Jiří Neuberg

Universality of the best determined terms method

Aplikace matematiky, Vol. 24 (1979), No. 6, 401--405

Persistent URL: http://dml.cz/dmlcz/103823

UNIVERSALITY OF THE BEST DETERMINED TERMS METHOD*)

JIŘÍ NEUBERG

(Received April 7, 1976)

1. INTRODUCTION

Let X, Y be real separable Hilbert Spaces and let $T \in [X, Y]$ be a compact linear operator. Let us consider $y \in R(T)$. Then the problem

$$Tx = y \tag{1.1}$$

has a solution, which need not be uniquely determined in general. A vector $x_0 \in X$ is called the normal solution to (1.1), if the following conditions are satisfied:

$$Tx_0 = y$$

and

$$\|x_0\|_X = \min \{\|x\|_X : Tx = y\} .$$

Obviously, x_0 is uniquely determined. In practical calculations it is quite usual that the right hand side in (1.1) is not known exactly; we are given a vector $y^* = y + \varepsilon$ such that $\varepsilon \in Y$ and $\|\varepsilon\|_Y \leqq \Delta$, where $\Delta \geqq 0$ is an a priori given bound. Our aim is to determine an approximation of the normal solution x_0. We denote

$$\mathfrak{N} = \{y \in R(T) : \|y - y^*\|_Y \leqq \Delta\} .$$

Definition 1.1. *The set $\{\mathfrak{N}_i\}_{i\in P}$ is called an a priori decomposition of $\mathfrak{N}$, if*

(i) $\emptyset \neq \mathfrak{N}_i \subset \mathfrak{N}$ *for* $i \in P$,

(ii) $\mathfrak{N}_i \subset \mathfrak{N}_j$ *for* $i \leqq j$, $i, j \in P$,

(iii) $(\bigcup_{i\in P} \mathfrak{N}_i)^c \supseteq \mathfrak{N}$,

where $\emptyset \neq P \subset \mathbb{N}$ and $\mathbb{N}$ is the set of positive integers.

It is well known (see [1, p. 328]) that the operator T has a canonical decomposition:

$$T = \sum_{i\in\mathscr{K}} d_i(\cdot, v_i)_X\, u_i\,, \tag{1.2}$$

where $d_i \geqq 0$ are the singular values of T (with out loss of generality we assume that $d_i \geqq d_j$ if $i \geqq j$ and $i, j \in \mathscr{K}$), u_i and v_i (for $i \in \mathscr{K}$) are the corresponding singular

*) See [2].

vectors which are constructed so that $\{u_i\}_{i\in I}$ and $\{v_i\}_{i\in J}$ respectively form complete orthonormal bases of X and Y while $\mathscr{K} = I \cap J$, where I (and similarly J) is either a set of the type $I = \{1, 2, \ldots, m\}$, where $m \in \mathbb{N}$, or $I = \mathbb{N}$. Further, let us define an operator $T^+ : R(T) \to X$ as follows:

$$T^+ = \sum_{i\in\mathscr{K}} d_i^+ (\cdot, u_i)_Y \, v_i \,, \tag{1.3}$$

where

$$d_i^+ = \begin{cases} d_i^{-1} & \text{if} \quad d_i > 0 \,, \\ 0 & \text{if} \quad d_i = 0 \,. \end{cases}$$

Definition 1.2. *Let $\{\mathfrak{N}_i\}_{i\in P}$ be an a priori decomposition of $\mathfrak{N}$. Then we denote:*

(i) $\omega(W, \mathfrak{N}_i) = \sup \{\|Wy^* - T^+y\|_X : y \in \mathfrak{N}_i\}$,

where $W \in [Y, X]$, *see* [3],

(ii) $\Omega(\mathfrak{N}_i) = \inf \{\omega(W, \mathfrak{N}_i) : W \in [Y, X]\}$.

Definition 1.3. *A vector $\hat{x} \in X$ is called a universal approximation to the normal solution x_0, if*

(i) $\hat{x} = \hat{W}y^*$, $\hat{W} \in [Y, X]$,

(ii) *there exists* $i(o) \in I$ *so that* $\omega(\hat{W}_{i(o)}) = \Omega(\mathfrak{N}_{i(o)}) \leqq \Omega(\mathfrak{N}_j)$ *for* $j \in P$,

(iii) $\omega(\hat{W}, \mathfrak{N}_i) \leqq d\, \Omega(\mathfrak{N}_i)$ *for* $i \in P$, *where* $d \geqq 1$ *is a constant independent of* i.

