A priori inverse operator estimation for guaranteed error estimate

Akitoshi Takayasu[1,3)], Shin'ichi Oishi[2,3)] and Takayuki Kubo[4)]
[1)] *Graduate School of Fundamental Science and Engineering, Waseda University,*
[2)] *Faculty of Science and Engineering, Waseda University*
[3)] *JST CREST*
[4)] *Institute of Mathematics, University of Tsukuba*
takitoshi@suou.waseda.jp

Abstract. A guaranteed error estimate procedure for linear or nonlinear two-point boundary value problems is established by authors. 'Guaranteed' error estimate is rigorous, i.e. it takes into account every error such as the discretization error and the rounding error when we compute an approximate solution. We can also prove the existence and the uniqueness of the exact solution. Namely, we can solve the problem with mathematically rigorous. In order to bound the guaranteed error, an inverse operator norm estimation is needed. In the previous work of authors, the inverse operator estimate needs the norm of the inverse matrix, which is a posteriori constant. So we need much time to compute the inverse operator norm estimation. In this paper, we propose a priori estimation concerning the inverse operator. The proposed estimate is given without the norm of the inverse matrix. By using a priori estimation, we can estimate the inverse operator as a priori constant. An extremely improvement of the computational speed is expected. On the other hand, it needs a condition to use a priori estimation. Finally, we present some numerical results.

Keywords: A priori estimates; Guaranteed error estimate; Numerical verification.

1. Introduction

Authors have established a guaranteed error estimate method for two-point boundary value problems in (Takayasu, Oishi and Kubo, 2009a; Takayasu, Oishi and Kubo, 2009b). Proposed method is based on the theorem with respect to the inverse operator norm estimation of a class of linear operators. This theorem has proved in 1994 on an unpublished paper by the one of authors (S.Oishi) and presented in his book (Oishi, 2000). This paper is concerned with a priori inverse operator norm estimation in the verified computation method. In the theorem, the inverse operator norm estimation needs the norm of the inverse matrix. It is a posteriori constant. So we need much time to compute the inverse operator norm estimation. It becomes the bottleneck of the computation speed. We propose a priori estimate concerning the inverse operator norm estimation. It is given without the norm of the inverse matrix. It is expected that a priori estimation induce an extremely improvement of the computational speed.

'Guaranteed' error estimate method is a method of computer assisted proof for the existence of solutions to differential equations. The method enables us to prove the existence and the uniqueness of the exact solution. The guaranteed error is rigorous, i.e. it includes all computational errors such

4th International Workshop on Reliable Engineering Computing (REC 2010)
Edited by Michael Beer, Rafi L. Muhanna and Robert L. Mullen

ISBN: 978-981-08-5118-7. *Published by* Research Publishing Services.
doi:10.3850/978-981-08-5118-7_054

as the discretization error and the rounding error. The purpose of our verified computation method is to get the guaranteed error between the exact solution u and a computational result $\hat{u}$ in the form of

$$\|u - \hat{u}\|_X \leq Const.$$

Here, X is a suitable functional space and $Const$ is a computable constant. Our method is based on the computation of some constants in order to obtain the enclosure property of a base theorem. For example, the enclosure property is the convergence condition in the Newton-Kantorovich theorem or the contraction property of some operator. The feature of our established method is pure analytic and its proof is easily understandable.

Studies on two-point boundary value problems by a computer assisted proof have been developed. Generally speaking, the differential equations are equivalent to the simultaneous equations with infinite dimension. Hence it is the important point to analyze the error caused by the truncation or the discretization of the original problem. There are some computer assisted proof methods in (Kedem, 1981) and (Plum, 1991) for example. Especially, M.T. Nakao has presented an another computer assisted proof method for two decades in (Nakao, 1988). It is shown that this method is useful to bound the tight numerical inclusion of solutions. The method is based on a kind of interval FEM. It was also described as the verification procedure for two-point boundary value problems in (Nakao, 1992). The enclosure methods in functional spaces are classified into an analytic method, an interval method and a mixed method. Our verified computation method is one of the analytic method.

In this paper, we first introduce the inverse operator norm estimation in Theorem 1. A proof of the theorem is given. It needs an estimation of a finite dimensional operator. A priori inverse operator estimate is proposed in Section 3. Instead of the finite dimensional operator, the infinite dimensional operator is used in the proof. In the next section, we briefly sketch some verified computation algorithms. The method for linear problems is explained in Section 4. A solution of nonlinear problems is guaranteed by our verification algorithm in Section 5. Finally, we present some numerical results. The effectiveness of a priori estimation is confirmed.

2. Previous estimation

Let X be a Banach space and X_n be an approximate finite dimensional subspace of X. Assuming that $\mathcal{P}_n$ denotes a projection. In the previous work, we use the following theorem for the inverse operator norm estimation. This is a fundamental theorem proved by the one of authors (S.Oishi).

Theorem 1 (Oishi 1994 (unpublished preprint), Oishi 2000 (book)). *Let $\mathcal{K} : X \to X$ be a compact operator and $\mathcal{P}_n : X \to X_n$ be the projection operator on $X_n \subset X$. We assume that $\mathcal{P}_n\mathcal{K}$ is bounded by*

$$\|\mathcal{P}_n\mathcal{K}\|_{\mathcal{L}(X,X_n)} \leq K,$$

the difference between $\mathcal{K}$ and $\mathcal{P}_n\mathcal{K}$ is bounded by

$$\|\mathcal{K} - \mathcal{P}_n\mathcal{K}\|_{\mathcal{L}(X,X)} \leq L,$$

and the finite dimensional operator $(I - \mathcal{P}_n\mathcal{K})|_{X_n} : X_n \to X_n$ *is invertible with*

$$\|(I - \mathcal{P}_n\mathcal{K})|_{X_n}^{-1}\|_{\mathcal{L}(X_n,X_n)} \le M.$$

If three assumptions are obtained and $(1 + MK)L < 1$*, then the infinite dimensional operator* $(I - \mathcal{K}) : X \to X$ *is invertible and*

$$\|(I - \mathcal{K})^{-1}\|_{\mathcal{L}(X,X)} \le \frac{1 + MK}{1 - (1 + MK)L}. \tag{1}$$

It shows the existence and the uniqueness of the inverse operator. It is based on the Fredholm alternative theorem and the Riesz-Schauder theorem in appendix. We also get the inverse operator estimate with respect to the operator norm. We shall give the proof of Theorem 1.

