

$$\|S_{h,r(\delta,h)}P_h(u_0-u_*)\| \le cQ\,[\xi_{h,*}^{\min(p,p_0)} + \xi_{h,p}].$$

So as $\xi_{h,*}^{p} \le c\xi_{h,p}$ $(\forall p>0)$, estimates of $\|S_{h,r(\delta,h)}P_h(u_0-u_*)\|$ are in harmony with (1.3), (5.8).

Consider the last term in (5.9). Let Rule 1 gives $r = r(\delta,h)>0$. Then from the left-hand side of (5.1),(5.2) and (3.17) due to $r\le\bar r$ we have

$$(b_0-1)r\delta \le r\|B_{h,r}S_{h,r}A_h(u_0-u_*)\| + r\|B_{h,r}S_{h,r}P_hA(I-P_h)\|\,\|(I-P_h)u_*\| \le$$

$$\le r\|B_{h,r}S_{h,r}A_h(u_0-u_*)\| + c\|(I-P_h)u_*\| \to 0 \quad (\delta\to 0,\ h\to 0), \tag{5.10}$$

since $r\|B_{h,r}S_{h,r}A_h(u_0-u_*)\|\to 0$ $(r\to\infty)$ due to (3.5) and the Banach-Steinhaus theorem. Let (1.2) holds. By (5.10) and (3.5) we get

$$(b_0-1)r\delta \le r\|B_{h,r}S_{h,r}A_hP_hA^{P}\|Q + cQ\xi_{h,p} \le cQ(r^{-p'} + \xi_{h,p}) \tag{5.11}$$

with $p' = \min(p,\bar p)$. In case $r \ge (\delta/Q)^{-1/(p'+1)}$ from here we have

$$r\delta \le c[(Q\delta^{p'})^{1/(p'+1)} + Q\xi_{h,p}]. \tag{5.12}$$

In case $r< (\delta/Q)^{-1/(p'+1)}$ (5.12) is trivial. Inequality (5.12) implies ,that (5.7) holds and the last term in (5.9) has also estimate, needed for (1.3) and (5.8). Theorem 5.2 for Rule 1 is proved. For other rules in a similar way we get analogues of (5.10) - (5.12) with following substitutions there: $B_{h,r}\to I$ for Rule 1'; $b_0\to b$, $r\to r'$ for Rule 2; $B_{h,r}\to I$, $b_0\to b$, $r\to r'$ for Rule 2'. At that condition $p_0>1$ for Rules 1', 2' is needed. These analogues of (5.10), (5.12) give assertments of Theorem 5.2 , taking for Rules 2, 2' into account also inequality $r \le r'/\Theta$.

Remark 1. If $p\le\bar p$ and in (1.3) the discretization error does not dominate, i.e. $\xi_{h,p} \le c((\delta/Q)^{p/(p+1)})$, then (5.7) has form $r\le c((\delta/Q)^{-1/(p+1)})$.

Remark 2. In [10] Rules 1',2' for the choice of r with the upper bound $r\le r_{max} = \xi_{h,1}^{-1}$ were proposed and estimate (1.4) under condition (1.2) with $p\le p_0-1$ (provided $p_0>1$) was given. From proof of Theorem 5.2 follows that by upper bound $r_{max} = \xi_{h,\alpha}^{-1/\alpha}$, $\alpha>0$ it holds better than (1.4) estimate (1.3) under assumption (1.2) with $p\in[\alpha, p_0-1]$ (provided $p_0>1$, $\alpha<p_0-1$) for Rules 1',2', with $p\in[\alpha,p_0]$ (provided $\alpha<p_0$) for Rules 1,2. However, our upper bound $\bar r = \xi_{h,*}^{-1}$ is also too small for some cases: if (1.2) holds with $p>p_0$, due to bound $r\le\bar r$ we get estimate (5.8), but not (4.2) as by a-priori choice of $r>\bar r$. if information (1.2) is not known, we need bound $\bar r$ to guarantee convergence $u_{h,r}\to u_*$ $(\delta\to 0,$ $h\to 0)$ for nonsmooth u_* too.

Remark 3. Analogues of Rules 1,2 for the choice of r in regularization methods for the equations without discretization were proposed and analysed in [12,13]. Note that in the Lavrentiev method $(u_{h,r}=u_{h,r,1})$ and in iterated Lavrentiev method $B_{h,r}(A_h u_{h,r,m}-P_h f_\delta) = A_h u_{h,r,m+1}-P_h f_\delta$ $(m\ge 1)$ (see (2.7)).

Remark 4. It holds the following generalization of Theorem 5.2. Let $g_r(\lambda)$ satisfy 1)-4) and 4'), where $\beta_r(\lambda)$ is replaced by $\beta_r^s(\lambda)$ with $s\in[0,1]$, $s>1-p_0$. Let $r=r(\delta,h)$ be chosen by Rule 1" or Rule 2", which we get from Rules 1,2, respectively, substituting $B_{h,r}$ by $B_{h,r}^s$. If $P_h\to I$ pointwise and $\xi_{h,1}\to 0$ $(h\to 0)$, then $u_{h,r}\to u_*$ $(\delta\to 0, h\to 0)$. If (1.2) holds, then it holds (5.7) with $\bar p=p_0+s-1$ and estimates (1.3), (5.8) in cases $p\le\bar p$ and $p>\bar p$ respectively. The proof is the same as for Theorem 5.2, using analogue of Lemma 5.1.

Remark 5. A more careful analysis replaces term $\xi_{h,*}$ in Theorems 4.1 and 5.2 by $\bar\xi_{h,*}=\inf\limits_{\alpha>0}(2\alpha)^{-1/(2\alpha)}\xi_{h,\alpha}^{1/\alpha}$.

6. **Choice of h**. Consider such choice of h that the discretization error in (1.3) does not dominate above the term caused by the data error. By known information (1.2) the natural choice of h is by rule $\xi_{h,p}\sim(\delta/\varrho)^{p/(p+1)}$ $(p\le p_0)$. For the case, if there is no information like (1.2) about the "smoothness" of u_*, u_0-u_*, in [10] the choice of h by rule

$$\xi_{h,\lambda}^{1/\lambda}\sim\delta \tag{6.1}$$

with $\lambda=1$ was proposed. Here we propose to choose λ in (6.1) by rule

$$\lambda=1 \text{ for } p_0=\infty;\quad \text{if } p_0<\infty, \text{ then } \left\{\begin{array}{l}\lambda=p_0/(p_0+1) \text{ for Rules 1,2,}\\ \lambda=(p_0-1)/p_0 \text{ for Rules 1',2'.}\end{array}\right\} \tag{6.2}$$

Note that due to (3.3) rule (6.1) with smaller λ generally gives a greater h (see also (7.3)). The proposed choice of h ensures that if (1.2) holds (with unknown p and ϱ), then under the assumptions of Theorem 5.2 it holds $\|u_{h,r}-u_*\|\le c\delta^{p/(p+1)}$ $(p\le\bar p)$. Namely, in case $p\le\lambda$ we have by (3.3) $\xi_{h,p}\le\xi_{h,\lambda}^{p/\lambda}\sim\delta^p\le\delta^{p/(p+1)}$, in case $p>\lambda$ we have $\xi_{h,p}\le c\xi_{h,\lambda}\sim c\delta^\lambda$.

7. **Application to integral equations**. Let $Au(t)=\int_0^1 K_1(t,s)u(s)ds$, where $K_1(t,s)=K_1(s,t)$ is such that $A=A^*\ge 0$ in $L_2=L_2[0,1]$. Operator A^n is also the integral operator with iterated kernel $K_n(t,s)=K_n(s,t)=\int_0^1 K_{n-1}(\sigma,t)K_1(\sigma,s)d\sigma$, $n=2,3,\dots$. Let l_n, $n=1,2,\dots$ are greatest numbers for which

$$\int_0^1\int_0^1 \left|\frac{\partial^{l_n}\mathcal{K}_n(t,s)}{\partial t^{l_n}}\right|^2 dt\,ds<\infty, \quad n=1,2,\dots . \tag{7.1}$$

If (7.1) holds for every l_n, define $l_n=\infty$. Then $\mathcal{R}(A^n)\subset H^{l_n}$, where $H^{l_n}=H^{l_n}[0,1]$ is a Sobolev space. Namely, if $z(t)\in\mathcal{R}(A^n)$, then $z(t)=\int_0^t \mathcal{K}_n(t,s)v(s)ds$ with $v\in L_2$ and $z^{(l_n)}(t)=\int_0^t \partial^{l_n}\mathcal{K}_n(t,s)/\partial t^{l_n}\, v(s)ds\in L_2$ due to (7.1), hence $z(t)\in H^{l_n}$. From $\mathcal{R}(A^n)\subset H^{l_n}$ follows $\mathcal{R}(A^p)\subset H^{pl_n/n}$ $(\forall p\le n)$ (see e.g. [14] (Lemma 2.4)).

Let $\mathcal{R}(P_h)$ be a spline space with degree k-1. Then $\|(I-P_h)u\|=O(h^{\min(\tau,k)})$ $(\forall u\in H^\tau)$, hence

$$\xi_{h,p}=\|(I-P_h)A^p\|=O(h^{\min(pl_n/n,k)}) \quad (p\le n). \tag{7.2}$$

Now let us compare the estimates of discretization error $\xi_{h,p}$ and $\xi_{h,1}^{\min(p,1)}$ in (1.3),(1.4) respectively. If $p\le 1$, (3.3) gives $\xi_{h,p}\le\xi_{h,1}^p$, if $p>1$, we have $\xi_{h,p}\le\|A^{p-1}\|\xi_{h,1}$. The use of (7.2) gives more expressive comparative examples:

$$\text{if } l_1^{-1}\le p<k=1, \quad \text{then } \xi_{h,p}=O(h), \quad \xi_{h,1}^{\min(p,1)}=O(h^p),$$

$$\text{if } 1=l_1<p\le l_n=n\le k, \quad \text{then } \xi_{h,p}=O(h^p), \quad \xi_{h,1}^{\min(p,1)}=O(h).$$

Now consider the choice of h and r. So as (7.2) gives

$$\xi_{h,\lambda}^{1/\lambda}=O(h^{\min(l_1,k/\lambda)}) \quad (\forall\lambda\le 1), \tag{7.3}$$

on the basis of Section 6 the choice of h by rule

$$h\sim\delta^{\max(\lambda/k,\,l_1^{-1})}$$

may be suggested with λ from (6.2). From (7.3) in case $l_1<\infty$ we have also $\xi_h=O(h^{l_1})$, $\xi_{h,*}=O(h^{l_1})$, $\bar\xi_{h,*}=O(h^{l_1})$, in case $l_1=\infty$ we have $\xi_h=0$ (provided $h^k<2^{-1/2}$), $\xi_{h,*}=O(h^{k^2|\ln h|/\ln 2})$, $\bar\xi_{h,*}=O(\exp(-h^{-2k}/e))$. Hence in the a-priori choice of r condition $r\bar\xi_{h,*}\le c$ is fulfilled, if $l_1<\infty$ and $r=O(h^{-l_1})$ or if $l_1=\infty$ and $r=O(\exp(h^{-2k}/e))$; condition $r\xi_{h,1}\le c$ in [10] is fulfilled, if $r=O(h^{-\min(l_1,k)})$. Upper bound $\bar r=\bar\xi_{h,*}^{-1}$ in Rules 1,2,1',2' for the a-posteriori choice of r may be replaced by $\tilde r=ch^{-l_1}$ in case $l_1<\infty$ and by $\tilde r=c\exp(h^{-2k}/e)$ in case $l_1=\infty$.

Note, that l_n, n=1,2,... may be considered real numbers, if (7.1) is replaced by any condition guaranteeing $\mathcal{R}(A^n)\subset H^{l_n}$ for real l_n (e. g. using analogue of (7.1) with fractional differentation).

Acknowledgement. The author thanks Dr. T.Raus, who was drawing the author's attention to Lemma 4.4 in [13] for his kind advice.

References

1. Gfrerer H. An a-posteriori parameter choice for ordinary and iterated Tikhonov regularization of ill-posed problems leading to optimal convergence rates, Math. Comput. 49 (1987), 507-522, S5-S12.
2. Groetsch S.W. , King J.T., Murio D. Asymptotic analysis of a finite element method for Fredholm equations of the first kind. In: C.T.H. Baker, G.F. Miller, eds., Treatment of Integral Equations by Numerical Methods (Academic Press, London, 1982), 1-11.
3. Groetsch C.W. The Theory of Tikhonov Regularization for Fredholm Equations of the first kind (Pitman, Boston, 1984).
4. Hämarik U. Regularized projection methods for ill-posed problems. Proc. Acad. Sciences Estonian SSR 33 (1984), No. 3, 266-276.
5. Hämarik U. On the discretization error in the regularized Ritz-Galerkin method for solving ill-posed problems. In: Proc. of Symposium on modeling, inverse problems and numerical methods, Tallinn, 1991, to appear.
6. Hämarik U. On the discretization error in regularized projection methods with parameter choice by discrepancy principle. In: Proc. of inernat. conf. Ill-posed problems in natural sciences (Moscow, Aug. 19-25, 1991), to appear 1992.
7. King J.T., Neubauer A. A variant of finite-dimensional Tikhonov regularization with a-posteriori parameter choice, Computing 40 (1988), 91-109.
8. Krasnoselskii M. et al, Integral operators in spaces of summable functions (Noordjoff Int. Publ., Leydem, 1976).
9. Neubauer A.An a posteriori parameter choice for Tikhonov Regularization in the presence of modeling error, Applied Numerical Mathematics 4 (1988), 507-519.
10. Plato R., Vainikko G.On the regularization of the Ritz-Galerkin method for solving ill-posed problems, Acta et comment. Univers. Tartuensis, 1989, 863, 3-17.
11. Plato R.,Vainikko G. On the regularization of projection methods for solving ill-posed problems , Numer. Math. 57, No. 1 (1990), 63-79.
12. Raus T. On the discrepancy principle for the solution of ill-posed problems. Acta et comment. Univers. Tartuensis, 1984, 672, 16-26.
13. Raus T. An a-posteriori choice of the regularization parameter in case of approximately given error bound of data. Acta et comment. Univers. Tartuensis, 1990, 913, 73-87.