2 A SPECIAL CASE OF AN A PRIORI DECOMPOSITION OF $\mathfrak{N}$

Let $j \in I$. Let $A(j) \subset I$ be the sets such that

(i) $j \in A(j)$ and $A(j) \cup B(j) = I$,

(ii) *if* $i \in A(j)$ then $i \leqq j$,

(iii) if $k \in B(j)$ then $j < k$.

Let us define the set $\mathfrak{N}_j$ $(j \in I)$ by setting

$$\mathfrak{N}_j = \{y \in \mathfrak{N} : (y, u_i)_Y = 0 \,, \quad i \in B(j)\} \,.$$

For $B(j) = \emptyset$ we put $\mathfrak{N}_j = \mathfrak{N}$. Let us assume that there exists an index $k(\Delta) \in \mathscr{K}$ such that $d_{k(\Delta)} \neq 0$ and

$$\sum_{i\in B(k(\Delta))} |(y^*, u_i)_Y|^2 \leqq \Delta^2 \,, \tag{2.1}$$

$$\text{if} \quad p \in I \quad \text{is such that} \quad \sum_{i\in B(p)} |(y^*, u_i)_Y|^2 \leqq \Delta^2 \quad \text{then} \quad k(\Delta) \leqq p \,. \tag{2.2}$$

Remark. In this paper we use the following notation:

$$\sum_{i\in B(p)} |(y^*, u_i)_Y|^2 = 0 \quad \text{if} \quad B(p) = \emptyset \,.$$

Now, let us introduce the set $P = B(k(\Delta) - 1) \cap \mathscr{K}$. P is not empty in the case of the best determined terms method.

Theorem 2.1.

(i) *If* $A(k(\Delta) - 1) \neq \emptyset$ *then* $\mathfrak{N}_i = \emptyset$ *for* $i \in A(k(\Delta) - 1)$,

(ii) $\{\mathfrak{N}_i\}_{i \in P}$ *is an a priori decomposition of* $\mathfrak{N}$.

Proof.

(i) Let $i \in A(k(\Delta) - 1)$. For $y \in \mathfrak{N}_i$ we have $y \in \mathfrak{N}$ and $(y, u_j)_Y = 0$ for $j \in B(i)$. This implies

$$\|y^* - y\|_Y^2 = \sum_{j \in A(i)} |(y^* - y, u_j)_Y|^2 + \sum_{j \in B(i)} |(y^*, u_j)_Y|^2 \leqq \Delta^2 . \tag{2.3}$$

By (2.1), (2.2) and by the assumption $A(k(\Delta) - 1) \neq \emptyset$ it follows that

$$\sum_{j \in A(i)} |(y^*, u_j)_Y|^2 > \Delta^2 .$$

Thus we obtain a contradiction with (2.3).

(ii) Evidently, (ii) of Definition 1.1 holds and $\mathfrak{N}_i \subset \mathfrak{N}$ for $i \in P$. Let us show that $\mathfrak{N}_i \neq \emptyset$. We define $\tilde{y} = \sum_{j \in A(k(\Delta))} (y^*, u_j)_Y u_j$. Then $\tilde{y} \in \mathfrak{N}_{k(\Delta)} \subset \mathfrak{N}_i$.

Now, let us prove (iii) of Definition 1.1. It is easy to verify that (iii) holds if card $P < \infty$. Let $y_0 \in \mathfrak{N}$ be such that $y_0 \notin (\bigcup_{i \in P} \mathfrak{N}_i)^c$. Then there exists $\delta > 0$ so that

$$\inf\{\|y_0 - y\|_Y : y \in (\bigcup_{i \in P} \mathfrak{N}_i)^c\} \geqq \delta > 0 .$$

Obviously,

$$y_0 = \sum_{j \in A(i)} (y_0, u_j)_Y u_j + \sum_{j \in B(i)} (y_0, u_j)_Y u_j$$

and

$$\lim_{i \to \infty} \| \sum_{j \in B(i)} (y_0, u_j)_Y u_j \|_Y = 0 .$$

This completes the proof.