Proof. Assume that there exists the finite dimensional operator $(I - \mathcal{P}_n\mathcal{K})|_{X_n}^{-1} : X_n \to X_n$. We can show that the inverse operator of $(I - \mathcal{P}_n\mathcal{K}) : X \to X$ is

$$(I - \mathcal{P}_n\mathcal{K})^{-1}\phi = (I + (I - \mathcal{P}_n\mathcal{K})|_{X_n}^{-1}\mathcal{P}_n\mathcal{K})\phi, \quad (\phi \in X) \tag{2}$$

by direct computations. Multiplying the $(I - \mathcal{P}_n\mathcal{K})$ from the both side of (2), we check that these become the identity operator $I : X \to X$. In fact, we have

$$\begin{aligned}
&(I + (I - \mathcal{P}_n\mathcal{K})|_{X_n}^{-1}\mathcal{P}_n\mathcal{K})(I - \mathcal{P}_n\mathcal{K}) \\
=\ & I - \mathcal{P}_n\mathcal{K} + (I - \mathcal{P}_n\mathcal{K})|_{X_n}^{-1}\mathcal{P}_n\mathcal{K} - (I - \mathcal{P}_n\mathcal{K})|_{X_n}^{-1}\mathcal{P}_n\mathcal{K}\mathcal{P}_n\mathcal{K} \\
=\ & I - \mathcal{P}_n\mathcal{K} + (I - \mathcal{P}_n\mathcal{K})|_{X_n}^{-1}(I - \mathcal{P}_n\mathcal{K} + \mathcal{P}_n\mathcal{K})|_{X_n}\mathcal{P}_n\mathcal{K} - (I - \mathcal{P}_n\mathcal{K})|_{X_n}^{-1}\mathcal{P}_n\mathcal{K}\mathcal{P}_n\mathcal{K} \\
=\ & I - \mathcal{P}_n\mathcal{K} + (I - \mathcal{P}_n\mathcal{K})|_{X_n}^{-1}(I - \mathcal{P}_n\mathcal{K})|_{X_n}\mathcal{P}_n\mathcal{K} + (I - \mathcal{P}_n\mathcal{K})|_{X_n}^{-1}\mathcal{P}_n\mathcal{K}\mathcal{P}_n\mathcal{K} - (I - \mathcal{P}_n\mathcal{K})|_{X_n}^{-1}\mathcal{P}_n\mathcal{K}\mathcal{P}_n\mathcal{K} \\
=\ & I - \mathcal{P}_n\mathcal{K} + \mathcal{P}_n\mathcal{K} \\
=\ & I.
\end{aligned}$$

Furthermore, since $X_n \subset X$

$$\begin{aligned}
(I - \mathcal{P}_n\mathcal{K})(I + (I - \mathcal{P}_n\mathcal{K})|_{X_n}^{-1}\mathcal{P}_n\mathcal{K}) &= I - \mathcal{P}_n\mathcal{K} + (I - \mathcal{P}_n\mathcal{K})(I - \mathcal{P}_n\mathcal{K})|_{X_n}^{-1}\mathcal{P}_n\mathcal{K} \\
&= I - \mathcal{P}_n\mathcal{K} + (I - \mathcal{P}_n\mathcal{K})|_{X_n}(I - \mathcal{P}_n\mathcal{K})|_{X_n}^{-1}\mathcal{P}_n\mathcal{K} \\
&= I - \mathcal{P}_n\mathcal{K} + \mathcal{P}_n\mathcal{K} \\
&= I.
\end{aligned}$$

Therefore, $(I - \mathcal{P}_n\mathcal{K})$ has the inverse operator as (2). Assumptions imply that the inverse operator of $(I - \mathcal{P}_n\mathcal{K})$ is bounded by

$$\begin{aligned}
\|(I - \mathcal{P}_n\mathcal{K})^{-1}\|_{\mathcal{L}(X,X)} &= \|I + (I - \mathcal{P}_n\mathcal{K})|_{X_n}^{-1}\mathcal{P}_n\mathcal{K}\|_{\mathcal{L}(X,X)} \\
&\le 1 + \|(I - \mathcal{P}_n\mathcal{K})|_{X_n}^{-1}\|_{\mathcal{L}(X_n,X_n)}\|\mathcal{P}_n\mathcal{K}\|_{\mathcal{L}(X,X_n)} \\
&\le 1 + MK,
\end{aligned}$$

where implicitly $\|I\|_{\mathcal{L}(X_n,X)} \le 1$ was used. Moreover, for $(I-\mathcal{K})\phi = \psi$ and using

$$\phi = \phi - (I-\mathcal{P}_n\mathcal{K})^{-1}(I-\mathcal{K})\phi + (I-\mathcal{P}_n\mathcal{K})^{-1}\psi,$$

we have

$$\begin{aligned}
\|\phi\|_X &\le \|I-(I-\mathcal{P}_n\mathcal{K})^{-1}(I-\mathcal{K})\|_{\mathcal{L}(X,X)}\|\phi\|_X + \|(I-\mathcal{P}_n\mathcal{K})^{-1}\psi\|_X \\
&\le \|(I-\mathcal{P}_n\mathcal{K})^{-1}\|_{\mathcal{L}(X,X)}\|\mathcal{K}-\mathcal{P}_n\mathcal{K}\|_{\mathcal{L}(X,X)}\|\phi\|_X + \|(I-\mathcal{P}_n\mathcal{K})^{-1}\|_{\mathcal{L}(X,X)}\|\psi\|_X \\
&\le (1+MK)L\|\phi\|_X + (1+MK)\|\psi\|_X.
\end{aligned}$$