14. Vainikko G.M., Hämarik U.A. Projection methods and self-regularization in ill-posed problems. Sov. Math. 29, No. 10 (1985), 1-20.
15. Vainikko G.M., Veretennikov Yu.A. Iteration procedures in ill-posed problems (Nauka, Moscow, 1986).

КВАЗИОПТИМАЛЬНАЯ ОЦЕНКА ПОГРЕШНОСТИ ДЛЯ РЕГУЛЯРИЗОВАННОГО МЕТОДА РИТЦА-ГАЛЕРКИНА С АПОСТЕРИОРНЫМ ВЫБОРОМ ПАРАМЕТРА

У.Хямарик

Резюме

В гильбертовом пространстве H рассматривается уравнение $Au=f$, $A=A^*\geq 0$. Вместо $f\in\mathcal{R}(A)$ задан $f_\delta\in H$, $\|f_\delta-f\|\leq\delta$. Пусть P_h $(h>0)$ - ортопроекторы в H, $\|A(I-P_h)\|\to 0$ $(h\to 0)$. Метод Ритца-Галеркина для $Au=f$ имеет вид $A_h u_h=P_h f_\delta$, $A_h=P_h A P_h$. Регуляризованное решение для этого уравнения строится в виде

$$u_{h,r}=(I-A_h g_r(A_h))P_h u_0+g_r(A_h)P_h f_\delta,$$

где $g_r:[0,\|A\|]\to\mathbb{R}$ измеримая по Борелю функция, удовлетворяющая условиям (2.1),(2.2) с $p_0>0$. Эта схема включает метод Лаврентьева ($p_0=1$), итерированный метод Лаврентьева порядка m ($p_0=m$), итерационные методы ($p_0=\infty$) и т.д. В теореме 4.1 указывается правило априорного выбора r, гарантирующее при предположении истокообразности

$$u_*=A^p v,\quad \|v\|\leq\varrho,\quad u_0-u_*=A^p w,\quad \|w\|\leq\varrho,\quad p>0. \qquad (*)$$

с $p\leq p_0$ оценку

$$\|u_{h,r}-u_*\|=\mathcal{O}((\varrho\delta^p)^{1/(p+1)}+\varrho\|(I-P_h)A^p\|). \qquad (**)$$

имеющую оптимальный порядок как в отношении погрешности правой части уравнения, так и в отношении погрешности дискретизации. Для апостериорного выбора r рассматриваются правила $\|A_h u_{h,r}-P_h f_\delta\|=b\delta$ (предполагая $p_0>1$) и $\|B_{h,r}(A_h u_{h,r}-P_h f_\delta)\|=b\delta$ $(b>1$, $B_{h,r}$ определен в (3.1)) с априорным ограничением $r\leq\sup\limits_{\alpha>0}[(2\alpha)^{1/(2\alpha)}\|(I-P_h)A^\alpha\|^{-1/\alpha}]$. В теореме 5.1 доказано, что если выполнено (*), то эти правила обеспечивают оценку (**) с $p\leq p_0-1$ и $p\leq p_0$ соответственно.

Acta et commentationes universitatis Tartuensis, 937, 1992 , 77-89.

ABOUT REGULARIZATION PARAMETER CHOICE IN CASE OF APPROXIMATELY GIVEN ERROR BOUNDS OF DATA

Toomas Raus

In this paper we consider the ill-posed problem $A_0u=f_0$ with an approximately given operator and the right-hand side . We suppose that the error bounds are unknown and only some guess about the errors is available. Under this condition we propose an a-posteriori parameter choice for a class of regularization methods and give the theorem of convergence. The error estimation of the approximate solution is deduced only in the case of properly determined error bounds.

1. Introduction . We consider the equation

$$A_0u = f_0, \quad f_0 \in R(A_0), \tag{1.1}$$

where u and f_0 are elements of real Hilbert spaces H and F, respectively, and A_0 is a continuous linear operator. We assume that range $R(A_0)$ is non-closed and so our problem is ill-posed. The kernel $N(A_0)$ may be non-trivial. We also suppose that instead of f_0 and A_0 we have at our disposal only their approximations $f \in F$ and $A \in \mathcal{L}(H,F)$. At that the error bounds are given approximately. It means that the supposed error levels $\delta>0$ and $\eta>0$ are given, but we do not know exactly if $\|f_0 - f\| \le \delta$ and $\|A_0 - A\| \le \eta$ or not.

To determine on approximation to the solution, it is necessary to regularize the equation (1.1). A more favourable way (see [1-3,9,10]) is the use of regularization methods which are generated by Borel measurable functions

$$g_r : [0,a] \to \mathbb{R},$$

$r \ge 0$, $\|A\|^2 \le a$. We assume that the functions g_r satisfy the following conditions:

$$\sup_{0 \le \lambda \le a} \sqrt{\lambda}\, |g_r(\lambda)| \le \gamma_* \sqrt{r} \ , \ r \ge 0, \tag{1.2}$$

$$\sup_{0 \le \lambda \le a} \lambda^p |1 - \lambda g_r(\lambda)| \le \gamma_p r^{-p} \ , \ r > 0, \ 0 \le p \le p_0, \tag{1.3}$$

where p_0, γ_* and γ_p are positive constants. Note that the greatest value p_0, for which the inequality (1.3) holds, is named a qualification of method.

Let $u_0 \in H$ be an initial approximation. Then an approximation to the solution u_* of equation (1.1), which is the nearest to u_0, is given by the formula

$$u_r = (I - A^*A g_r(A^*A))u_0 + g_r(A^*A)A^*f, \qquad (1.4)$$

where I is the unique operator and r - regularization parameter. In finding the approximate solution a question arises how to choose the parameter r. In the case of the known error bounds (i.e., if $\|f - f_0\| \le \delta$, $\|A - A_0\| \le \eta$ and, hence, $\|f - Au_*\| \le \delta + \eta\|u_*\|$) a-posteriori parameter choice by discrepancy principle or its modifications (see e.g. [2,4-7,9, 10]) is well known and effective. But if $\|f - Au_*\| > b_1(\delta + \eta\|u_*\|)$ ($b_1 > 1$ - a constant used in the discrepancy principle; usually $b_1 \in [1, 2]$), then the choice of the parameter r by a discrepancy principle is unstable in this sence that the error of approximate solution may be arbitrarily great independent of the value of relation $\|f - Au_*\| / (\delta + \eta\|u_*\|)$. Therefore, such a choice of the parameter in the case of approximately given error bounds is not reasonable. The aim of our paper is to present some rules for the stable parameter choice which guarantees the convergence of the approximate solution to the exact solution if only relation $\|f - Au_*\| / (\delta + \eta\|u_*\|)$ is bounded in the process $\delta \to 0$, $\eta \to 0$. Note that the case of exactly given and self-adjoint operator is considered in [8].

We denote

$$\beta_r(\lambda) := \begin{cases} 1, & \text{if } p_0 = \infty, \\ (1 - \lambda g_r(\lambda))^{1/p_0}, & \text{if } p_0 < \infty, \end{cases}$$

and suppose in this paper that in addition to inequalities (1.2) and (1.3) the functions g_r satisfy the conditions 1) – 3):

1) for $0 \le \lambda \le a$, $0 \le r_2 \le r_1$

$$0 \le 1 - \lambda g_{r_1}(\lambda) \le 1 - \lambda g_{r_2}(\lambda) \le 1; \qquad (1.5)$$

2) the function $g_r(\lambda)$ is continuously differentiable and for $0 \le \lambda \le a$, $r > 0$

$$\frac{\partial(g_s(\lambda))}{\partial s}\Big|_{s=r} \le \bar{\gamma}\beta_r(\lambda)(1 - \lambda g_r(\lambda)), \quad \bar{\gamma} = \text{const}; \qquad (1.6)$$

3) there exists a constant R_0, so that, for $r \ge R_0$ and ε, $0 < \varepsilon < \bar{\gamma}_{1/2}$, we find a constant $W(\varepsilon)$ such that

$$\sqrt{r\lambda}\,\beta_r(\lambda)(1 - \lambda g_r(\lambda)) \le \varepsilon, \qquad (1.7)$$

if $\lambda \ge W(\varepsilon)/r$.

Some examples of regularization methods, for which (1.2), (1.3) and the conditions 1) - 3) are satisfied, are presented in Section 2. In Section 3 and 4 we propose four rules for a-posteriori parameter choice and prove the convergence theorem. The error estimation of the approximate solution is deduced under the condition that our guess about errrors $\|f - f_0\|$ and $\|A - A_0\|$ is true.

2. Examples of methods.

1. The Tikhonov method.

$$u_r = (r^{-1}I + A^*A)^{-1}A^*f.$$

The method is of the form (1.4) with $u_0 = 0$ and $g_r(\lambda) = (\lambda + r^{-1})^{-1}$. Conditions (1.2), (1.3), (1.6) and (1.7) hold with $p_0 = 1$, $\gamma_* = 1/2$, $\gamma_p = p^p(1-p)^{1-p}$, $\bar{\gamma} = 1$, $R_0 = 0$, $W(\varepsilon) = \varepsilon^{-3/2}$.

2. The iterative variant of the Tikhonov method. Let $m \in \mathbb{N}$, $m \geq 1$. We determine consequtively the solution $u_{1,r}, u_{2,r}, \ldots, u_{m,r}$ of the equations

$$r^{-1}u_{n,r} + A^*A\,u_{n,r} = r^{-1}u_{n-1,r} + A^*f, \quad n = 1,2,\ldots,m.$$

As the approximation u_r to the solution u_* we take an element $u_{m,r}$. The method is of the form (1.4) with $g_r(\lambda) = (1-(1+r\lambda)^{-m})/\lambda$. Conditions (1.2), (1.3), (1.6) and (1.7) hold with $p_0 = m$, $\gamma_* = \sqrt{m}$, $\gamma_p = (p/m)^p(1-p/m)^{m-p}$, $\bar{\gamma} = m$, $R_0 = 0$, $W(\varepsilon) = \varepsilon^{-2/(2m+1)}$.

3. The method of successive approximation (explicit scheme). Suppose that $0 < \mu < 1/\|A^*A\|$ and compute consequtively

$$u_r = u_{r-1} - \mu(A^*Au_{r-1} - A^*f), \quad r = 1,2,\ldots.$$

The method is of the form (1.4) with $g_r(\lambda) = (1-(1-\mu\lambda)^r)/\lambda$. Conditions (1.2), (1.3), (1.6) and (1.7) hold with $p_0 = \infty$, $\gamma_* = \sqrt{\mu}$, $\gamma_p = (p/\mu e)^p$, $\bar{\gamma} = \mu/(1-\mu\|A^*A\|)$, $R_0 = 0$ and $W(\varepsilon) = x_0$, where x_0 is the greatest solution of the equation $\sqrt{x}e^{-\mu x} = \varepsilon$.

4. Implicit scheme. Let $\alpha > 0$ be a constant. We determine consequtively u_r as the solutions of the equations

$$\alpha u_r + A^*Au_r = \alpha u_{r-1} + A^*f, \quad r = 1,2,\ldots.$$

The method is of the form (1.4) with $g_r(\lambda) = (1-(\alpha/(\alpha+\lambda))^r)/\lambda$. Conditions (1.2), (1:3), (1.6) and (1.7) hold with $p_0 = \infty$, $\gamma_* = b/\sqrt{\alpha}$ ($b_0 = 0{,}6382$), $\gamma_p = (\alpha p)^p$, $\bar{\gamma} = 1/\alpha$, $R_0 = 3/2$ and $W(\varepsilon) = x_0$, where x_0 is the greatest solution of the equation $\sqrt{x}(1+2x/3\alpha)^{-3/2} = \varepsilon$.

5. The method of the Cauchy problem (a continuous analog of

iterative methods). We take an approximation u_r to the solution (1.1) the solution of the Cauchy problem

$$u'(r) + A^*A\,u(r) = A^*f, \quad u(0) = u_0$$

The method is of the form (1.4) with $g_r(\lambda) = (1 - e^{-r\lambda})/\lambda$. Conditions (1.2), (1.3), (1.6) and (1.7) hold with $p_0 = \infty$, $\gamma_* = b_0$, $\gamma_p = (p/e)^p$, $\overline{\gamma} = 1$, $R_0 = 0$ and $W(\varepsilon) = x_0$, where the x_0 is the greatest solution of the equation $\sqrt{x}\, e^{-x} = \varepsilon$.