We denote

$$\Delta_j^2 = \Delta^2 - \sum_{i \in B(j)} |(y^*, u_i)_Y|^2 \quad \text{for} \quad j \in P ,$$

and

$$T^j = \sum_{i \in A(j)} d_i^+(\cdot, u_i)_Y v_i \quad \text{for} \quad j \in P .$$

Theorem 2.2. *For* $j \in P$,

$$\Omega(\mathfrak{N}_j) = d_{j(o)}^+ \Delta_{j(o)} ,$$

where $j(o) \in P$ *is such that* $d_{j(o)}^+ = \max\{d_i^+ : i \in A(j) \setminus A(k(\Delta)) - 1)\}$ *and if* $p \in A(j) \setminus A(k(\Delta) - 1)$ *is such that* $d_{j(o)}^+ = d_p^+$ *then* $p \leqq j(o)$.

Proof. First we prove that for $j \in P$ it holds

$$\omega(T^j, \mathfrak{N}_j) = d_{j(o)}^+ \Delta_{j(o)} . \tag{2.4}$$

Obviously,

$$\omega(T^j, \mathfrak{N}_j)^2 = (d_{j(o)}^+)^2 \sup\{ \sum_{i \in A(j(o))} |(y^* - y, u_i)_Y|^2 : y \in \mathfrak{N}_j\} . \tag{2.5}$$

It is easy to verify that

$$(2.6)\qquad \sup\{\sum_{i\in A(j(o))} |(y^* - y, u_i)_Y|^2 : y \in \mathfrak{R}_j\} = $$
$$= \sup\{\sum_{i\in A(j(o))} |(y^* - y, u_i)_Y|^2 : y \in \mathfrak{R}_{j(o)}\}$$

and for all $y \in \mathfrak{R}_{j(o)}$ it holds $\sum_{i\in A(j(o))} |(y^* - y, u_i)_Y|^2 \leqq \Delta^2_{j(o)}$. It follows that

$$(2.7)\qquad \omega(T^j, \mathfrak{R}_j) = d^+_{j(0)}\Delta_{j(o)}\,.$$

Let us denote

$$\tilde{y} = \sum_{i\in A(j(o)),} (y^*, u_i)_Y\, u_i + \Delta_{j(o)} u_{j(o)}\,.$$

Evidently $\tilde{y} \in \mathfrak{R}_{j(o)} \subset \mathfrak{R}_j$ and thus (2.4) is fulfilled because

$$(2.8)\qquad \|T^j y^* - T^+ \tilde{y}\|_X = d^+_{j(o)}\, \Delta_{j(o)}\,.$$

Now let us prove Theorem 2.2. For $W \in [Y, X]$ and $y \in \mathfrak{R}_j$ we have

$$(2.9)\qquad \|Wy^* - T^j y\|^2_X = \sum_{i\in B(j(o))} |(Wy^*, v_i)_X|^2 +$$
$$+ \sum_{i\in A(j(o))} |(Wy^*, v_i)_X - d^+_i (y, u_i)_Y|^2\,.$$

We denote

$$\beta = \sum_{i\in A(j(o))} T(Wy^*, v_i)_X\, v_i$$

and for $t \geqq 0$,

$$y(t) = \beta + t u_{j(o)}\,.$$

We put

$$y' = \sum_{i\in A(j(o))} (y^*, u_i)_Y\, u_i - \operatorname{sgn}\{d_{j(o)}((Wy^*, v_{j(o)})_X - (y^*, u_{j(o)})_X\}\, \Delta_{j(o)} u_{j(o)}\,,$$

where we use the notation sgn 0 = 1.