If $(1+MK)L<1$, then we obtain

$$\|\phi\|_X \le \frac{1+MK}{1-(1+MK)L}\|(I-\mathcal{K})\phi\|_X. \tag{3}$$

From (3), $\phi=0$ holds if and only if $(I-\mathcal{K})\phi = 0$. This implies the operator $(I-\mathcal{K}):X\to X$ is injective. Accordingly, by the Fredholm alternative theorem (Theorem 3 in appendix) and the Riesz-Schauder theorem (Theorem 4 in appendix), it has a bounded inverse $(I-\mathcal{K})^{-1}$, which satisfies $\phi = (I-\mathcal{K})^{-1}\psi$. Hence, we have

$$\|(I-\mathcal{K})^{-1}\|_{\mathcal{L}(X,X)} := \sup_{\psi\neq 0}\frac{\|(I-\mathcal{K})^{-1}\psi\|_X}{\|\psi\|_X} \le \frac{1+MK}{1-(1+MK)L}.$$

The inverse operator norm estimation is given by (1). □

Theorem 1 uses the finite dimensional operator to estimate the infinite dimensional operator $I-\mathcal{P}_n\mathcal{K}:X\to X$ in (2). However the inverse norm estimation of the finite operator is equivalent to the norm of the inverse matrix. It costs the $O(n^3)$ computation. It takes a substantial amount of time in this part. We want to get round the weak point of our verification method. In the next section, a priori estimate of the inverse operator is proposed. Under a certain condition, we can get the inverse operator norm estimation without calculating the norm of the inverse matrix. The inverse operator estimate is given as a priori estimate.

3. A priori inverse operator estimate

The bottleneck of the previous estimation is the estimation of the inverse finite dimensional operator in (2). In this section, we have a priori norm estimation regarding the inverse operator. Theorem 1 is modified by the one of authors (A. Takayasu). In the part of (2), we use an infinite dimensional operator itself instead of the finite operator. Then, we have the following proposition.

Proposition 1 (Takayasu). *Let $\mathcal{K}:X\to X$ be a compact operator and $\mathcal{P}_n:X\to X_n$ be the projection operator on $X_n\subset X$, an approximate finite dimensional subspace of X. Assuming that $\mathcal{P}_n\mathcal{K}$ is bounded by*

$$\|\mathcal{P}_n\mathcal{K}\|_{\mathcal{L}(X,X_n)} \le K,$$

and the difference between $\mathcal{K}$ *and* $\mathcal{P}_n\mathcal{K}$ *is bounded by*

$$\|\mathcal{K} - \mathcal{P}_n\mathcal{K}\|_{\mathcal{L}(X,X)} \leq L.$$

If these assumptions are obtained and satisfy $K + L < 1$*, then the infinite dimensional operator* $(I - \mathcal{K}) : X \to X$ *is invertible and*

$$\|(I - \mathcal{K})^{-1}\|_{\mathcal{L}(X,X)} \leq \frac{1}{1 - K - L}.$$

The proof follows the proof of Theorem 1. Meanwhile, we don't use the finite dimensional operator. The constant M in Theorem 1 is unnecessary in a priori estimate.

Proof. Assume that the inverse operator of $(I - \mathcal{P}_n\mathcal{K}) : X \to X$ is denoted as

$$(I - \mathcal{P}_n\mathcal{K})^{-1}\phi = (I + (I - \mathcal{P}_n\mathcal{K})^{-1}\mathcal{P}_n\mathcal{K})\phi, \quad (\phi \in X).$$

The assumptions imply that the inverse operator is bounded by

$$\|(I - \mathcal{P}_n\mathcal{K})^{-1}\|_{\mathcal{L}(X,X)} \leq 1 + \|(I - \mathcal{P}_n\mathcal{K})^{-1}\|_{\mathcal{L}(X,X)}\|I\|_{\mathcal{L}(X_n,X)}\|\mathcal{P}_n\mathcal{K}\|_{\mathcal{L}(X,X_n)}.$$

Implicitly $\|I\|_{\mathcal{L}(X_n,X)} \leq 1$, then we have the norm estimation of the inverse operator for $K < 1$

$$\|(I - \mathcal{P}_n\mathcal{K})^{-1}\|_{\mathcal{L}(X,X)} \leq \frac{1}{1 - K}.$$

Moreover, for $(I - \mathcal{K})\phi = \psi$ and using

$$\phi = \phi - (I - \mathcal{P}_n\mathcal{K})^{-1}(I - \mathcal{K})\phi + (I - \mathcal{P}_n\mathcal{K})^{-1}\psi.$$

we have

$$\begin{aligned}
\|\phi\|_X &\leq \|I - (I - \mathcal{P}_n\mathcal{K})^{-1}(I - \mathcal{K})\|_{\mathcal{L}(X,X)}\|\phi\|_X + \|(I - \mathcal{P}_n\mathcal{K})^{-1}\psi\|_X \\
&\leq \|(I - \mathcal{P}_n\mathcal{K})^{-1}\|_{\mathcal{L}(X,X)}\|\mathcal{K} - \mathcal{P}_n\mathcal{K}\|_{\mathcal{L}(X,X)}\|\phi\|_X + \|(I - \mathcal{P}_n\mathcal{K})^{-1}\|_{\mathcal{L}(X,X)}\|\psi\|_X \\
&\leq \frac{L}{1 - K}\|\phi\|_X + \frac{1}{1 - K}\|\psi\|_X.
\end{aligned}$$

If $K + L < 1$, then we obtain

$$\|\phi\|_X \leq \frac{1}{1 - K - L}\|(I - \mathcal{K})\phi\|_X.$$

This implies that $(I - \mathcal{K}) : X \to X$ is injective, and by the Fredholm alternative theorem (Theorem 3 in appendix) and the Riesz-Schauder theorem (Theorem 4 in appendix) it has a bounded inverse satisfying $\phi = (I - \mathcal{K})^{-1}\psi$. Hence, it follows

$$\|(I - \mathcal{K})^{-1}\|_{\mathcal{L}(X,X)} := \sup_{\psi \neq 0} \frac{\|(I - \mathcal{K})^{-1}\psi\|_X}{\|\psi\|_X} \leq \frac{1}{1 - K - L}.$$