3. A posteriori parameter choice. Let us define operator B_r depending on the qualification of method p_0, as follows:

$$B_r = \int_0^{\|AA^*\|} \beta_r(\lambda)\, dQ(\lambda) = \begin{cases} I, & \text{if } p_0 = \infty, \\ (I - AA^* g_r(AA^*))^{1/p_0}, & \text{if } p_0 < \infty, \end{cases}$$

where $Q(\lambda)$ is the spectral family of the projectors of operator AA^*. Note that for methods 1 and 2 $B_r = (I + rAA^*)^{-1}$ and for methods 3-5 $B_r = I$. Denote

$$\varphi(r) = \sqrt{r}\, \|A^* B_r (Au_r - f)\|,$$

$$\tilde{\gamma}_p := \begin{cases} \gamma_p, & \text{if } p_0 = \infty, \\ (\gamma_{pp_0/(1+p_0)})^{(p_0+1)/p_0}, & \text{if } p_0 < \infty \end{cases}$$

and consider the following rules for the choice of the regularization parameter r.

Rule 1. Let b_2 and b_1 be the constants such that $b_2 \ge b_1 > \tilde{\gamma}_{1/2}$. If

$$\varphi(1) \le b_2(\delta + \eta\, \|u_*\|), \tag{3.1}$$

then choose $r(\delta,\eta) = 1$. In the contrary case choose $r = r(\delta,\eta) > 1$ such that

$$\varphi(r(\delta,\eta)) \le b_2(\delta + \eta\, \|u_*\|), \tag{3.2}$$

$$\varphi(r) \ge b_1(\delta + \eta\, \|u_*\|) \quad \forall r \in [1, r(\delta,\eta)]. \tag{3.3}$$

The next parameter selection rule may be applied to iteration procedures.

Rule 2. Let $b > \tilde{\gamma}_{1/2}$. We choose the least integer parameter $r = r(\delta,\eta) \ge 1$ such that

$$\varphi(r(\delta,\eta)) \le b(\delta + \eta\, \|u_*\|). \tag{3.4}$$

The application of Rules 1 and 2 is difficult as we must know the estimation of norm $\|u_*\|$. Therefore we present some modifications of these Rules where quantity $\|u_*\|$ is substituted by $\|u_r - u_0\| + \|u_0\|$.

Rule 3 is the same as Rule 1 with substitution (3.1)–(3.3) by the conditions

$$\varphi(1) \le b_2(\delta + \eta\, \|u_0\|),$$

$$\varphi(r(\delta,\eta)) \le b_2(\delta+\eta(\|u_{r(\delta,\eta)}-u_0\|+\|u_0\|)), \tag{3.5}$$

$$\varphi(r) \ge b_1(\delta+\eta(\|u_r-u_0\|+\|u_0\|)) \quad \forall r\in[1,r(\delta,\eta)]. \tag{3.6}$$

Rule 4 is the same as Rule 2 with substitution (3.4) by the condition

$$\varphi(r(\delta,\eta)) \le b(\delta+\eta(\|u_{r(\delta,\eta)}-u_0\|+\|u_0\|)). \tag{3.7}$$

To show that our parameter selection rules are practicable, we shall prove some properties of function $\varphi(r)$. First we define

$$K_r = I-AA^*g_r(AA^*), \qquad \tilde{K}_r = I-A^*Ag_r(A^*A),$$

$$\tilde{B}_r = \begin{cases} I, & \text{if } p_0=\infty, \\ (I-A^*Ag_r(A^*A))^{1/p_0}, & \text{if } p_0<\infty. \end{cases}$$

Lemma 3.1. Suppose that the condition (1.3) holds for functions g_r. Then for each $f\in F$ and $A\in\mathcal{L}(H,F)$ we have

$$\lim_{r\to\infty}\varphi(r) = 0 \tag{3.8}$$

Proof. Let Q be an orthogonal projection in F onto $\overline{R(A)}$. Applying the equalities

$$A^*g_r(AA^*) = g_r(A^*A)A^*, \quad Ag_r(A^*A) = g_r(AA^*)A, \tag{3.9}$$

it is easy to show that

$$Au_r-f = K_r(Au_0-f), \tag{3.10}$$

$$\varphi(r) = \sqrt{r}\,\|A^*B_rK_rQ(Au_0-f)\|. \tag{3.11}$$

If the condition (1.3) holds for functions g_r, then using the polar expansion of the operator A^* ($A^*=U^*(AA^*)^{1/2}$, $\|U^*\|=1$) we obtain

$$\sqrt{r}\,\|A^*B_rK_r\| \le \sqrt{r}\max_{0\le\lambda\le a}\sqrt{\lambda}\,(1-\lambda g_r(\lambda))^{1+1/p_0} = \tag{3.12}$$

$$= \sqrt{r}\max_{0\le\lambda\le a}\left[\lambda^{(p_0+1)/2p_0}(1-\lambda g_r(\lambda)\right]^{1+1/p_0} \le \tilde{\gamma}_{1/2} = \text{const}$$

and for elements $Q(Au_0-f) = Av \in R(A)$, $v\in H$,

$$\sqrt{r}\,\|A^*B_rK_rQ(Au_0-f)\| \le \sqrt{r}\|A^*B_rK_rA\|\cdot\|v\| \le \tag{3.13}$$

$$\le \tilde{\gamma}_1\|v\|/\sqrt{r} \to 0, \quad r\to\infty.$$

Using now the Banach-Steinhaus theorem, we get with regard (3.11) the convergence (3.8). ■

Lemma 3.2. Let $\|f-Au_*\| \le \delta+\eta\|u_*\|$, $\|u_0-u_*\|\le M$ and $b>\tilde{\gamma}_{1/2}$. Suppose that the condition (1.3) holds for functions g_r. Then for each r, $r\ge R_{M,\delta,\eta}=(\tilde{\gamma}_1M/((b-1)\,\tilde{\gamma}_{1/2}(\delta+\eta\|u_*\|)))^2$, holds

$$\varphi(r) \le b(\delta+\eta\|u_*\|). \tag{3.14}$$

Proof. Using (3.10) we have

$$A^*B_r(Au_r-f) = A^*B_rK_r(Au_0-f) = A^*B_rK_rA(u_0-u_*) + A^*B_rK_r(Au_*-f). \quad (3.15)$$

As $\|A^*B_rK_rA\| \le \tilde{\gamma}_1/r$, $\|A^*B_rK_r\| \le \tilde{\gamma}_{1/2}/\sqrt{r}$ (see (3.12),(3.13)), then

$$\varphi(r) \le \tilde{\gamma}_1 M/\sqrt{r} + \tilde{\gamma}_{1/2}(\delta+\eta\|u_*\|),$$

from which after non-complicated calculations follows the assertion of the lemma. ■

From condition 2) it follows that the function $r \to \varphi(r)$ is continuous and then Lemma 3.1 yields that the choice of parameter $r(\delta,\eta)$ according to Rule 1–Rule 4 is possible. If we know a constant $M>0$ such that $\|u_0-u_*\| \le M$, then it is sufficient to search the parameter $r(\delta,\eta)$ in the finite interval $[1, R_{M,\delta,\eta}]$. If $\varphi(r) > b_2\delta$ for $r \in [1, R_{M,\delta,\eta}]$, then from Lemma 3.2 it follows that $\|f-Au_*\| \ge \delta+\eta\|u_*\|$ and, consequently, instead of $q = \delta+\eta\|u_*\|$ it is necessary to take in Rules 1–4 some $q' > q$. Note that the function $\varphi(r)$ is non-monotone and therefore in Rule 1 we must use the conditions (3.2) and (3.3) instead of inequalities $b_1(\delta+\eta\|u_*\|) \le \varphi(r) \le b_2(\delta+\eta\|u_*\|)$.

4. Convergence and error estimation of the approximate solution

Theorem 4.1. Let $A_0, A \in \mathcal{L}(H,F)$, $\|A\|^2 \le a$, $f_0 \in R(A_0)$ and $u_* \ne 0$ in the case of Rule 3 and 4. Suppose that conditions (1.2), (1.3), (1.5) and (1.7) hold for functions g_r and the parameter $r = r(\delta,\eta)$ is chosen according to one of Rules 1-4. If in the process $\delta \to 0$, $\eta \to 0$

$$\|f-Au_*\| \le C(\delta+\eta\|u_*\|), \quad \|A-A_0\| \le C_1\eta, \quad C, C_1 = \text{const},$$

then

$$\|u_{r(\delta,\eta)} - u_*\| \to 0 \quad (\delta \to 0, \eta \to 0). \qquad (4.1)$$

To prove the theorem, we need the following lemmas.

Lemma 4.2. [10, p.99]. Let $A, A_0 \in \mathcal{L}(H,F)$, $\|A_0-A\| \le \eta$, $\|A\|^2 \le a$ and for functions g_r holds (1.3). Then for each $v \in N(A_0)^\perp$

$$\|\tilde{K}_r v\| \to 0 \quad (r \to \infty, \eta \to 0).$$

Lemma 4.3. [10, p.109]. Let $A, A_0 \in \mathcal{L}(H,F)$, $\|A-A_0\| \le \eta$, $\|A\|^2 \le a$ and for functions g_r holds (1.3). If for $v \in N(A_0)^\perp$, $r_n \le \bar{r} = \text{const}$, we have

$$\|A^*B_{r_n}K_{r_n}Av\| \to 0 \quad (n \to \infty, \eta \to 0),$$

then

$$\|\tilde{K}_{r_n} v\| \to 0 \quad (n \to \infty, \eta \to 0).$$

Proof of Theorem 4.1. 1. At first we give some auxiliary results. We have

$$u_r - u_* = K_r(u_0-u_*) + g_r(A^*A)A^*(f-Au_*), \qquad (4.2)$$

$$A^*B_r(Au_r - f) = A^*B_rK_rA(u_0 - u_*) + A^*B_rK_r(Au_* - f) \tag{4.3}$$

From (1.2) and (1.3) we obtain

$$\|g_r(A^*A)A^*(f - Au_*)\| \le \gamma_* C\sqrt{r}(\delta + \eta\|u_*\|) \tag{4.4}$$

$$\|A^*B_rK_r\| \le \tilde{\gamma}_{1/2}/\sqrt{r}. \tag{4.5}$$

From (4.2) and (4.4) we get

$$\|u_{r(\delta,\eta)} - u_*\| \le \|\tilde{K}_{r(\delta,\eta)}(u_0 - u_*)\| + \gamma_*\sqrt{r(\delta,\eta)}\, C(\delta + \eta\|u_*\|). \tag{4.6}$$

To prove the theorem, it suffices to show the convergence of the right-hand side of (4.6).

2. First we consider the case if the parameter $r(\delta,\eta)$ is chosen according to Rule 1 or Rule 2 and show the convergence of the first term of the right-hand side of (4.6). From (3.2)-(3.4), (4.3) and (4.5) it follows that

$$\sqrt{r(\delta,\eta)}\,\|A^*B_{r(\delta,\eta)}K_{r(\delta,\eta)}A(u_0 - u_*)\| \le \sqrt{r(\delta,\eta)}\,\|A^*B_{r(\delta,\eta)}(Au_{r(\delta,\eta)} - f)\| + \tag{4.7}$$

$$+ \sqrt{r(\delta,\eta)}\,\|A^*B_{r(\delta,\eta)}K_{r(\delta,\eta)}(f - Au_*)\| \le (b_2 + \tilde{\gamma}_{1/2}C)(\delta + \eta\|u_*\|),$$

where $b_2 = b$ in case of Rule 2. As $r(\delta,\eta) \ge 1$ then from (4.7) we get

$$\|A^*B_{r(\delta,\eta)}K_{r(\delta,\eta)}A(u_0 - u_*)\| \to 0 \quad (\delta \to 0,\ \eta \to 0).$$

If at that $r(\delta,\eta) \to \infty$, then the convergence

$$\|\tilde{K}_{r(\delta,\eta)}(u_0 - u_*)\| \to 0 \qquad (\delta \to 0,\ \eta \to 0)$$

follows from Lemma 4.2. If $r(\delta,\eta) \le \bar{r} = \text{const}$ $(\delta \to 0,\ \eta \to 0)$, then the convergence follows from Lemma 4.3.

3. To show the convergence of the second term of the right-hand side of (4.6), we prove at first the convergence

$$r_{\delta,\eta}(\delta + \eta) \to 0 \quad (\delta \to 0, \eta \to 0), \tag{4.8}$$

where $r_{\delta,\eta}$ is the greatest parameter, for which

$$\sqrt{r_{\delta,\eta}}\,\|A^*B_{r_{\delta,\eta}}K_{r_{\delta,\eta}}A(u_0 - u_*)\| = (b_1 - \tilde{\gamma}_{1/2})(\delta + \eta\|u_*\|) \tag{4.9}$$

(in case of Rule 2 $b_1 = b$). If $r_{\delta,\eta} \le \bar{r} = \text{const}$ in the process $\delta \to 0$, $\eta \to 0$, then (4.8) is obvious. If $r_{\delta,\eta} \to \infty$ $(\delta \to 0,\ \eta \to 0)$, then applying the Banach-Steinhaus theorem, we can prove similarly as in Lemma 3.1 that

$$r_{\delta,\eta}\,\|A^*B_{r_{\delta,\eta}}K_{r_{\delta,\eta}}A(u_0 - u_*)\| \to 0 \quad (r_{\delta,\eta} \to \infty) \tag{4.10}$$

and then the convergence (4.8) follows from (4.9) and (4.10).

