Numerische Mathematik (2020) 146:907–926
https://doi.org/10.1007/s00211-020-01155-7

Numerische
Mathematik

A new formulation using the Schur complement for the numerical existence proof of solutions to elliptic problems: without direct estimation for an inverse of the linearized operator

Kouta Sekine[1] · **Mitsuhiro T. Nakao**[2] · **Shin'ichi Oishi**[3]

Received: 31 May 2019 / Revised: 4 August 2020 / Accepted: 13 October 2020 /
Published online: 22 October 2020

Abstract

Infinite-dimensional Newton methods can be effectively used to derive numerical proofs of the existence of solutions to partial differential equations (PDEs). In computer-assisted proofs of PDEs, the original problem is transformed into the infinite-dimensional Newton-type fixed point equation $w = -\mathcal{L}^{-1}\mathcal{F}(\hat{u}) + \mathcal{L}^{-1}\mathcal{G}(w)$, where $\mathcal{L}$ is a linearized operator, $\mathcal{F}(\hat{u})$ is a residual, and $\mathcal{G}(w)$ is a nonlinear term. Therefore, the estimations of $\|\mathcal{L}^{-1}\mathcal{F}(\hat{u})\|$ and $\|\mathcal{L}^{-1}\mathcal{G}(w)\|$ play major roles in the verification procedures . In this paper, using a similar concept to block Gaussian elimination and its corresponding 'Schur complement' for matrix problems, we represent the inverse operator $\mathcal{L}^{-1}$ as an infinite-dimensional operator matrix that can be decomposed into two parts: finite-dimensional and infinite-dimensional. This operator matrix yields a new effective realization of the infinite-dimensional Newton method, which enables a more efficient verification procedure compared with existing Nakao's methods for the solution of elliptic PDEs. We present some numerical examples that confirm the usefulness of the proposed method. Related results obtained from the representation of the operator matrix as $\mathcal{L}^{-1}$ are presented in the "Appendix".

Mathematics Subject Classification 65G20 · 65N30 · 35J25

✉ Kouta Sekine
k.sekine@computation.jp

[1] Faculty of Information Networking for Innovation and Design, Toyo University, 1-7-11 Akabanedai, kita-ku, Tokyo 115-0053, Japan

[2] Faculty of Science and Engineering, Waseda University, 3-4-1 Okubo, Tokyo 169-8555, Japan

[3] Department of Applied Mathematics, Faculty of Science and Engineering, Waseda University, 3-4-1 Okubo, Tokyo 169-8555, Japan

1 Introduction

In this paper, we study a new approach to proving the existence of solutions to elliptic problems. The proposed approach offers an improvement over existing Nakao's methods that use a finite-dimensional projection, that is, FN and IN methods in [11] and [9, Part I]. Particularly, an important aspect of our approach is the formulation using the Schur complement without direct estimates of an inverse of the linearized operator. Our approach inherits the advantages of both Nakao's FN and IN methods using a finite-dimensional projection, which also indicates that the disadvantages of both methods are resolved.

We consider computer-assisted existence proofs for the nonlinear elliptic boundary value problem

$$\begin{cases} -\Delta u = f(u) & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases} \tag{1}$$

where $\Omega \subset \mathbb{R}^n (n = 1, 2, 3)$ is a bounded domain with a Lipschitz boundary, and $f : H_0^1(\Omega) \to H^{-1}(\Omega)$ is a given nonlinear function which is assumed to be Fréchet differentiable. Equation (1) is a basic case of a semi-linear elliptic partial differential equation (PDE), for which many computer-assisted proof methods have been developed [4–6,8–11,13–17]. These methods are intended to prove the existence of solutions based on the fixed point theorem in Sobolev spaces. Throughout this paper, $H^1(\Omega)$ is defined as the first-order L^2 Sobolev space, $H_0^1(\Omega) := \{u \in H^1(\Omega) : u = 0 \text{ on } \partial\Omega\}$ with the inner product $(u, w)_{H_0^1} := (\nabla u, \nabla w)_{L^2}$, and $H^{-1}(\Omega)$ is the topological dual of $H_0^1(\Omega)$. We now categorize the verification methods developed to date to clarify the significance and advantages of the method described in this paper:

- FN method: Applies Newton's method only for the finite-dimensional part (e.g., FN-Int [5,6,9,11], FN-Norm [9–11]).
- IN method: Uses the infinite-dimensional Newton method (e.g., Newton–Kantorovich-like theorem [9,15–17], IN-Linz [8,9,11,13], Newton–Kantorovich theorem [18]).