4. Linear problem

The guaranteed error estimation is given by a verified computation. In this section, we consider the following linear two-point boundary value problems of the form

$$\begin{cases} -u'' = ru + f & 0 < x < 1, \\ u(0) = u(1) = 0 \end{cases} \tag{4}$$

where $r \in L^\infty([0,1])$, which denotes the space of all functions that are essentially bounded, and $f \in L^2([0,1])$. Here, we follow (Takayasu, Oishi and Kubo, 2009a). We shall introduce the verification method. The linear problem uses the inverse operator norm estimation directly. The finite element method is used in order to get an approximate solution. The verification method transforms the problem into an operator equation with a solution operator to the Poisson equation

$$\begin{cases} -u'' = f & 0 < x < 1, \\ u(0) = u(1) = 0. \end{cases} \tag{5}$$

A solution operator leads the solution $u = \mathcal{K}f$ of Poinsson equation (5). Let $X = L^2([0,1])$ denote the L^2 functional space with L^2-inner product, and let $H^m([0,1])$ denote the L^2-Sobolev space of order m. We define

$$D := H_0^1([0,1]) = \{u \in H^1 : u(0) = u(1) = 0\}$$

with the inner product (u', v') and norm $\|u\|_D = \|u'\|_X$. The solution operator $\mathcal{K} \in \mathcal{L}(X, D)$ is linear and bounded, so $\mathcal{K} : X \to H^2 \cap H_0^1 \hookrightarrow D$ is compact by the Sobolev embedding theorem (for example, see (Adams, 1975)). The original problem is transformed into the operator equation

$$(4) \iff (I - \mathcal{K}r)u = \mathcal{K}f.$$

Since the solution operator is compact operator, it can be obtained that the operator $(I - \mathcal{K}r)$ has its inverse operator by Theorem 1 or Proposition 1. Then we have the existence and the uniqueness of the exact solution to (4). A norm estimation of the inverse operator is also given. We arrive at the guaranteed error bounds

$$\begin{aligned} \|u - \hat{u}\|_X &= \|(I - \mathcal{K}r)^{-1}(I - \mathcal{K}r)(u - \hat{u})\|_X \\ &\le \|(I - \mathcal{K}r)^{-1}\|_{\mathcal{L}(X,X)} \|\mathcal{K}f - (I - \mathcal{K}r)\hat{u}\|_X. \end{aligned} \tag{6}$$

There are no assumptions on the quality of $\hat{u}$. Two constants are needed to bound the guaranteed error. One is the inverse operator norm estimation

$$\|(I - \mathcal{K}r)^{-1}\|_{\mathcal{L}(X,X)} \le C'. \tag{7}$$

The other is the residual of the operator equation

$$\|\mathcal{K}f - (I - \mathcal{K}r)\hat{u}\|_X \le C. \tag{8}$$

In order to approximate $\mathcal{K}$, let $x_i := ih$, $(i = 1, ..., n)$ with $h := 1/(n+1)$ be an equidistant partition of the interval $[0,1]$. Let S_h denote a class of finite element basis functions, we define a discrete functional space

$$X_n = \operatorname{span}\{\phi_{h_1}, \phi_{h_2}, ..., \phi_{h_n}\}, \quad \phi_{h_i} \in S_h$$

equipped with the norm

$$\|u_h\|_{X_n} := \sqrt{\frac{1}{n}\sum_{j=1}^{n} u_j^2},$$

for $u_h = \sum_{j=1}^{n} u_j \phi_{h_j}$. A discrete projection $\mathcal{P}_n : D \to X_n$ is defined by

$$(u' - (\mathcal{P}_n u)', v) = 0 \quad \text{for } v \in S_h.$$

4.1. Inverse operator estimate

The constant C' in (7) is bounded by Proposition 1. In the established procedure, the inverse operator estimate is based on the computation of constants K and L. We first compute the constant K. It is theoretically known that the eigenvalue problem

$$\begin{cases} -u'' = \lambda u & 0 < x < 1, \\ u(0) = u(1) = 0 \end{cases}$$

has eigenvalues $\lambda = n^2\pi^2$, so the minimal eigenvalue is $\lambda_{min} = \pi^2$. By using Rayleigh quotient,

$$\lambda_{min}(u, u) \leq (u', u') = (-u'', u) \leq \|u''\|_X \|u\|_X,$$

hence $\|u\|_X \leq \frac{1}{\pi^2}\|u''\|_X$ and therefore

$$(u', u') \leq \frac{1}{\pi^2}\|u''\|_X^2.$$

It follows

$$\|\mathcal{P}_n\mathcal{K}\|_{\mathcal{L}(X,X_n)} \leq \|\mathcal{K}\|_{\mathcal{L}(X,D)} = \sup_{f\in X}\frac{\|\mathcal{K}f\|_D}{\|f\|_X} = \sup_{f\in X}\frac{\|u'\|_X}{\|f\|_X} \leq \sup_{f\in X}\frac{\pi^{-1}\|u''\|_X}{\|f\|_X} = \frac{1}{\pi} =: K.$$

For later use, we have

$$\|\mathcal{K}\|_{\mathcal{L}(X,X)} = \sup_{f\in X}\frac{\|\mathcal{K}f\|_X}{\|f\|_X} \leq \sup_{f\in X}\frac{\frac{1}{\pi^2}\|u''\|_X}{\|f\|_X} \leq \frac{1}{\pi^2}.$$

Secondly, the constant L is bounded by the Aubin-Nitsche's trick to estimate the error of the FEM. We obtain

$$\|u - \mathcal{P}_n u\|_X \leq \frac{h^2}{\pi^2}\|u''\|_X,$$

and therefore

$$\|\mathcal{K} - \mathcal{P}_n\mathcal{K}\|_{\mathcal{L}(X,X)} = \sup_{f\in X}\frac{\|\mathcal{K}f - \mathcal{P}_n\mathcal{K}f\|_X}{\|f\|_X} = \sup_{f\in X}\frac{\|u - \mathcal{P}_n u\|_X}{\|f\|_X} \leq \frac{h^2}{\pi^2} =: L.$$