To complete the proof, we shall show that

$$r(\delta,\eta) - q \le d_C \cdot r_0 \tag{4.11}$$

where $r_0 = \max\{r_{\delta,\eta}, 1, R_0\}$, $d_C = (W_0/\varepsilon_0)^m = \text{const}$, $m = \text{entier}(3C^2 + 1)$,

$\varepsilon_0 = (\tilde{\gamma}_{1/2}/(\sqrt{3}\,C))^2$, $W_0 = \max\{W(\sqrt{\varepsilon_0}), \varepsilon_0 + 1\}$.

Here the constants $W(\varepsilon)$ and R_0 are defined by condition 3), $q=0$ in case of Rule 1 and $q=1$ in case of Rule 2. As $d_C > 1$, then in case of $r(\delta,\eta) - q \le r_0$ the inequality (4.11) holds. In the following we consider the case $r(\delta,\eta) - q > r_0$.

If the parameter $r = r(\delta,\eta)$ is chosen according to Rule 1 or 2, then

$$\sqrt{r}\,\|A^* B_r K_r (Au_r - f)\| \ge b_1(\delta + \eta\|u_*\|) \quad 1 \le r \le r(\delta,\eta) - q \qquad (4.12)$$

As $r_{\delta,\eta}$ is the greatest parameter, for which (4.9) holds, then we have for r, $r \ge r_0 \ge r_{\delta,\eta}$, that

$$\sqrt{r}\,\|A^* B_r K_r A(u_0 - u_*)\| \le (b_1 - \tilde{\gamma}_{1/2})(\delta + \eta\|u_*\|) \qquad (4.13)$$

Now applying (4.3), (4.12) and (4.13) we obtain for r, $r_0 \le r < r(\delta,\eta) - q$, that

$$\begin{aligned} \sqrt{r}\,\|A^* B_r K_r (f - Au_*)\| &\ge \sqrt{r}\,\|A^* B_r K_r (Au_r - f)\| - \\ &- \sqrt{r}\,\|A^* B_r K_r A(u_0 - u_*)\| \ge \tilde{\gamma}_{1/2}(\delta + \eta\|u_*\|). \end{aligned} \qquad (4.14)$$

Further we give the upper bound for $\sqrt{r}\,\|A^* B_r K_r (f - Au_*)\|$. From (1.3), (1.5) and (1.7) we get that

$$T(r,\lambda) := \sqrt{r\lambda}\,\beta_r(\lambda)(1 - \lambda g_r(\lambda)) \le \sqrt{\varepsilon_0} \quad (\lambda \le \varepsilon_0/r)$$

$$T(r,\lambda) \le \sqrt{\varepsilon_0} \quad (\lambda \ge W_0/r), \quad \sup_{0 \le \lambda < a} T(r,\lambda) \le \tilde{\gamma}_{1/2}$$

Then we can estimate for $r \ge r_0$

$$\begin{aligned} r\,\|A^* B_r K_r (f - Au_*)\|^2 &\le 2\varepsilon_0 \|f - Au_*\|^2 + \\ &\int_{\varepsilon_0/r}^{W_0/r} T^2(r,\lambda)\, d\langle Q(\lambda)(f - Au_*), f - Au_*\rangle \le \end{aligned} \qquad (4.15)$$

$$2\tilde{\gamma}_{1/2}^2 (\delta + \eta\|u_*\|)^2/3 + \tilde{\gamma}_{1/2}^2 \|(Q(W_0/r) - Q(\varepsilon_0/r))(f - Au_*)\|^2,$$

where $Q(\lambda)$ is the spectral family of the projectors of operator AA^*.

Now from (4.14) and (4.15) it follows that

$$\|(Q(W_0/r) - Q(\varepsilon_0/r))(f - Au_*)\| \ge (\delta + \eta\|u_*\|)/\sqrt{3}, \quad r_0 \le r < r(\delta,\eta) - q. \qquad (4.16)$$

Let $r_j := r_0 (W_0/\varepsilon_0)^j$, $j = 0,1,2,\ldots,m$. Then $\varepsilon_0/r_j = W_0/r_{j+1}$, $j = 0,1,\ldots,m-1$ and, hence

$$(Q(W_0/r_j)F \setminus Q(\varepsilon_0/r_j)F) \cap (Q(W_0/r_i)F \setminus Q(\varepsilon_0/r_i)F) = \emptyset, \quad i \ne j, \qquad (4.17)$$

where $B_1 \setminus B_2$ denotes difference of sets B_1 and B_2.

If we now suppose in contradiction to (4.11) that $r(\delta,\eta) - q > d_C r_0$, then with regard (4.16), (4.17) we have

$$\|f - Au_*\|^2 \ge \sum_{j=0}^{m} \|(Q(W_0/r_j) - Q(\varepsilon_0/r_j))(f - Au_*)\|^2 \ge (m+1)(\delta + \eta\|u_*\|)^2/3 =$$

$$= (1+\text{entier}(3C^2+1))(\delta+\eta\|u_*\|)^2/3 > C^2(\delta+\eta\|u_*\|)^2,$$

what contradicts to the assumption of the theorem. Hence (4.11) holds, which together with (4.8) proves the convergence of the second term of the right-hand side of (4.6).

4. Finally we consider the case if the parameter $r(\delta,\eta)$ is chosen according to Rule 3 or Rule 4. Let r_0 be the parameter, which is chosen according to Rule 1 or Rule 2, respectively.

If $r(\delta,\eta)\le r_0$, then the convergence of the second term of right-hand side of (4.6) is obvious. To prove the convergence $\|\tilde{K}_{r(\delta,\eta)}(u_0-u_*)\|\to 0$ $(\delta\to 0,\ \eta\to 0)$, it suffices to show (see Part 2) that

$$\sqrt{r(\delta,\eta)}\ \|A^*B_{r(\delta,\eta)}(Au_{r(\delta,\eta)}-f)\| \le b_2'(\delta+\eta\|u_*\|),\ b_2'=\text{const}. \quad (4.18)$$

Really, applying the equality $u_r-u_0=g_r(A^*A)A^*(f-Au_0)$ and (1.5) we get

$$\|u_{r(\delta,\eta)}-u_0\|+\|u_0\| \le \|u_{r_0}-u_0\|+\|u_0\| \le$$

$$\le \|u_{r_0}-u_*\|+\|u_*-u_0\|+\|u_0\| \le c\|u_*\|;\quad c=\text{const},$$

which together with (3.5), (3.7) gives (4.18), where $b_2'=b_2c$.

If $r(\delta,\eta)>r_0$, then from (1.5) follows

$$\|\tilde{K}_{r(\delta,\eta)}(u_0-u_*)\| \le \|\tilde{K}_{r_0}(u_0-u_*)\|\to 0 \quad (\delta\to 0,\ \eta\to 0).$$

To prove the convergence of the second term of right-hand side of (4.6), it suffices to show (see Part 3) that for sufficiently smalls δ and η

$$\sqrt{r}\ \|A^*B_r(A_r-f)\| \ge b_1'(\delta+\eta\|u_*\|), \quad (4.19)$$

$$1\le r\le r(\delta,\eta)-q,\ b_1'=\text{const}>\tilde{\gamma}_{1/2}.$$

Really, if $1\le r\le r_0-q$, then (4.19) holds with $b_1'=b_1$ (see Rules 1 and 2). If $r_0-q\le r\le r(\delta,\eta)-q$, then

$$\|u_r-u_0\|+\|u_0\| \ge \|u_{r_0-q}-u_*\|+\|u_0\| \ge \quad (4.20)$$

$$\ge \|u_*-u_0\|+\|u_0\|-\|u_{r_0-q}-u_*\| \ge c\|u_*\|,\ c>\tilde{\gamma}_{1/2}/b_1,$$

if δ and η are sufficiently small. Now (4.19) follows from (3.6), (3.7) and (4.20) with $b_1'=cb_1>\tilde{\gamma}_{1/2}$. ■

In the next theorem we give the error estimation of the approximate solution in case, if our supposition about errors $\|f-Au_*\|$ and $\|A-A_0\|$ is found to be true.

Theorem 4.4. Let $A,A_0\in\mathcal{L}(H,F)$, $\|A\|^2\le a$, $f_0\in R(A_0)$, $\|f-Au_*\|\le \delta+\eta\|u_*\|$, $\|A-A_0\|\le\eta$ and $u_*\ne 0$ in case of Rules 3 and 4. Suppose that conditions (1.2), (1.3), (1.5), (1.6) hold for functions g_r and the

parameter $r = r(\delta,\eta)$ is chosen according to one of Rules 1-4. If $u_0 - u_* = |A_0|^p v$, $\|v\| \le \varrho$, $0 \le p \le 2p_0$ ($|A_0| = (A_0^* A_0)^{1/2}$), then

$$\|u_{r(\delta,\eta)} - u_*\| \le d_p \Big[((b_* + \tilde\gamma_{1/2})^{\frac{p}{p+1}} + (b_1' - \tilde\gamma_{1/2})^{-\frac{1}{p+1}}) \varrho^{\frac{1}{p+1}} (\delta + \eta\|u_*\|)^{\frac{p}{p+1}} +$$

$$+ (b_1' - \tilde\gamma_{1/2})^{-1} \varrho\, |\ln\eta|\, \eta^{\min\{1,p\}} \Big], \quad d_p = \text{const},$$

where

$$b_* = \begin{cases} b_2 b', & \text{if } r(\delta,\eta) \ge R(\delta,\eta), \\ \max\limits_{r(\delta,\eta) \le r \le R(\delta,\eta)} \varphi(r)/\tau_r(\delta,\eta), & \text{if } r(\delta,\eta) < R(\delta,\eta) \end{cases} \tag{4.21}$$

and $R(\delta,\eta)$ is the greatest parameter, for which $\varphi(r) = b_2 \tau_r(\delta,\eta)$. Here $\tau_r(\delta,\eta) = \delta + \eta\|u_*\|$, $b' = 1$, $b_1' = b_1$ in case of Rules 1, 2 and $b_1 = b_2 = b$, $\tau_r(\delta,\eta) = \delta + \eta(\|u_r - u_0\| + \|u_0\|)$, $b_1' \to b_1$, $b' \to (\|u_* - u_0\| + \|u_0\|)/\|u_*\|$ $(\delta,\eta \to 0)$ in case of Rules 3,4.

We see that in case of known error bounds the parameter choice by Rules 1-4 leads to optimal convergence rates with respect to data errors δ and η. Note that the numerical examples show that in most cases $r(\delta,\eta) \ge R(\delta,\eta)$, $b_* = b_2 b'$ and the choice of the parameter by Rules 1-4 give approximately the same result as the choice by discrepancy principle (see also [8]).

To prove Theorem 4.4, we need the following lemmas.

Lemma 4.5. [10, p.93]. Let $\|A - A_0\| \le \eta$. Then

$$\| |A|^p - |A_0|^p \| \le c_p' (1 + |\ln\eta|)\, \eta^{\min\{1,p\}}, \quad c_p' = \text{const}.$$

Lemma 4.6. Let $c > 0$, $v \in H$, $A \in \mathcal{L}(H,F)$, $\|A\|^2 \le a$ and the condition (1.6) holds for functions g_r. If the parameter r_c is such that

$$\sqrt{r}\, \|A^* B_r K_r A v\| \le c \quad \forall r \ge r_c, \tag{4.22}$$

then

$$r_c \ge r_*,$$

where r_* is the parameter, for which the function

$$\psi(r) := \|\tilde K_r v\| + 2\sqrt{\bar c}\sqrt{r}, \quad \bar c > c,$$

has a global minimum in the interval $[0,\infty)$.