In the FN method, with an appropriate setting of the finite-dimensional subspace $V_h \subset H_0^1(\Omega)$, we first consider the Ritz projection $R_h : H_0^1(\Omega) \to V_h$ defined as

$$((I - R_h)u, v_h)_{H_0^1} = 0 \quad \forall v_h \in V_h$$

for $u \in H_0^1(\Omega)$, where I denotes the identity operator on $H_0^1(\Omega)$. As a property of R_h, we assume that there exists a positive constant $C(h)$ which can be numerically estimated satisfying

$$\|u - R_h u\|_{H_0^1} \le C(h)\|\Delta u\|_{L^2}, \ \forall u \in \{v \in H_0^1(\Omega) \mid \Delta v \in L^2(\Omega)\} \tag{2}$$

with the property that $C(h) \to 0$ as the parameter $h \to 0$. Using this projection, the problem is decomposed into two parts: finite-dimensional and infinite-dimensional.

Let $\psi_1,\ldots,\psi_N$ be a basis of V_h, and let $V_\perp := \{u \in H_0^1(\Omega) \mid (u, v_h)_{H_0^1} = 0,\ v_h \in V_h\}$ be an orthogonal complement of V_h. For a given approximate solution $\hat{u} \in V_h$, setting $w := u - \hat{u}$, $w_h := R_h w$, and $w_\perp := (I - R_h)w$, the FN method uses the following fixed point formulation:
find $w_h \in V_h,\ w_\perp \in V_\perp$ satisfying

$$\begin{cases} w_h = R_h \mathcal{A}^{-1}(f(w_h + w_\perp + \hat{u}) - \mathcal{A}\hat{u}), \\ w_\perp = (I - R_h)\mathcal{A}^{-1}(f(w_h + w_\perp + \hat{u}) - \mathcal{A}\hat{u}), \end{cases} \tag{3}$$

where $\mathcal{A} : H_0^1(\Omega) \to H^{-1}(\Omega)$ is equal to minus the weak Laplace operator. In [4], the candidate set of solutions was set to

$$W_h := \left\{ \sum_{i=1}^{N} W_i \psi_i \subset V_h \mid W_i \text{ is a closed interval in } \mathbb{R} \right\}, \tag{4}$$

$$W_\perp := \{w_\perp \in V_\perp \mid \|w_\perp\|_{H_0^1}, \leq \alpha\}, \tag{5}$$

and the fixed point theorem was applied to (3) to verify a solution in the set $W = W_h + W_\perp$. To confirm the verification condition in the fixed point theorem, the verified computation of solutions for a linear system of equations and the constructive a priori error estimates (2) for the Ritz projection play an essential role. This was the first approach to numerical verification, but is based on a contraction assumption which is more restrictive compared with the FN method.

The FN method applies Newton's method to the finite-dimensional part of (3) as follows:

Let $f'[\hat{u}]$ be the Fréchet derivative at $\hat{u}$ of the nonlinear term $f(u)$, and let $\mathcal{L} : H_0^1(\Omega) \to H^{-1}(\Omega)$ be a linear operator defined as

$$\mathcal{L} := \mathcal{A} - f'[\hat{u}]. \tag{6}$$

Furthermore, the finite-dimensional operator $T : V_h \to V_h$ is defined as

$$T := R_h \mathcal{A}^{-1} \mathcal{L}|_{V_h}, \tag{7}$$

and T is assumed to be invertible. Then, we can rewrite the finite-dimensional part of (3) as

$$w_h = T^{-1} R_h \mathcal{A}^{-1}(f(w_h + w_\perp + \hat{u}) - \mathcal{A}\hat{u} - f'[\hat{u}]w_h). \tag{8}$$