Since $\mathcal{K}$ is a compact operator and $r \in L^\infty([0,1])$, the operator $\mathcal{K}r$ is compact as well. Thus constants K' and L' can be induced such that

$$\|\mathcal{P}_n\mathcal{K}r\|_{\mathcal{L}(X,X_n)} \le K\|r\|_\infty =: K',$$

and

$$\|\mathcal{K}r - \mathcal{P}_n\mathcal{K}r\|_{\mathcal{L}(X,X)} \le L\|r\|_\infty =: L'.$$

Finally, the constant C' in (7) is calculated by Proposition 1. If $K' + L' < 1$, then

$$C' := \frac{1}{1 - K' - L'}.$$

In case of $K' + L' \not< 1$, the constant C' needs to be bounded by Theorem 1. Thus, we compute the constant M' concerning the inverse of the finite dimensional operator. In (Takayasu, Oishi and Kubo, 2009a), the discrete solution operator is denoted as the matrix form. By using the inverse matrix norm estimation, the computable constant M' is yielded by a matrix norm. However, it is expected that it takes much time to compute the constant M'. In this case, the constant C' in (7) is calculated by

$$C' := \frac{1 + M'K'}{1 - (1 + M'K')L'}.$$

4.2. Residual of the operator equation

For the constant C in (8), the residual of the operator equation follows

$$\begin{aligned}
&\|\mathcal{K}f - (I - \mathcal{K}r)\hat{u}\|_X \\
=\ &\|\mathcal{K}f - \mathcal{P}_n\mathcal{K}f + \mathcal{P}_n\mathcal{K}f - \mathcal{P}_n\mathcal{K}f_h + \mathcal{P}_n\mathcal{K}f_h - (I - \mathcal{P}_n\mathcal{K}r)\hat{u} + \mathcal{K}r\hat{u} - \mathcal{P}_n\mathcal{K}r\hat{u}\|_X \\
\le\ &\|\mathcal{K} - \mathcal{P}_n\mathcal{K}\|_{\mathcal{L}(X,X)}\|f\|_X + \|\mathcal{P}_n\mathcal{K}\|_{\mathcal{L}(X,X_n)}\|f - f_h\|_X + \|Res\|_X + \|\mathcal{K} - \mathcal{P}_n\mathcal{K}\|_{\mathcal{L}(X,X)}\|r\|_\infty\|\hat{u}\|_X \\
\le\ &\|Res\|_X + K\|f - f_h\|_X + L(\|f\|_X + \|r\|_\infty\|\hat{u}\|_X) \\
=:\ &C.
\end{aligned}$$

Here, K and L are constants of assumptions in Theorem 1. $\|Res\|_X$ is the residual of the finite linear system. The norm $\|f - f_h\|_X$ can be estimated by the interpolation theory. $\|\hat{u}\|_X$ is the L^2-norm of an approximate solution $\hat{u}$. Then, the residual of the operator equation is given by the standard norm estimation.

By Proposition 1 or Theorem 1, $K_1 + L_1 < 1$ or $(1 + M'K')L < 1$ implies that there exists an unique solution u satisfying the linear problem (4) in the guaranteed error (6).

Algorithm 1 (Linear Two-point bounary value problem)**.** *We present the verification algorithm as follows:*

1. *Get an approximate solution by FEM.*
2. *Inverse operator norm estimation (Constant C') satifying (7).*

- *Compute the constants* K' *and* L'.
- *Check* $K' + L' < 1$.
- *If not compute* M'.
- *Check* $(1 + M'K')L' < 1$.

3. *The residual of the operator equation (Constant* C*).*
4. *Guaranteed error estimate is bounded by*

$$\|u - \hat{u}\|_X \le C'C.$$

5. Nonlinear problem

Verified computation for a nonlinear problem is different from the one for a linear problem. In this section, we introduce a verification method of a nonlinear problem by using a priori inverse norm estimation. In particular we are here concerned with a nonlinear two-point boundary value problems of the form

$$\begin{cases} -u'' = ru^N + f & 0 < x < 1, \\ u(0) = u(1) = 0, \end{cases} \tag{9}$$

for $N \in \mathbb{N}$, $r \in L^\infty([0,1])$ and $f \in L^2([0,1])$. Here, we assume $N \ge 2$ and follow (Takayasu, Oishi and Kubo, 2009b). In the verification method, we also transform the problem into a nonlinear operator equation with the solution operator, which is defined in the former section. We have $u = \mathcal{K}(ru^N + f)$, or equivalently the nonlinear operator equation

$$F(u) = u - \mathcal{K}ru^N - \mathcal{K}f = 0. \tag{10}$$

Based on an approximate finite element solution $\hat{u}$, the Newton-Kantorovich Theorem is applicable. We note that, there are no assumptions on the quality of $\hat{u}$.

Theorem 2 (Newton-Kantorovich Theorem)**.** *Let* F *be given in (10), and assume the Fréchet derivative* $F'(u)$ *is nonsingular and satisfies*

$$\|F'(\hat{u})^{-1}F(\hat{u})\|_X \le \alpha,$$

for a certain positive α*. Furthermore, assuming that*

$$\|F'(\hat{u})^{-1}(F'(v) - F'(w))\|_X \le \omega\|v - w\|_X$$

for a certain positive ω *and for any* $v, w \in B(\hat{u}, \delta) := \{v \in X : \|v - \hat{u}\| \le \delta\} \subset X$*, which is a closed ball in* X *centered at* $\hat{u}$ *with the radius* δ*. If*

$$\alpha\omega \le \frac{1}{2}, \tag{11}$$

then F *has a unique solution* u *satisfying*

$$\|u - \hat{u}\|_X \le \rho := \frac{1 - \sqrt{1 - 2\alpha\omega}}{\omega}. \tag{12}$$