Proof. If $r_* = 0$, then the assertion of the lemma is obvious. If $r_* \ne 0$, then

$$\psi'(r_*) = 0{,}5\, \|\tilde K_{r_*} v\|^{-1} \cdot \frac{\partial}{\partial r} \left(\|\tilde K_r v\|^2 \right)\Big|_{r_* = r} + \sqrt{\bar c}/\sqrt{r_*} = 0. \tag{4.23}$$

Using the condition (1.6) and equality

$$\frac{\partial((1 - \lambda g_r(\lambda))^2)}{\partial r} = -2\lambda(1 - \lambda g_r(\lambda)) \frac{\partial g_r(\lambda)}{\partial r},$$

we obtain

$$-\frac{\partial}{\partial r}(\|\tilde{K}_r v\|^2) = -\frac{\partial}{\partial r}\left(\int_0^{\|A^*A\|}(1-\lambda g_r(\lambda))^2\, d\langle P(\lambda)v,v\rangle\right) =$$

$$= \int_0^{\|A^*A\|} 2\lambda(1-\lambda g_r(\lambda))\frac{\partial g_r(\lambda)}{\partial r}\, d\langle P(\lambda)v,v\rangle \le \qquad (4.24)$$

$$\le 2\bar{\gamma}\int_0^{\|A^*A\|}\lambda\beta_r(\lambda)(1-\lambda g_r(\lambda))^2 d\langle P(\lambda)v,v\rangle = 2\bar{\gamma}\,\|(A^*A)^{1/2}\tilde{B}_r^{1/2}\tilde{K}_r v\|^2.$$

Applying the inequality of the moments $(\|B^p v\| \le \|B^q v\|^{p/q}\,\|v\|^{1-p/q}$, $0<p\le q)$, we have

$$\|(A^*A)^{1/2}\,\tilde{B}_r^{1/2}\,\tilde{K}_r v\| \le \|A^*A\,\tilde{B}_r\tilde{K}_r v\|^{1/2}\,\|\tilde{K}_r v\|^{1/2}. \qquad (4.25)$$

Now using (4.23)–(4.25) and (3.9), we get

$$\bar{c} = -\sqrt{r_*}\,(2\bar{\gamma}\,\|\tilde{K}_{r_*}v\|)^{-1}\cdot\frac{\partial}{\partial r}\left(\|\tilde{K}_r v\|^2\right)\Big|_{r_*=r} \le$$

$$\le \sqrt{r_*}\,\|(A^*A\,\tilde{B}_{r_*})^{1/2}\,\tilde{K}_{r_*}v\|^2\cdot\|\tilde{K}_{r_*}v\|^{-1} \le$$

$$\le \sqrt{r_*}\,\|A^*A\,\tilde{B}_{r_*}\tilde{K}_{r_*}v\| = \sqrt{r_*}\,\|A^*\,\tilde{B}_{r_*}\tilde{K}_{r_*}Av\|,$$

from which together with (4.22) follows the assertion of the lemma. ■

Proof of Theorem 4.4. 1. Similarly as in the proof of Theorem 4.1 we have

$$\|u_{r(\delta,\eta)} - u_*\| \le \|\tilde{K}_{r(\delta,\eta)}(u_0-u_*)\| + \gamma_*\sqrt{r(\delta,\eta)}\,(\delta+\eta\|u_*\|) \qquad (4.26)$$

$$\sqrt{r}\,\|A^*B_rK_r(f-f_0)\| \le \tilde{\gamma}_{1/2}(\delta+\eta\|u_*\|) \qquad (4.27)$$

$$A^*B_r(Au_r - f) = A^*B_rK_rA(u_0-u_*) + A^*B_rK_r(Au_*-f). \qquad (4.28)$$

If the parameter $r(\delta,\eta)$ is chosen according to Rule 1 or Rule 2, then using (3.1)-(3.4),(4.21),(4.27) and (4.28), we get for $r \ge r(\delta,\eta)$

$$\sqrt{r}\,\|A^*B_r\,K_r\,A(u_0-u_*)\| \le (b_* + \tilde{\gamma}_{1/2})(\delta+\eta\|u_*\|), \qquad (4.29)$$

and for $r(\delta,\eta) > 1$

$$\sqrt{r'}\,\|A^*B_{r'}\cdot K_{r'}\cdot A(u_0-u_*)\| \ge (b_1 - \tilde{\gamma}_{1/2})(\delta+\eta\|u_*\|), \qquad (4.30)$$

where $r' = r(\delta,\eta)$ in case of Rule 1 and $r' = r(\delta,\eta) - 1$ in case of Rule 2.

2. We estimate the first term of the right-hand side of (4.26). Using Lemma 4.6 and inequality (4.29), we get $r(\delta,\eta) \ge r_*$, where r_* is the global minimum point of the function

$$\psi(r) = \|\tilde{K}_r(u_0-u_*)\| + 2\bar{\gamma}\,(b_* + \tilde{\gamma}_{1/2})\sqrt{r}\,(\delta+\eta\|u_*\|).$$

If $u_0 - u_* = |A_0|^p v$, $\|v\| \le \varrho$, $0 < p \le 2p_0$, then from (1.3) and Lemma 4.5 follows

$$\|\tilde{K}_r(u_0 - u_*)\| = \|\tilde{K}_r |A_0|^p v\| \le \|\tilde{K}_r |A|^p v\| + \|\tilde{K}_r (|A_0|^p - |A|^p) v\| \le$$

$$\le \gamma_{p/2} r^{-p/2} \varrho + c'_p (1 + |\ln \eta|) \eta^{\min\{1,p\}} \varrho.$$

Now applying the last inequality and (1.5), we get after non-complicated calculations that

$$\|\tilde{K}_{r(\delta,\eta)}(u_0 - u_*)\| \le \|\tilde{K}_{r_*}(u_0 - u_*)\| \le \psi(r_*) \le \min_r \psi(r) \le \tag{4.31}$$

$$\le c_p \left[(b_* + \tilde{\gamma}_{1/2})^{p/(p+1)} \varrho^{1/(p+1)} (\delta + \eta \|u_*\|)^{p/(p+1)} + |\ln \eta| \, \eta^{\min\{1,p\}} \varrho \right].$$

3. In the following we shall show that

$$\gamma_* \sqrt{r'} (\delta + \eta \|u_*\|) \le q_p \left[(b_1 - \tilde{\gamma}_{1/2})^{-1/(p+1)} \varrho^{1/(p+1)} (\delta + \eta \|u_*\|)^{p/(p+1)} + \right.$$

$$\left. + (b_1 - \tilde{\gamma}_{1/2})^{-1} |\ln \eta| \, \eta^{\min\{1,p\}} \varrho \right], \quad q_p = \text{const}. \tag{4.32}$$

If $r' \le R_* = (b_1 - \tilde{\gamma}_{1/2})^{-2/(p+1)} \varrho^{2/(p+1)} (\delta + \eta \|u_*\|)^{-2/(p+1)}$, then (4.32) is obvious. Consider now the case $r' > R_*$. If $u_0 - u_* = |A_0|^p v$, $\|v\| \le \varrho$, $0 < p \le 2p_0$, then using Lemma 4.5 and (1.3), we obtain

$$r' \|A^* B_r \cdot K_r \cdot A(u_0 - u_*)\| \le r' \|A^* B_r \cdot K_r \cdot A |A|^p v\| +$$

$$+ r' \|A^* B_r \cdot K_r \cdot A (|A_0|^p - |A|^p) v\| \le$$

$$\le \tilde{\gamma}_{(p+1)/2} r^{-p/2} \varrho + c'_p (1 + |\ln \eta|) \eta^{\min\{1,p\}} \varrho,$$

which together with (4.30) gives (4.32). Now the assertion of the Theorem in case of Rules 1 and 2 follows immediately from (4.26), (4.31), (4.32) and inequality $\sqrt{r(\delta,\eta)} \le \sqrt{r'} + 1$.

The proof in case of Rules 3 and 4 easily follows from proof in case of Rules 1,2 and from Part 4 of the proof of Theorem 4.1.■

References

1. Bakushinski A.B. About a general method for construction of the regularizing algorithms for linear ill-posed problem in Hilbert space. U.S.S.R. Comp. Maths. Mat. Phys. 7, №7 (1967), 672-677.
2. Engl H.W. and Gfrerer H. A posteriori parameter choice for general regularization methods for solving linear ill-posed problems. Appl. Numer. Math. 4 (1988), 395-417.
3. Groetsch C.W. On the rate of convergence for approximations of the

generalized inverse. Numer. Funct. Anal. and Optimiz. №1 (1979), 195-201.

4. Ivanov V.K. Approximate solution of operator equations of the first kind. U.S.S.R. Comp. Maths. Mat. Phys. 6, №6 (1966), 197-205.

5. King J.T. and Neubauer A. A variant of finite-dimensional Tikhonov regularization with a posteriori parameter choice. Computing 40 (1988), 91-109.

6. Morozov V.A. On the solution of functional equations by the methods of regularization. Soviet Math. Doklady 7 (1966), 414-417.

7. Raus T. On the discrepancy principle for the solution of ill-posed problems. Acta et comment. Univers. Tartuensis, 1984, 672, 16-26.

8. Raus T. An a-posteriori choice of the regularization parameter in case of approximately given error bound of data. Acta et comment. Univers. Tartuensis, 1990, 913, 73-87.

9. Vainikko G.M. The discrepancy principle for a class of regularization methods. U.S.S.R. Comp. Maths. Mat. Phys. 22, №3 (1982), 1-19.

10. Vainikko G.M. and Veretennikov A.Y. Iteration procedures in ill-posed problems. Nauka, Moscow, 1986.

О ВЫБОРЕ ПАРАМЕТРА РЕГУЛЯРИЗАЦИИ В СЛУЧАЕ ПРИБЛИЖЕННО ЗАДАННЫХ УРОВНЯХ ПОГРЕШНОСТЕЙ ИСХОДНЫХ ДАННЫХ

Т. Раус

Резюме

Рассматривается линейное уравнение $A_0u = f_0$ в гильбертовых пространствах H и F. Предполагается, что оператор A_0 и элемент f_0 заданы лишь приближенно: вместо них заданы некоторые A и f. Также предполагается, что нам известны некоторые предполагаемые уровня ошибок δ и η, но мы не знаем, действительно ли $\|A_0 - A\| \le \eta$, $\|f_0 - f\| \le \delta$ или нет. При таких условиях даются для класса регуляризационных методов (1.4) правила выбора параметра регуляризации. Доказывается, что приближенное решение сходится к точной, если только отношения $\|f - Au_*\|/(\delta + \eta\|u_*\|)$ и $\|A_0 - A\|/\eta$ остаются ограниченными в процессе $\delta \to 0$, $\eta \to 0$. В частном случае $\|f - Au_*\| \le \delta + \eta\|u_*\|$, $\|A_0 - A\| \le \eta$ дается оценка погрешности приближенного решения.

Acta et commentationes universitatis Tartuensis, 937, 1992, 90-102.

SOME NUMERICAL SCHEMES FOR THE IDENTIFICATION OF THE FILTRATION COEFFICIENT

Eero Vainikko and Gennadi Vainikko

In this paper, we deal with the coefficient inverse problem describing the filtration of ground water in a region $\Omega\subset\mathbb{R}^n$, $n\geq 2$. Introducing a weak formulation of the problem, discretization and regularization methods are constructed in a natural way. We present some numerical schemes and numerical results in the case where $n=2$, $\Omega\subset\mathbb{R}^2$ is a rectangle and Courant elements are used.

1. Inverse problem

1.1. Boundary value problem formulation. Let $\Omega\subset\mathbb{R}^n$ $(n\geq 2)$ be an open bounded region with a piecewise smooth boundary $\partial\Omega$; by ν we denote the outer unit normal to $\partial\Omega$. Let $\Gamma\subset\partial\Omega$ be a relatively open set having a piecewise smooth boundary on $\partial\Omega$. We shall deal with the following inverse problem: find a coefficient $a\in L^2(\Omega)$ such that

$$\begin{aligned} -div(a(x)\nabla u(x)) &= f(x), \quad x\in\Omega, \\ [a(x)\nabla u(x)]\cdot\nu(x) &= g(x), \quad x\in\Gamma, \end{aligned} \tag{1.1}$$

where $u\in W^{1,\infty}(\Omega)$, $f\in L^2(\Omega)$, $g\in L^2(\Gamma)$ are given functions. Physically, u can be interpreted as the piezometrical head of ground water in Ω; functon f characterizes the sources and sinks in Ω and function g characterizes the inflow and the outflow through $\Gamma\subset\partial\Omega$. The filtration (transmissivity) coefficient a is, physically, positive and piecewise smooth with possible discontinuities of the first kind on some surfaces in Ω.

We do not exclude the case of $\Gamma=\emptyset$. Boundary condition is omitted in (1.1) in this case.

Conditions (1.1) can be understood in the sense of distributions. We prefer to deal with the weak formulation of the problem.

1.2. Weak formulation. Let us provisionally assume that functions a and u are smooth (e.g. $a\in H^1(\Omega)$, $u\in W^{2,\infty}(\Omega)$). Multiplying the first equation of (1.1) by $w\in H^1(\Omega)$, integrating by parts and using the second equation we obtain

$$\int_\Omega a\nabla u\cdot\nabla w dx = \int_\Omega fw dx + \int_\Gamma gw dS + \int_{\partial\Omega\setminus\Gamma} (a\nabla u)\cdot\nu w dS.$$

Introduce the subspace

$$H^1(\Omega,\Gamma) = \{w\in H^1(\Omega): w(x)=0 \text{ for } x\in\partial\Omega\setminus\Gamma\} \subseteq H^1(\Omega).$$

We obtain the following weak formulation of inverse problem (1.1): find $a\in L^2(\Omega)$ such that

$$\int_\Omega a\nabla u\cdot\nabla w dx = \int_\Omega fw dx + \int_\Gamma gw dS, \quad \forall w\in H^1(\Omega,\Gamma). \tag{1.2}$$

The same formulation can be obtained in case of piecewise smooth a and u. Problem (1.2) has sense for $u\in W^{1,\infty}(\Omega)$, $a\in L^2(\Omega)$.