Obviously $y' \in \mathfrak{R}_{j(o)} \subset \mathfrak{R}_j$. We choose $t_0 \geqq 0$ such that

$$(2.10)\qquad \|Wy^* - T^+\, y(t_0)\|_X = \|Wy^* - T^+ y'\|_X\,.$$

By (2.10) we obtain

$$(2.11)\qquad t_0^2 (d^+_{j(o)})^2 = \sum_{i\in A(j(o))} \|(Wy^*, v_i)_X\, v_i - d^+_i (y', u_i)_Y\, v_i\|^2_X\,.$$

Evidently $d_{j(o)} > 0$. By (2.11),

$$t_o^2 = \sum_{i\in A(j(o))} d^2_{j(o)} d_i^{-2} |d_i(Wy^*, v_i)_X - (y', u_i)_Y|^2 \geqq$$
$$\geqq |d_{j(o)}(Wy^*, v_{j(o)})_X - (y', u_{j(o)})_Y|^2 \geqq \Delta^2_{j(o)}$$

and therefore

$$(2.12)\qquad t_o^2 \geqq \Delta^2_{j(o)}\,.$$

By (2.12) and (2.10) we have

$$(2.13)\qquad \|Wy^* - T^+ y'\|^2_X \geqq (d^+_{j(o)})^2\, \Delta^2_{j(o)} + \sum_{i\in B(j(o))} |(Wy^*, v_i)_X|^2\,.$$

By (2.12) and (2.13) we obtain

$$(2.14)\qquad \omega^2(T^j, \mathfrak{R}_j) \leqq \Delta^2_{j(o)} (d^+_{j(o)})^2 + \sum_{i\in B(j(o))} |(Wy^*, v_i)_X|^2 \leqq \omega^2(W, \mathfrak{R}_j)\,,$$

because $y' \in \mathfrak{N}_j$. Then it is easy to verify

$$\omega(T^j, \mathfrak{N}_j) = \inf\{\omega(W, \mathfrak{N}_j) : W \in [Y, X]\}\,, \tag{2.15}$$

which completes the proof.

Corollary 2.1. *Let $j, k \in P$ be such that $j \leqq k$. Then $\Omega(\mathfrak{N}_j) \leqq \Omega(\mathfrak{N}_k)$.*

Proof. We have $d^+_{j(o)}\, \Delta_{j(o)} = d^+_{k(o)}\, \Delta_{k(o)}$ because $j(o) \leqq k(o)$, where $j(o)$ (and similarly $k(o)$) satisfies

(i) $d^+_{j(o)} = \max\{d^+_i : i \in A(j) \setminus A(k(\Delta) - 1)\}$;

(ii) if $d^+_p = d^+_{j(o)}$ for some $p \in A(j) \setminus A(k(\Delta) - 1)$ then $p \leqq j(o)$.

Thus, the validity of the relation $\Omega(\mathfrak{N}_j) \leqq \Omega(\mathfrak{N}_k)$ is a consequence of Theorem 2.2.

Theorem 2.3. *The element $\hat{x} = T^{k(\Delta)}y^*$ is a universal approximation to the normal solution x_0.*

Proof. With respect to the above results and to (2.15) it is enough to show that there exists a constant $d \geqq 1$ independent of $j \in I$ such that $\omega(T^{k(\Delta)}, \mathfrak{N}_j) \leqq d\, \Omega(\mathfrak{N}_j)$. Since

$$\omega(T^{k(\Delta)}, \mathfrak{N}_j) \leqq \|T^{k(\Delta)}y^* - T^j y^*\|_X + \sup\{\|T^j y^* - T^+ y\|_X : y \in \mathfrak{N}_j\}\,, \tag{2.16}$$

we obtain by (2.6), (2.7) that

$$\omega(T^{k(;\Delta)}, \mathfrak{N}_j) \leqq 2d^+_{j(o)}\, \Delta_{j(o)}\,.$$

References

[1] *T. Kamo:* Теория возмущений линейных операторов изд. Мир, Москва 1972

[2] *R. J. Hanson:* A numerical method for solving Fredholm integral equations of the first kind using singular values, Siam J. Numer. Anal., Vol. 8, 1970, p. 616—622.

[3] *B. A. Морозов:* Линейные и нелинейные некорректные задачи, Математический анализ, том 11, Итоги науки и техники, Москва 1973, стр. 129—178

Souhrn

UNIVERZALITY METODY NEJLÉPE URČENÝCH TERMŮ

JIŘÍ NEUBERG

Jsou studovány vlastnosti metody nejlépe určených termů vzhledem k jednomu apriornímu rozkladu $R(T)$ s cílem určit univerzální aproximaci normálního řešení Fredholmových integrálních rovnic prvního druhu.

Author's address: Dr. *Jiří Neuberg,* Oblastní výpočetní středisko vysokých škol, Zikova 4, 160 00 Praha 6.