Through the Newton-Kantorovich theorem, we can show the existence and the uniqueness of the exact solution to (9). Additionally, the guaranteed error is bounded. The verification method is based on the computation of three constants C_1, C_2 and C_3 with

$$\|F'(\hat{u})^{-1}\|_{\mathcal{L}(X,X)} = \|(I - \mathcal{K}Nr\hat{u}^{N-1})^{-1}\|_{\mathcal{L}(X,X)} \leq C_1,$$

$$\|F(\hat{u})\|_X = \|\hat{u} - \mathcal{K}r\hat{u}^N - \mathcal{K}f\|_X \leq C_2$$

and

$$\begin{aligned}\|F'(v) - F'(w)\|_X &= \|\mathcal{K}Nr(v^{N-2} + v^{N-3}w + ... + vw^{N-3} + w^{N-2})(v - w)\|_X \\ &\leq C_3\|v - w\|_X.\end{aligned}$$

If the condition (11) is satisfied for $\alpha := C_1C_2$ and $\omega := C_1C_3$, then (12) is obtained.

5.1. Some constants

There are three parts of the guaranteed error estimation for nonlinear problems. The inverse operator norm estimation, the residual of the operator equation and the Lipschitz constant of the Fréchet derivative $F'(u)$ are needed. The inverse operator norm estimation is obtained by Proposition 1. Since $\mathcal{K}$ is a compact operator, $r \in L^\infty([0,1])$ and $\hat{u} \in X_n$, the operator $\mathcal{K}Nr\hat{u}^{N-1}$ is compact as well. Then constants K_1 and L_1 can be computed such that

$$\|\mathcal{P}_n\mathcal{K}Nr\hat{u}^{N-1}\|_{\mathcal{L}(X,X_n)} \leq KN\|r\|_\infty\|\hat{u}^{N-1}\|_X =: K_1,$$

and

$$\|\mathcal{K}Nr\hat{u}^{N-1} - \mathcal{P}_n\mathcal{K}Nr\hat{u}^{N-1}\|_{\mathcal{L}(X,X)} \leq LN\|r\|_\infty\|\hat{u}^{N-1}\|_X =: L_1.$$

Therefore the constant C_1 is given by Proposition 1.

$$C_1 := \frac{1}{1 - K_1 - L_1}.$$

When $K_1 + L_1 \not< 1$, the constant C_1 requires to be bounded by Theorem 1. Thus, we compute the constant M_1 concerning the inverse of the finite dimensional operator. In (Takayasu, Oishi and Kubo, 2009b), the discrete solution operator is denoted as the matrix form. By using the inverse matrix estimation, the computable constant M_1 is yielded by a matrix norm. However, it is expected that it takes much time to compute the constant M_1. In this case, the constant C_1 is given by

$$C_1 := \frac{1 + M_1K_1}{1 - (1 + M_1K_1)L_1}.$$

For the constant C_2, the residual of the operator equation (10) follows

$$\begin{aligned}\|F(\hat{u})\|_X &= \|\hat{u} - \mathcal{K}r\hat{u}^N - \mathcal{K}f\|_X \\ &= \|\hat{u} - \mathcal{P}_n\mathcal{K}r\hat{u}^N - \mathcal{P}_n\mathcal{K}f_h - (\mathcal{K} - \mathcal{P}_n\mathcal{K})r\hat{u}^N - \mathcal{P}_n\mathcal{K}(f - f_h) - (\mathcal{K} - \mathcal{P}_n\mathcal{K})f\|_X \\ &\leq \|Res\|_X + K\|f - f_h\|_X + L(\|r\|_\infty\|\hat{u}^N\|_X + \|f\|_X) \\ &=: C_2,\end{aligned}$$

where $\|Res\|_X$ is the residual of the finite nonlinear system, and the norm $\|f - f_h\|_X$ can be estimated by the interpolation theory.

Finally, the constant C_3 is computable. For any $v, w \in B(\hat{u}, \delta)$, we have

$$\|v\|_X \leq \|\hat{u}\|_X + \delta, \quad \|w\|_X \leq \|\hat{u}\|_X + \delta.$$

Therefore

$$\begin{aligned}\|F'(v) - F'(w)\|_X &= \|\mathcal{K}Nr(v^{N-2} + v^{N-3}w + ... + vw^{N-3} + w^{N-2})(v - w)\|_X \\ &\leq N(N-1)\|\mathcal{K}\|_{\mathcal{L}(X,X)}\|r\|_\infty(\|\hat{u}\|_X + \delta)^{N-2}\|v - w\|_X.\end{aligned}$$

Accordingly, we may use

$$C_3 := \frac{N(N-1)}{\pi^2}\|r\|_\infty(\|\hat{u}\|_X + \delta)^{N-2}.$$

By Newton-Kantorovich theorem, $\alpha\omega = C_1^2 C_2 C_3 \leq \frac{1}{2}$ implies that there exists an unique solution $u \in \overline{B(\hat{u}, \rho)}$ satisfying the two-point boundary value problem (9).

Algorithm 2 (Nonlinear Two-point boudary value problem)**.** *The verification algorithm for a nonlinear problem is presented as follows.*

1. *Get an approximate solution by FEM.*
2. *Inverse operator norm estimation (Constant C_1).*
   - *Get K_1 and L_1.*
   - *Check $K_1 + L_1 < 1$.*
   - *If not compute M_1,*
   - *Check $(1 + M_1 K_1)L_1 < 1$.*
3. *The residual of the operator equation (Constant C_2).*
4. *Lipschitz constant (Constant C_3).*
5. *Set $\alpha = C_1 C_2,\ \omega = C_1 C_3$.*
6. *Check the condition $\alpha\omega \leq \frac{1}{2}$.*
7. *Guaranteed error estimate is bounded by*

$$\|u - \hat{u}\|_X \leq \frac{1 - \sqrt{1 - 2\alpha\omega}}{\omega}.$$

6. Computational results

For an application of our verification procedure, we present some computational results in this section. All computations are carried out on Intel Core2 Duo 1.86 GHz by using MATLAB 2009a with a toolbox for verified computations, INTLAB (Rump). IEEE 754 double precision, which means a relative precision is $2^{-53} \approx 10^{-16}$, is used. An approximate solution is led by the FEM. As a finite element subspace S_h, we choose one-dimensional piecewise hat functions.