1.3. Operator equation formulation. Let us denote by G the space of gradients of functions $w\in H^1(\Omega,\Gamma)$:

$$G = G(\Omega,\Gamma) = \{\nabla w: w\in H^1(\Omega,\Gamma)\} \subset (L^2(\Omega))^n.$$

Let Q_G denote the orthoprojector in $(L^2(\Omega))^n$ corresponding to G. We observe that problem (1.2) is equivalent to equation

$$Ta = \nabla\psi \tag{1.3}$$

where operator $T=T_u\in\mathcal{L}(L^2(\Omega),G)$ is defined via the formula

$$Ta = Q_G(a\nabla u), \quad a\in L^2(\Omega), \tag{1.4}$$

and $\psi=\psi_{f,g}$ is a solution to the direct problem

$$\begin{aligned} -\nabla\psi(x) &= f(x), \quad x\in\Omega, \\ \nabla\psi(x)\cdot\nu(x) &= g(x), \quad x\in\Gamma, \qquad \psi(x)=0, \quad x\in\partial\Omega\setminus\Gamma. \end{aligned} \tag{1.5}$$

Indeed, ψ satisfies variational equality

$$\int_\Omega \nabla\psi\cdot\nabla w dx = \int_\Omega fw dx + \int_\Gamma gw dS, \quad \forall w\in H^1(\Omega,\Gamma), \tag{1.6}$$

thus (1.2) takes the form

$$\int_\Omega a\nabla u\cdot\nabla w dx = \int_\Omega \nabla\psi\cdot\nabla w dx, \quad \forall w\in H^1(\Omega,\Gamma),$$

and this is equivalent to (1.3) while $Q_G\nabla\psi = \nabla\psi$.

In case $\Gamma\neq\partial\Omega$, problem (1.5) is uniquely solvable. In case $\Gamma=\partial\Omega$, (1.5) is solvable if $\int_\Omega f dx + \int_{\partial\Omega} g dS = 0$; this condition is necessary for the solvability of inverse problem (1.2), too, if $\Gamma=\partial\Omega$.

1.4. Ill-posedness of the inverse problem. Operator $T\in\mathcal{L}(L^2(\Omega),G)$ has a very simple adjoint operator $T^*\in\mathcal{L}(G,L^2(\Omega))$:

$$T^*\nabla w = \nabla u\cdot\nabla w, \quad \nabla w\in G. \tag{1.7}$$

It is easy to see that the range $R(T^*)\subset L^2(\Omega)$ is non-closed in $L^2(\Omega)$ even if $|\nabla u|\geq c_0>0$ in Ω (here our case $n\geq 2$ essentially differs from case $n=1$). It is also clear that T^* is non-compact. Consequently, $T\in\mathfrak{L}(L^2(\Omega),G)$ has a non-closed range $R(T)\subset G$ and is non-compact, too. Thus, (1.3) is an ill-posed problem with a non-compact operator. This circumstance essentially influences the construction of discretization and regularization schemes for problem (1.1).

2. Discretization, regularization and identifiability (a survey)

2.1. Discretization. A natural way to discretize the inverse problem (1.1) is to apply finite element approximations to weak formulation (1.2) of the problem. Introduce finite dimensional subspaces $S_h\subset H^1(\Omega,\Gamma)$ depending on a discretization parameter $h>0$; we assume that S_h is full in $H^1(\Omega,\Gamma)$ as $h\to 0$, i.e. for every $w\in H^1(\Omega,\Gamma)$, there exists $w_h\in S_h$ such that $w_h\to w$ in $H^1(\Omega)$ as $h\to 0$. We introduce the following discrete version of problem (1.2):

$$\left.\begin{array}{l}\text{find } a_h\in L^2(\Omega) \text{ of minimal } L^2(\Omega) \text{ norm such that}\\ \displaystyle\int_\Omega a_h\nabla u\cdot\nabla w_h dx = \int_\Omega f w_h dx + \int_\Gamma g w_h dS, \quad \forall w_h\in S_h.\end{array}\right\} \tag{2.1}$$

If problem (1.2) is solvable then (2.1) is solvable, too, whereby uniquely, and the solutions are in the relation $a_h=P_{h,u}a$ where $P_{h,u}$ is the orthoprojector in $L^2(\Omega)$ corresponding to the subspace $\{a_h\in L^2(\Omega):\ a_h=\nabla u\cdot\nabla v_h,\ v_h\in S_h\}$; a consequence is that $a_h\to a_0$ in $L^2(\Omega)$ as $h\to 0$ where $a_0\in L^2(\Omega)$ is the normal solution (the solution of minimal $L^2(\Omega)$ norm) of problem (1.2). Conversely if (1.2) is non-solvable in $L^2(\Omega)$ but (2.1) is solvable, then $\|a_h\|_{L^2(\Omega)}\to\infty$ as $h\to 0$.

Choosing a basis $w_j=w_{j,h}$ $(j=1,\dots,l=l_h)$ of S_h, problem (2.1) can be formulated as follows: find

$$a_h = \sum_{j=1}^{l} c_j\ \nabla u\cdot\nabla w_j \tag{2.2}$$

solving the system of linear equations

$$Ac = d \tag{2.3}$$

where c is l-vector with components c_j, d is l-vector with components

$$d_i = \int_\Omega f w_i dx + \int_\Gamma g w_i dS, \quad i=1,\dots,l, \tag{2.4}$$

and $A=(a_{ij})$ is $l\times l$-matrix with elements

$$a_{ij} = \int_{\Omega} (\nabla u \cdot \nabla w_j)(\nabla u \cdot \nabla w_i)\, dx, \quad i,j=1,..,l. \tag{2.5}$$

2.2. Regularization. Consider a case where, instead of exact data denoted here by u_0, f_0 and g_0, we have polluted data $u=u_\eta \in W^{1,\infty}(\Omega)$, $f=f_\delta \in L^2(\Omega)$ and $g=g_\delta \in L^2(\Gamma)$ at our disposal. Then it is very risky to solve (2.3) directly, and a precedent regularization of problem (2.1) is needed. The Tikhonov regularization yields the following numerical scheme (cf. (2.2)-(2.3)):

$$a_{\alpha,h} = \sum_{j=1}^{l} c_{j,\alpha} \nabla u \cdot \nabla w_j \ , \quad (\alpha B + A) c_\alpha = d. \tag{2.6}$$

Here d is l-vector with components d_i defined in (2.4), c_α is l-vector with components $c_{j,\alpha}$ $(j=1,\dots,l)$, $A=(a_{ij})$ and $B=(b_{ij})$ are $l\times l$-matrices with elements a_{ij} defined in (2.5) and

$$b_{ij} = \int_{\Omega} \nabla w_j \cdot \nabla w_i dx, \quad i,j=1,\dots,l. \tag{2.7}$$

A suitable value of regularization parameter $\alpha>0$ depends on the error level of data. Assume that

$$\|\nabla \psi_\delta - \nabla \psi_0\|_{(L^2(\Omega))^n} \le \delta, \tag{2.8}$$

$$\sup_{x\in\Omega} |\nabla u_\eta(x) - \nabla u_0(x)| \le \eta \tag{2.9}$$

where ψ_0 and ψ_δ are the solutions to direct problem (1.5) with the right hand terms f_0, g_0 and f_δ, g_δ, respectively, and δ and η are small positive numbers. Then an a priori choice $\alpha=\alpha(h,\delta,\eta)$ such that

$$\alpha(h,\delta,\eta)\to 0, \quad (\delta^2+\eta^2)/\alpha(h,\delta,\eta)\to 0 \text{ as } h\to 0,\ \delta\to 0,\ \eta\to 0$$

guarantees the convergence $a_{\alpha(h,\delta,\eta),h}\to a_0$ in $L^2(\Omega)$ norm as $h\to 0$, $\delta\to 0$, $\eta\to 0$ where a_0 is the normal solution to (1.2) corresponding to the exact data u_0, f_0, g_0 (we assume that (1.2) with the exact data is solvable in $L^2(\Omega)$). The same result holds if $\alpha=\alpha(h,\delta,\eta)$ is chosen, according to the residual principle, so that

$$\delta + \langle Ac_\alpha, c_\alpha\rangle^{1/2}\eta < \langle Ac_\alpha - d, B^{-1}(Ac_\alpha - d)\rangle^{1/2} \le \beta(\delta + \langle Ac_\alpha, c_\alpha\rangle^{1/2}\eta)$$

where $\langle\cdot,\cdot\rangle$ denotes the scalar product in $\mathbb{R}^l$ and $\beta\ge 1$ is a constant not depending on h, δ, η.

The convergence result concerning the a priori parameter choice remains valid if (2.9) is replaced by conditions

$$\alpha_0 \in L^\infty(\Omega), \ \sup|u_\eta(x)| \le c, \ \|\nabla u_\eta - \nabla u_0\|_{(L^2(\Omega))^n} \le \eta$$

where constant c does not depend on η.

We refer to [6] for the more detailed exposition of discretization

and regularization methods for problem (1.2), including iterative regularization, to [8] for the general theory of regularization, to [2,3,5,7] for other methods (without regularization) to solve an inverse problem of type (1.1).

2.3. Identifiability. We say that the filtration coefficient a is L^2-identifiable from problem (1.2) on a subregion $\Omega'\subseteq\Omega$ if, for any solutions $a_1\in L^2(\Omega)$ and $a_2\in L^2(\Omega)$ to (1.2), $a_1(x)=a_2(x)$ for almost every $x\in\Omega'$. We now describe subregions $\Omega'\subseteq\Omega$ where a is L^2-identifiable assuming that

$$u \in W^{1,\infty}(\Omega) \cap W^{2,\infty}(\Omega_\varepsilon), \quad \forall \varepsilon>0. \tag{2.10}$$

Here Ω_ε consists of all points $x\in\Omega$ such that the distance from x to a nearest non-smoothness point of $\partial\Omega$ exceeds ε; if $\partial\Omega$ is smooth then (2.10) means that $u\in W^{2,\infty}(\Omega)$. Introduce a flow curve $x=\varphi(t,y)$ through a point $y\in\Omega$ as the maximal solution (the solution on the maximal time interval) to the Cauchy problem

$$dx/dt = -\nabla u(x), \quad x(0)=y. \tag{2.11}$$

Physically, ground water flows along those curves but with a speed depending on ∇u and the transmissivity coefficient (Darcy's law).

Due to (2.10), ∇u is bounded and locally Lipschitz continuous on Ω, with possible singularities of the Lipschitz coefficient only as x tends to a non-smoothness point of $\partial\Omega$. Therefore, problem (2.11) is uniquely solvable and $\varphi(t,y)$ is defined on a finite or infinite time interval (t_y^-,t_y^+); if t_y^- or t_y^+ is finite then $\varphi(t,y)$ tends to a point on $\partial\Omega$ as $t\searrow t_y^-$, respectively, $t\nearrow t_y^+$; if $t_y^-=-\infty$ or $t_y^+=\infty$ then $\varphi(t,y)$ tends to a critical point of u as $t\to-\infty$, respectively, $t\to\infty$.

Introduce the following subsets of Ω:

$\Omega^+ = \{y\in\Omega:\ \nabla u(y)\neq 0,\ t_y^+=\infty\}$,

$\Omega^- = \{y\in\Omega:\ \nabla u(y)\neq 0,\ t_y^-=-\infty\}$,

$\Omega_\Gamma^+ = \{y\in\Omega:\ \nabla u(y)\neq 0,\ t_y^+<\infty,\ \varphi(t,y)$ transversely reaches a smoothness point of $\Gamma\subseteq\partial\Omega$ as $t\nearrow t_y^+\}$,

$\Omega_\Gamma^- = \{y\in\Omega:\ \nabla u(y)\neq 0,\ t_y^->-\infty,\ \varphi(t,y)$ transversely reaches a smoothness point of $\Gamma\subseteq\partial\Omega$ as $t\searrow t_y^-\}$.

While $\Gamma\subset\partial\Omega$ is relatively open, Ω_Γ^+ and Ω_Γ^- are open subsets of Ω. The interior of $\Omega^\pm$ will be denoted by $int\Omega^\pm$.

It can be proved that, under condition (2.10), the filtration coefficient a is L^2-identifiable (and even L^1-identifiable) on sets $int\Omega^+$, $int\Omega^-$, Ω_Γ^+ and Ω_Γ^-. Thereby, on $int\Omega^+$ and $int\Omega^-$ the identifiability takes place even in the case $\Gamma=\emptyset$. These results can be extended to the case where ∇u has jumps on some (unknown) surfaces in Ω.

We refer to [6] for more detailed formulations; see [5,1], too, where the C^1-identifiability is examined.

3. Numerical schemes based on Courant elements

3.1. Introduction. Now we restrict ourselves to the case $n=2$ and

$$\Omega = \{(x_1,x_2)\in\mathbb{R}^2:\ 0<x_1<1,\ 0<x_2<1\}.$$

Introduce a standard uniform triangulation of Ω corresponding to a mesh size $h=1/N$ where N is an integer (see Figure 1, where $N=4$). Denote $x_{i_1i_2}=(i_1h, i_2h)$ and introduce the Courant element $w_{i_1i_2}$ which is a linear function on every triangle of the triangularization, continuous on Ω, equals 1 at $x_{i_1i_2}$ and 0 at other knots. Figures 2 and 3 describe the support of $w_{i_1i_2}$ in the cases $1<i_1,i_2<N-1$ and $i_1=0,\ 1<i_2<N-1$ respectively.