6.1. Example 1

Consider the two-point boundary value problem of the form

$$\begin{cases} -u'' = cu + 1 & 0 < x < 1, \\ u(0) = u(1) = 0, \end{cases} \tag{13}$$

for $c = \pm 3.14, \pm 10$. Note that if $c = \pi^2 = 9.86...$, then this equation has no unique solution. An approximate solution $\hat{u}$ is obtained by a finite linear system with $n = 32$ grid points in Figure 1.

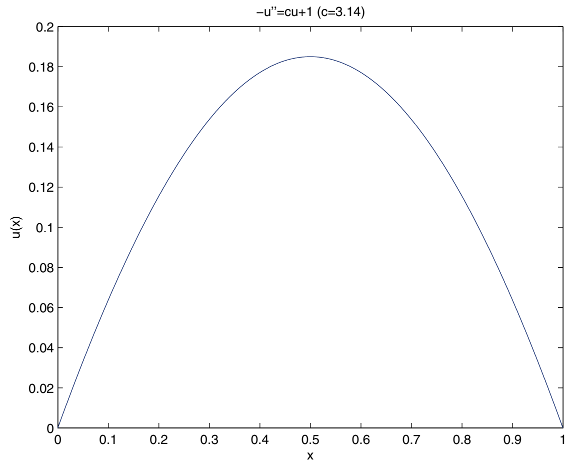

Figure 1. A shape of an approximate solution ($c = 3.14$).

The significant feature of a priori estimate is the computational time. In the previous mehtod, it took the $O(10) \sim O(10^3)$ times longer than the approximate computation. Whereas we need the inverse matrix estimation (dense), the finite linear system is sparse matrix equation. Then we have to compute much greater than the cost of the approximate calculation. By our a priori estimate the computational cost is extremely improved. It costs almost the same as the approximate time. The inverse matrix norm estimation is skipped. We present the ratio of computational costs in Table I. Let $t1$ in Table I be the computation time of the approximate calculation.

On the other hand, for $n = 32$, $c = 3.14$ and an approximate solution, our verification method compute the constant

$$K' + L' = 0.9998 < 1. \tag{14}$$

Table I. The ratio of computational costs $c = 3.14$.

Grid Points (2^x)	Approximate Solver ($/t1$)	Previous method ($/t1$)	New method ($/t1$)
5	1	16.7622	1.1747
6	1	18.5198	1.0974
7	1	24.9291	1.0537
8	1	48.4662	1.0312
9	1	92.1778	1.0191
10	1	188.8086	1.0118
11	1	394.5264	1.0107
12	1	943.6785	1.0066

Table II. The guaranteed error estimate 1

Grid Points (2^x)	$c = -10$	$c = -3.14$	$c = 3.14$	$c = 10$
5	8.4351133e-03	1.9271977e+00	2.2474712e+00	3.5905640e+01
6	2.1496523e-03	2.2029344e-01	2.5694899e-01	5.9783394e+00
7	5.4318474e-04	4.8493811e-02	5.6565425e-02	1.3739643e+00
8	1.3656193e-04	1.1771869e-02	1.3731406e-02	3.3941910e-01
9	3.4239070e-05	2.9217843e-03	3.4081520e-03	8.4945965e-02
10	8.5722734e-06	7.2913409e-04	8.5050826e-04	2.1274442e-02
11	2.1446433e-06	1.8220182e-04	2.1253190e-04	5.3257029e-03
12	5.3635880e-07	4.5545587e-05	5.3127492e-05	1.3343458e-03
13	-	1.1386556e-05	1.3282454e-05	-
14	-	2.8475816e-06	3.3224486e-06	-
15	-	7.1382513e-07	8.3444184e-07	-
16	-	1.8230625e-07	2.1599190e-07	-
17	-	5.3308325e-08	6.8793232e-08	-
18	-	2.8623837e-08	4.6608701e-08	-

Hence, there exists a unique solution of (13) and the guaranteed error is

$$\|u - \hat{u}\|_X \leq 2.2475.$$

The guaranteed error estimation is not tight because the condition (14) is almost failed. In fact, when $c = \pm 10$, we cannot get a priori estimation. In case of the linear problem, the coefficient function r in (4) satisfies $\|r\|_\infty < \pi$ then we can operate a priori estimation. If $\|r\|_\infty \approx \pi$, the constant C' in (7) gets too large. Though the inverse estimation looks the overestimation in this case, the error is reductive by increasing grid points in Table II.

Moreover, when $c = \pm 10$ in (13), we have the guaranteed error by the previous method. We cannot apply a priori estimate. Especially for $c = 10$ the problem is nearly singular. The guaranteed error is wider than another results. It is implied that the numerical result is sensitive. In general, the difficulty of the original problem has an effect on the guaranteed error bounds. Instead, we can improve the guaranteed error by increasing grid points.

6.2. Example 2

Let us consider the following nonlinear two-point boundary value problem

$$\begin{cases} -u'' = u^5 - \cos 2\pi x & 0 < x < 1, \\ u(0) = u(1) = 0. \end{cases} \tag{15}$$

An approximate solution $\hat{u}$ is obtained by some Newton iterates with $n = 32$ grid points. Our verification method yields the constant

$$\alpha\omega = 8.6659 \times 10^{-8} < \frac{1}{2}.$$

Therefore there exists a unique solution of (15) and we have the guaranteed error bounds

$$\|u - \hat{u}\| \leq 1.1547 \times 10^{-3}.$$

By established procedure we have the guaranteed inclusion in Figure 2. In the inclusion, there exists the exact solution between two curves based on the Newton-Kantorovich theorem.

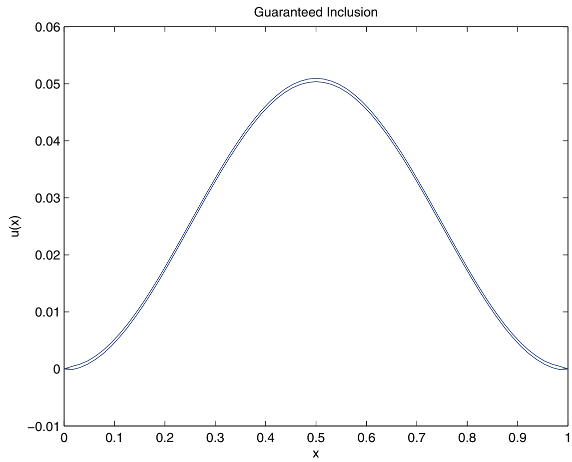

Figure 2. The guaranteed inclusion of the exact solution to (15).

Increasing the number of grid points, the guaranteed error bounds can be improved in Table III. As it happens, the guaranteed error in a priori estimate method is almost the same as the previous one. The improving ratio of the guaranteed error is dropped with the $O(1/n^2)$. It depends on the

Table III. The guaranteed error estimate 2.

Grid Points (2^x)	Previous method	New method
5	1.1546594e-03	1.1546595e-03
6	2.8866483e-04	2.8866484e-04
7	7.2166209e-05	7.2166211e-05
8	1.8041553e-05	1.8041554e-05
9	4.5103899e-06	4.5103901e-06
10	1.1275983e-06	1.1275984e-06
11	2.8190164e-07	2.8190165e-07
12	7.0476916e-08	7.0476917e-08

Table IV. The ratio of computational costs on example 2.

Grid Points (2^x)	Approximate Solver ($/t1$)	Previous method ($/t1$)	New method ($/t1$)
5	1	3.1199	1.0522
6	1	2.8691	1.0256
7	1	3.3120	1.0149
8	1	4.1511	1.0078
9	1	5.3455	1.0038
10	1	9.0196	1.0018
11	1	10.0452	1.0006
12	1	9.1956	1.0002

order of the constant L and the interpolation theory. If we choose the some piecewise polynomial spaces of higher order, then some refinement on good accuracy is expected.

The improvement of computational costs is also extraordinary. In the previous work, it costs several times as much as the approximate computation in nonlinear problems. The inverse matrix is also calculated in Newton iterates. We get round the norm estimation of the inverse matrix in the verification part by Proposition 1. Table IV shows the ratio of computational costs. By a priori inverse operator norm estimation, we have the guaranteed error bounds with several percents of additional costs. However, there are some condition to use a priori estimation. In nonlinear problem, $N\|r\|_\infty\|\hat{u}^{N-1}\|_X < \pi$ must be satisfied.

Appendix

We introduce two theorems with respect to Theorem 1 and Proposition 1. These can be found in most textbooks, for example, see (Atkinson and Han, 2001) or (Zeidler, 1995) etc.

Theorem 3 (Fredholm alternative theorem)**.** *Let X be a Banach space, and let $\mathcal{K} : X \to X$ be compact. Then the equation $(I-\mathcal{K})u = f$ has a unique solution $u \in X$ if and only if the homogeneous equation $(I-\mathcal{K})v = 0$ has only the trivial solution $v = 0$. In such a case, the bijective operator $(I-\mathcal{K}) : X \to X$ has a bounded inverse $(I-\mathcal{K})^{-1}$.*

Theorem 4 (Riesz-Schauder theorem)**.** *Let X be a Banach space, and let $\mathcal{K} : X \to X$ be compact. The operator $(I-\mathcal{K}) : X \to X$ is surjective if and only if it is injective*

References

R.A. Adams. Sobolev spaces. *Academic Press*, New York, 1975.

K. Atkinson. The Numerical Solution of Integral Equations of the Second Kind. *Cambridge University Press*, 1997.

K. Atkinson and W. Han. Theoretical Numerical Analysis. *Springer*, 2001.

S.C. Brenner and L.R. Scott The Mathematical Theory of Finite Element Methods. *Springer*, 2008.

G. Kedem. A posteriori error bounds for two-point boundary value problems. *SIAM J. Numer. Anal.*, 18, 1981, pp.431-448.

M.T. Nakao. A numerical approach to the proof of existence of solutions for elliptic problems. *Japan J. Appl. Math.*, 5, 1988, pp.313-332.

M.T. Nakao. A numerical verification method for the existence of weak solutions for nonlinear boundary value problems. *J. Math. Anal. Appl.*, 164, 1992, pp.489-507.

M.T. Nakao and N. Yamamoto. Numerical Verification. *Nihonhyouron-sya*, 1998, (Japanese).

S. Oishi. Numerical Methods with Guaranteed Accuracy. *Corona-sya*, 2000, (Japanese).

S. Oishi. Numerical Uniqueness and Existence Theorem for Solution of Lippmann-Schwinger Equation to Two Dimensional Sound Scattering Problem. *Proceedings of The 36th Numerical Analysis Symposium proceedings*, Japan, 2007, pp.23-26.

M. Plum. Computer-assisted existence proofs for two-point boundary value problems. *Computing*, 46, 1991, pp.19-34.

S.M. Rump. INTLAB-INTerval LABoratory, a Matlab toolbox for verified computations. Hamburg University of Technology, "`http://www.ti3.tu-harburg.de/rump/intlab/`".

A. Takayasu, S. Oishi and T. Kubo. Guaranteed error estimate for solutions to linear two-point boundary value problems with FEM. *Proceedings of ASIA SIMULATION CONFERENCE 2009 (JSST 2009)*, Japan, 2009a, Paper ID: 163 (8 pages).

A. Takayasu, S. Oishi and T. Kubo. Guaranteed error estimate for solutions to two-point boundary value problem. *Proceedings of 2009 International Symposium on Nonlinear Theory and its Applications (NOLTA 2009)*, Japan, 2009b, pp.214-217.

E. Zeidler. Applied Functional Analysis, Main Principles and Their Applications. *Springer-Verlag*, 1995.

E. Zeidler. Nonlinear Functional Analysis and Its Applications, Part I Fixed Point Theorems. *Springer-Verlag*, 1986.