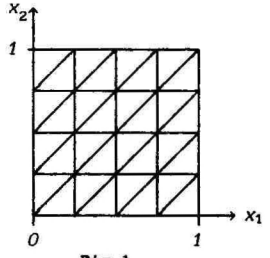


Fig.1

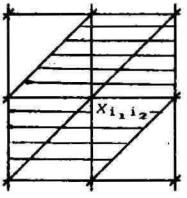


Fig.2

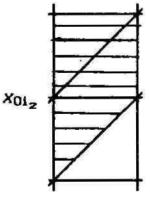


Fig.3

To concretize the numerical scheme of the Tikhonov method (2.6), we have to calculate the elements of the matrices A and B and the components of vector d (see (2.4), (2.5) and (2.7)); further, we have to simplify the formula for $a_{\alpha,h}$. The numerical schemes are slightly different in the cases $\Gamma=\emptyset$ and $\Gamma\neq\emptyset$.

3.2. Case $\Gamma=\emptyset$. In this case $H^1(\Omega,\Gamma)=\{w\in H^1(\Omega):\ w(x)=0$ for $x\in\partial\Omega\}=H_0^1(\Omega)$ and the subspace $S_h\subset H^1(\Omega,\Gamma)$ consists of the piecewise linear functions

$$w_h = \sum_{i_1=1}^{N-1}\sum_{i_2=1}^{N-1} c_{i_1i_2}w_{i_1i_2}$$

vanishing on $\partial\Omega$. We approximate

$$d_{i_1i_2} = \int_\Omega f w_{i_1i_2}dx \approx h^2 f(x_{i_1i_2}),\ 1\le i_1,i_2\le N-1.$$

The elements of the matrix B can be found exactly: for $1\le i_1,i_2,j_1,j_2\le N-1$,

$$b_{i_1i_2j_1j_2}=\int_\Omega \nabla w_{j_1j_2}\cdot\nabla w_{i_1i_2}dx=\begin{cases}4 \text{ if } (i_1,i_2)=(j_1,j_2),\\ -1 \text{ if } i_1=j_1,\ i_2=\pm1 \text{ or } i_1=\pm1,\ i_2=j_2,\\ 0 \text{ in other cases.}\end{cases}$$

To approximate the elements of the matrix A we first replace u by its

piecewise linear interpolant

$$u_h = \sum_{k_1=1}^{N-1} \sum_{k_2=1}^{N-1} u(x_{k_1 k_2}) w_{k_1 k_2}$$

and then integrate exactly: for $1 \le i_1, i_2, j_1, j_2 \le N-1$,

$$a_{i_1 i_2 j_1 j_2} = \int_\Omega (\nabla u \cdot \nabla w_{j_1 j_2})(\nabla u \cdot \nabla w_{i_1 i_2}) dx \approx \int_\Omega (\nabla u_h \cdot \nabla w_{j_1 j_2})(\nabla u_h \cdot \nabla w_{i_1 i_2}) dx =$$

$$= \tilde{a}_{i_1 i_2 j_1 j_2},$$

where

$$\tilde{a}_{i_1 i_2, i_1 i_2} = ([u(i_1+1, i_2) - u(i_1, i_2)]^2 + [u(i_1, i_2+1) - u(i_1, i_2)]^2 + [u(i_1-1, i_2) - u(i_1, i_2)]^2 + [u(i_1, i_2-1) - u(i_1, i_2)]^2 + [2u(i_1, i_2) - u(i_1-1, i_2) - u(i_1, i_2+1)]^2 + [2u(i_1, i_2) - u(i_1, i_2-1) - u(i_1+1, i_2)]^2)/2h^2,$$

$$\tilde{a}_{i_1-1 i_2, i_1 i_2} = [u(i_1, i_2) - u(i_1-1, i_2)]([u(i_1-1, i_2) - u(i_1-1, i_2-1)]/2 + [u(i_1, i_2+1) - u(i_1, i_2)]/2 - [u(i_1, i_2) - u(i_1-1, i_2)])/h^2,$$

$$\tilde{a}_{i_1 i_2-1, i_1 i_2} = [u(i_1, i_2) - u(i_1, i_2-1)]([u(i_1, i_2-1) - u(i_1-1, i_2-1)]/2 + [u(i_1+1, i_2) - u(i_1, i_2)]/2 - [u(i_1, i_2) - u(i_1, i_2-1)])/h^2,$$

$$\tilde{a}_{i_1-1 i_2-1, i_1 i_2} = ([u(i_1, i_2-1) - u(i_1-1, i_2-1)][u(i_1, i_2) - u(i_1, i_2-1)] + [u(i_1, i_2) - u(i_1-1, i_2)][u(i_1-1, i_2) - u(i_1-1, i_2-1)])/2h^2,$$

$$\tilde{a}_{i_1 i_2, i_1-1 i_2} = \tilde{a}_{i_1-1 i_2, i_1 i_2},$$

$$\tilde{a}_{i_1 i_2, i_1 1 i_2-1} = \tilde{a}_{i_1 i_2-1, i_1 i_2},$$

$$\tilde{a}_{i_1 i_2, i_1-1 i_2-1} = \tilde{a}_{i_1-1 i_2-1, i_1 i_2}$$

and $\tilde{a}_{i_1 i_2, i_1 1 i_2} = 0$ in other cases. We see that system (2.6) has a symmetric sparse matrix comparable with the matrix when the Dirichlet problem for the Poisson equation is solved using the Courant elements. Solving the system, we can specify the formula for $a_{\alpha,h}$ as follows: $a_{\alpha,h}$ is constant on any triangle of our triangulation; in the triangle with $x_{i_1} < x_1 < x_{i_1+1}$, $x_{i_2-1} < x_1 < x_{i_2}$,

$$a_{\alpha,h} = [(c_{i_1+1 i_2} - c_{i_1 i_2})(u_{i_1+1 i_2} - u_{i_1 i_2}) + (c_{i_1 i_2} - c_{i_1 i_2-1})(u_{i_1 i_2} - u_{i_1 i_2-1})]/h^2, \tag{3.1}$$

in the triangle with $x_{i_1-1} < x_1 < x_{i_1}$, $x_{i_2} < x_1 < x_{i_2+1}$,

$$a_{\alpha,h} = [(c_{i_1 i_2+1} - c_{i_1 i_2})(u_{i_1 i_2+1} - u_{i_1 i_2}) + (c_{i_1 i_2} - c_{i_1-1 i_2})(u_{i_1 i_2} - u_{i_1-1 i_2})]/h^2 \tag{3.2}$$

3.3. Case $\Gamma = ((0, x_2): 0 < x_2 < 1)$. **In this case the subspace** $S_h \subset H^1(\Omega, \Gamma)$ **consists of the piecewise linear functions**

$$w_h = \sum_{i_1=0}^{N-1} \sum_{i_2=1}^{N-1} c_{i_1 i_2} w_{i_1 i_2}.$$

The formulas corresponding to the inner knots are the same as in Section 3.2. In addition, we obtain the formulas corresponding to the boundary knots:

$$d_{0i_2} = \int_\Omega f w_{i_1 i_2} dx + \int_\Gamma g w_{i_1 i_2} dS \approx \frac{1}{2} h^2 f(x_{0,i_2}) + hg(x_{0,i_2}), \quad 1 \le i_2 \le N-1,$$

$$b_{0,i_2,j_1 j_2} = b_{j_1 j_2, 0 i_2} = \begin{cases} 2, & \text{if } j_1=0, \ j_2=i_2 \\ -1/2, & \text{if } j_1=0, \ j_2=i_2+1 \\ -1, & \text{if } j_1=1, \ j_2=i_2 \\ 0 & \text{in other cases,} \end{cases}$$

$$a_{0i_2,0i_2} = ([u(1,i_2)-u(0,i_2)]^2 + [u(0,i_2+1)-u(0,i_2)]^2 + [2u(0,i_2)-u(0,i_2-1)-u(1,i_2)]^2)/2h^2,$$

$$a_{0i_2,1i_2} = [u(1,i_2)-u(0,i_2)]([u(0,i_2)-u(0,i_2-1)]/2 + [u(1,i_2+1)-u(1,i_2)]/2 - [u(1,i_2)-u(0,i_2)])/h^2,$$

$$a_{0i_2-1,0i_2} = [2u(0,i_2)-u(1,i_2)-u(0,i_2-1)][u(0,i_2-1)-u(0,i_2)]/2h^2,$$

$$a_{0i_2-1,1i_2} = ([u(1,i_2-1)-u(0,i_2-1)][u(1,i_2)-u(1,i_2-1)] + [u(1,i_2)-u(0,i_2)][u(0,i_2)-u(0,i_2-1)])/2h^2,$$

$$a_{1i_2,0i_2} = a_{0i_2,1i_2},$$

$$a_{0i_2,0i_2-1} = a_{0i_2-1,0i_2},$$

$$a_{1i_2,0i_2-1} = a_{0i_2-1,1i_2}.$$

3.4. Solution of the system $(\alpha B+A)c=d$. Solving the system $(\alpha B+A)c=d$ is complicated because of the big dimensions of matrix $\alpha B+A$ (in case $\Gamma=\emptyset$: $(N-1)^2 \times (N-1)^2$ and in case $\Gamma=((0,x_2): 0<x_2<1)$: $N(N-1)\times N(N-1)$). At the same time, matrix $\alpha B+A$ is symmetric and has got quite a sparse structure. In Figure 4 we represent the nonzero elements of matrix B in case $\Gamma=((0,x_2): 0<x_2<1)$, $N=4$:

$$B = \begin{pmatrix}
2 & -1 & & & -\frac{1}{2} & + & & & & & & \\
-1 & 4 & -1 & & & -1 & + & & & & & \\
 & -1 & 4 & -1 & & & -1 & + & & & & \\
 & & -1 & 4 & & & & -1 & & & & \\
-\frac{1}{2} & & & & 2 & -1 & & & -\frac{1}{2} & + & & \\
+ & -1 & & & -1 & 4 & -1 & & & -1 & + & \\
 & + & -1 & & & -1 & 4 & -1 & & & -1 & + \\
 & & + & -1 & & & -1 & 4 & & & & -1 \\
 & & & & -\frac{1}{2} & & & & 2 & -1 & & \\
 & & & & + & -1 & & & -1 & 4 & -1 & \\
 & & & & & + & -1 & & & -1 & 4 & -1 \\
 & & & & & & + & -1 & & & -1 & 4
\end{pmatrix}$$

Fig. 4.

Matrix A has got nonzeroes in the same positions as matrix B and additionally the elements that are marked with "+" in Fig. 4. Wholly, there are $(2N-3)(2N-1)$ (in case $\Gamma=\emptyset$, $(2N-3)^2$) different nonzeroes in the matrix $\alpha B+A$ which has the following block tridiagonal form:

$$\alpha B-A = \begin{pmatrix} D_1 & G_{12} & 0 & \cdot & \cdot & 0 \\ G_{21} & D_2 & G_{23} & & \cdot & \cdot \\ 0 & G_{32} & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & & 0 \\ \cdot & & \cdot & & D_{N-2} & G_{N-2,N-1} \\ 0 & \cdot & \cdot & 0 & G_{N-1,N-2} & D_{N-1} \end{pmatrix},$$

where each block in turn is tridiagonal or twodiagonal,

$$D_k = \begin{pmatrix} \tilde{a}_{0k,0k}+2\alpha & \tilde{a}_{0k,1k}-\alpha & 0 & \cdot & \cdot & 0 \\ \tilde{a}_{1k,0k}-\alpha & \tilde{a}_{1k,1k}+4\alpha & \tilde{a}_{1k,2k}-\alpha & & \cdot & \cdot \\ 0 & \tilde{a}_{2k,1k}-\alpha & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & & 0 \\ \cdot & & \cdot & & \tilde{a}_{N-1\ k,N-1\ k}+4\alpha & \tilde{a}_{N-1\ k,Nk}-\alpha \\ 0 & \cdot & \cdot & 0 & \tilde{a}_{Nk,N-1\ k}-\alpha & \tilde{a}_{Nk,Nk}+4\alpha \end{pmatrix},$$

$$G_{k,k-1} = \begin{pmatrix} \tilde{a}_{0k,0k-1}-\frac{1}{2}\alpha & 0 & 0 & \cdot & \cdot & 0 \\ \tilde{a}_{1k,0k-1} & \tilde{a}_{1k,1k-1}-\alpha & 0 & & \cdot & \cdot \\ 0 & \tilde{a}_{2k,1k-1} & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & & 0 \\ \cdot & & \cdot & & \tilde{a}_{N-1\ k,N-1\ k-1}-\alpha & 0 \\ 0 & \cdot & \cdot & 0 & \tilde{a}_{Nk,N-1\ k-1} & \tilde{a}_{Nk,Nk-1}-\alpha \end{pmatrix}$$

and $G_{k-1,k}=G_{k,k-1}^{T}$. In case $\Gamma=\emptyset$ submatrixes D_k and $G_{k,k\pm1}$ lack the first column and the first row.

Due to the sparsity of matrix $\alpha B+A$, iterative methods are more convenient for solving the system because of the non-changing structure of the matrix during the solution process. We used the method of block successive overrelaxation (block SOR) iterations [4] keeping in memory and operating with only nonzero elements of matrix A. Block SOR iterations are obtained solving on each step the system

$$\begin{aligned} D_1\hat{c}_1^s &= -G_{1,2}c_2^{s-1} + d_1, \\ D_i\hat{c}_i^s &= -G_{i,i-1}c_{i-1}^s - G_{i,i+1}c_{i+1}^{s-1} + d_i, \quad i=2,\dots N-2 \\ D_{N-1}\hat{c}_{N-1}^s &= -G_{N-1,N-2}c_{N-1}^s + d_{N-1}, \end{aligned} \tag{3.3}$$

(the block Gauss-Seidel method, [4]) where c^s and d are partitioned commensurately with $\alpha B+A$, and then define

$$c_i^s = c_i^{s-1} + \omega(\hat{c}_i^s - c_i^{s-1}).$$

Here $\hat{c}_i^s$ is the Gauss-Seidel iterate produced by (3.3). Parameter ω is introduced to obtain quicker convergence; we found it out by experiment.

4. Numerical results

Here we consider a model problem defining

$$a(x) = e^{x_1+x_2} + \sin 2(x_1-x_2), \tag{4.1}$$

$$u(x) = (x_1-b_1)^2 + (x_2-1/3)^2 \tag{4.2}$$

where the parameter b_1 takes the values $-1/4$, 0 and $1/4$; the functions f and g are calculated in correspondence to (1.1). After that we perturbed the knot values of u and d and constructed $a_{\alpha,h}$ by the numerical schemes described in Sections 3.2 and 3.3, found the best regularization parameter α and compared the corresponding approximation $a_{\alpha,h}$ with the exact a.

For finding out the best value of regularization parameter α we used the method of golden section combined with SOR iterations as follows: we started each step with updating the values of α_1, α_2 and choosing a new parameter $\alpha\in[\alpha_1,\alpha_2]$ according to the method of golden section for minimizing the value of $\|a_{\alpha,h}-a\|_{L_2}$. After that we carried out SOR iterations (3.3) with the stopping condition $\|c^s-c^{s-1}\|_{R^n}\le q|\alpha_1-\alpha_2|$ and calculated a new value of $\|a_{\alpha,h}-a\|_{L_2}$. We repeated these steps until we had reached sufficiently short interval $[\alpha_1,\alpha_2]$. Note that on the first step we took $c^0=(0,0,\dots,0)$ and on the following steps we just continued the previous iterations with the new value of α. In our case it was sufficient to take $q=0.1$ that made up about 3...10 SOR iterations on each step.

Some numerical results are presented in Tables 1 - 6. They correspond to $N=4,8,16,32$; the perturbation level of the knot values of u and d is $\pm\varepsilon$ where $\varepsilon=0$, 0.01 or 0.1 whereby the sign + or - is chosen using a special randomization program.

It is slightly surprising that in some cases with $\varepsilon=0$ the best results were obtained by $\alpha=0$, i.e. without a regularization, although system (2.3) is perturbed by the approximate evaluation of $a_{i_1i_2j_1j_2}$ and $d_{i_1i_2}$ (see Sections 3.2 and 3.3). But these perturbations converge to 0 as $h\longrightarrow 0$. We represent some graphs of norm $\|a_{\alpha,h}-a\|_{L_2}$ depending on α on Figure 5.

The numerical results with the perturbation level $\varepsilon=0,1$ are bad, especially for great N. This is natural while the error of ∇u_h caused by the perturbation of knot values by $\pm\varepsilon$ may be $2\varepsilon/h$, e.g. 6.4 for $\varepsilon=0.1$, $N=32$. Thus, $\varepsilon=0,1$ is an extremely great perturbation level in this example.

The numerical results with the perturbation level $\varepsilon=0,1$ are bad, especially for great N. This is natural while the error of ∇u_h caused by the perturbation of knot values by $\pm\varepsilon$ may be $2\varepsilon/h$, e.g. 6.4 for $\varepsilon=0.1$, $N=32$. Thus, $\varepsilon=0,1$ is an extremely great perturbation level in this

example.

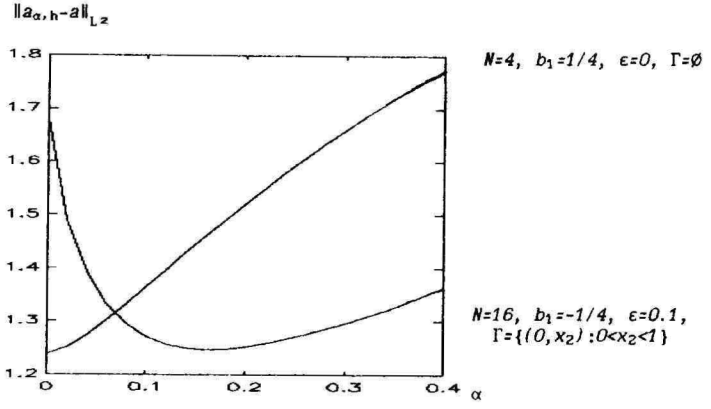


Fig.5.

We can see the divergence of the method in the case $\Gamma=\emptyset$, $b_1=-1/4$. It corresponds to non-identifiability of a. Indeed, the flow curves go through Ω during a finite time (see Figure 6), and $\Omega^+=\emptyset$, $\Omega^-=\emptyset$. In the other cases a is L-identifiable from problem (1.2): $\Omega^+=\Omega$ in cases $b_1=0$ and $b_1=1/4$; $\Omega_\Gamma^+=\Omega$ in case $b_1=-1/4$, $\Gamma=\{(0,x_2):\ 0<x_2<1\}$. We can see the convergence of the method in those cases.

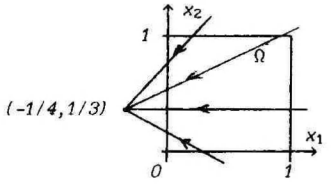


Fig.6

In the tables, α_{opt} means the value of the regularization parameter α giving the best result. The corresponding error $\|a_{\alpha,h}-a\|_{L_2}$ and the residual $\langle Ac^\alpha-d, B^{-1}(Ac^\alpha-d)\rangle^{1/2}=\alpha\|Bc_\alpha\|_{R_1}^{1/2}$ are also presented. Note that $\|a\|_{L_2}=3.256$ in this example, so the relative error $\|a_{\alpha,h}-a\|_{L_2}/\|a\|_{L_2}$ is much smaller than the error presented in the tables.

We also compared the concrete values of given function (4.1) with its approximations $a_{\alpha,h}$ (from (3.1) or (3.2), respectively) on triangles of our triagularization of Ω. It came out that in most triangles the approximation gave guite good results, but at the same time there were regions of Ω where approximation completely failed. One of such regions occured around the point $(x_1,x_2)=(1/4,1/3)$ in the case $b_1=1/4$: the values of $a_{\alpha,h}$ in one or two triangles there were even negative. (This is caused of the fact, that $\mathrm{grad}u=0$ in that point). Other such regions occured

mostly on boundary triangles of Ω.

Table 1: $b_1=-1/4$, $\Gamma=\emptyset$

		$\varepsilon=0$			$\varepsilon=0.01$			$\varepsilon=0.1$	
N	α_{opt}	Error	Residual	α_{opt}	Error	Residual	α_{opt}	Error	Residual
4	0.0	1.906	0.0	0.0	1.906	0.0	0.0	1.914	0.0
8	0.0	1.656	0.0	0.0	1.657	0.0	0.011	1.726	0.029
16	0.0	1.553	0.0	0.0	1.558	0.0	0.181	1.903	0.435
32	0.0	1.514	0.0	0.026	1.538	0.080	0.764	2.259	1.718

Table 2: $b_1=-1/4$, $\Gamma=\{(0,x_2):0<x_2<1\}$

		$\varepsilon=0$			$\varepsilon=0.01$			$\varepsilon=0.1$	
N	α_{opt}	Error	Residual	α_{opt}	Error	Residual	α_{opt}	Error	Residual
4	0.0	1.229	0.0	0.0	1.141	0.0	0.011	1.171	0.031
8	0.0	0.740	0.0	0.0	0.708	0.0	0.018	0.874	0.055
16	0.0	0.450	0.0	0.0	0.467	0.0	0.181	1.248	0.500
32	0.0	0.274	0.0	0.026	0.431	0.090	0.854	1.820	1.982

Table 3: $b_1=0$, $\Gamma=\emptyset$

		$\varepsilon=0$			$\varepsilon=0.01$			$\varepsilon=0.1$	
N	α_{opt}	Error	Residual	α_{opt}	Error	Residual	α_{opt}	Error	Residual
4	0.0	1.401	0.0	0.0	1.402	0.0	0.008	1.416	0.030
8	0.0	0.874	0.0	0.0	0.877	0.0	0.045	1.076	0.206
16	0.0	0.555	0.0	0.001	0.592	0.006	0.148	1.293	0.663
32	0.0	0.351	0.0	0.005	0.648	0.028	0.764	2.185	1.857

Table 4: $b_1=0$, $\Gamma=\{(0,x_2):0<x_2<1\}$

		$\varepsilon=0$			$\varepsilon=0.01$			$\varepsilon=0.1$	
N	α_{opt}	Error	Residual	α_{opt}	Error	Residual	α_{opt}	Error	Residual
4	0.032	1.275	0.101	0.0	1.153	0.0	0.020	1.176	0.065
8	0.012	0.761	0.048	0.001	0.698	0.004	0.077	0.984	0.294
16	0.005	0.476	0.025	0.005	0.506	0.020	0.292	1.219	1.020
32	0.002	0.293	0.011	0.015	0.604	0.067	0.854	1.999	2.047

Table 5: $b_1=1/4$, $\Gamma=\emptyset$

		$\varepsilon=0$			$\varepsilon=0.01$			$\varepsilon=0.1$	
N	α_{opt}	Error	Residual	α_{opt}	Error	Residual	α_{opt}	Error	Residual
4	0.0	1.239	0.0	0.0	1.240	0.0	0.0	1.279	0.0
8	0.0	0.750	0.0	0.0	0.756	0.0	0.023	1.089	0.118
16	0.0	0.443	0.0	0.003	0.518	0.020	0.133	1.522	0.586
32	0.0	0.266	0.0	0.014	0.615	0.080	0.528	2.005	1.885

Table 6: $b_1=1/4$, $\Gamma=\{(0,x_2):0<x_2<1\}$

		$\varepsilon=0$			$\varepsilon=0.01$			$\varepsilon=0.1$	
N	α_{opt}	Error	Residual	α_{opt}	Error	Residual	α_{opt}	Error	Residual
4	0.032	1.273	0.129	0.011	1.190	0.043	0.002	1.234	0.009
8	0.011	0.769	0.056	0.003	0.706	0.017	0.046	1.145	0.219
16	0.004	0.469	0.022	0.008	0.531	0.046	0.164	1.472	0.705
32	0.002	0.289	0.011	0.016	0.617	0.087	0.764	1.977	2.213

References

1. Chicone C. and Gerlach J. Identifiability of distributed parameters. In: Inverse and Ill-Posed Problems, Engl H.W. and Groetsch (Ed.). Academic Press, Boston etc.,1987, pp. 513-521.
2. Kohn R.V., Lowe B.C. A variational method for parameter estimation. RAIRO Math. Mod. and Num. Anal., 1988, V.22, pp. 119-198.
3. Kunisch K. A review of some recent results on the output least squares formulation of parameter estimation problem. Automatica 24, 1988, pp. 210-221.
4. Ortega J.M. Introduction to Parallel and Vector Solution of Linear Systems. New York and London, Plenum Press, 1988.
5. Richter G.R. An inverse problem for the steady state diffusion equation. SIAM J. Appl. Math. 41, 1981, No 2, pp. 210-221.
6. Vainikko G. Identification of filtration coefficient. In: Ill-Posed Problems in Natural Sciences, 1992 (to appear)
7. Yeh W.W.-G. Review of parameter identification procedures in ground-water hydrology: the inverse problem. Water Resources Res. 22, 1986, No 2, pp. 95-108.
8. Вайникко Г.М., Веретенников А.Ю. Итерационные процедуры в некорректных задачах. М., Наука, 1986.

НЕКОТОРЫЕ ЧИСЛЕННЫЕ СХЕМЫ ДЛЯ ИДЕНТИФИКАЦИИ КОЭФФИЦИЕНТА ФИЛЬТРАЦИИ

Ээро Вайникко и Геннадий Вайникко

Резюме

В статье рассматривается обратная задача определения коэффициента фильтрации грунтовых вод в области $\Omega \subset \mathbb{R}^n$, $n \geq 2$. Вводится слабая формулировка задачи и естественным образом конструируются методы дискретизации и регуляризации. Представляются некоторые численные схемы вместе с результатами вычислений в случае, когда $n=2$, а область $\Omega \subset \mathbb{R}^2$ является квадратом, используя при дискретизации элементы Куранта.

CONTENTS СОДЕРЖАНИЕ

Tartu Ülikooli toimetised.
Vihik 937.
METHODS FOR SOLUTION OF INTEGRAL EQUATIONS
AND ILL-POSED PROBLEMS.
Matemaatika- ja mehhaanika-alaseid töid.
Tartu Ülikool.
EV, 202400 Tartu, Ülikooli 18.
Vastutav toimetaja P. Oja.
6,42. 6,5. T. 644.300.
Hind rbl. 8.
TÜ trükikoda. EV, 202400 Tartu, Tiigi 78.