This FN method was developed in [5], and it has been confirmed that Newton's method is more effective for the verification of w_h than the method in [4]. Additionally, because the computation of w_h by the iterative use of (8) is sufficient to simply solve the linear system of equations, the cost is not large compared with the matrix norm estimation of T^{-1}. However, when $f(u)$ includes the first-order derivative ∇u, because of the property of interval arithmetic, it sometimes causes unexpected inefficiencies in the estimation of the right-hand side of (8); that is, the inefficiency appears as a result of the effect of the corresponding estimation of $\nabla w_\perp$ and leads to the failure of verification because of an explosive expansion of intervals in the iteration (e.g., see [9,10]). This implies that the effect of Newton's method on $w_\perp$ has not been achieved. To overcome this difficulty, in [9,10], the FN-Norm method was introduced. This ensures a more effective verification procedure by setting the candidate set W_h of the finite-dimensional part to

$$W_h := \left\{ w_h \in V_h \mid \|w_h\|_{H_0^1} \leq \gamma \right\}.$$

By contrast, the IN method assumes that the linearized operator $\mathcal{L}$ is invertible and considers the following fixed point equation:

find $w \in H_0^1(\Omega)$ such that

$$w = -\mathcal{L}^{-1}\mathcal{F}(\hat{u}) + \mathcal{L}^{-1}\mathcal{G}(w), \tag{9}$$

where

$$\mathcal{F}(\hat{u}) := \mathcal{A}\hat{u} - f(\hat{u}) \tag{10}$$

and

$$\mathcal{G}(w) := f'[\hat{u}]w + f(\hat{u}) - f(w + \hat{u}). \tag{11}$$

As (9) is a Newton-type formulation in the infinite-dimensional sense, the linear term with respect to w, which is a shortcoming of the FN method, is no longer present on the right-hand side of the equation. Therefore, a Newton–Kantorovich-like theorem can be derived using the fixed point equation (9). However, as the inverse operator $\mathcal{L}^{-1}$ cannot be calculated directly, it is necessary to show the invertibility of $\mathcal{L}$ and estimate the operator norm $\|\mathcal{L}^{-1}\|$. Thus, the evaluation of $||\mathcal{L}^{-1}||$ is the major task in the verification procedures of the IN method, and has been studied by many researchers since 1991. For example, [3,8,12,13,19–22] and [9, Part I] used the Ritz projection R_h and the projection error bound $C(h)$ by limiting Ω to bounded domains. Other methods exist that avoid the Ritz projection R_h and its error bound $C(h)$, and hence can be applied even for unbounded domains Ω, where the projection error bound $C(h)$ is not accessible or does not even exist (e.g., [14–16] and [9, Part II]). IN methods that do not need a projection error bound and methods that use projection error bounds have strengths and weaknesses, and a comparison of their qualities on a general level is impossible. Generally, the IN method defines the candidate set as

$$W := \{w \in H_0^1(\Omega) \mid \|w\|_{H_0^1} \leq \rho\}. \tag{12}$$

Therefore, the IN method overestimates the finite-dimensional parts compared with the FN method.

We now describe a new approach that incorporates the advantages of both Nakao's FN and IN methods using the Ritz projection R_h in [11] and [9, Part I]. In this paper, we essentially consider the IN method based on (9), but we propose a verification method *without* estimating the norm $\|\mathcal{L}^{-1}\|$. Through a computational procedure that avoids the direct evaluation of the operator norm $\|\mathcal{L}^{-1}\|$, highly accurate and efficient verification can be expected. We decompose (9) into finite-dimensional and infinite-dimensional parts; that is, $(w_h, w_\perp)^T \in V_h \times V_\perp$ such that

$$\begin{pmatrix} w_h \\ w_\perp \end{pmatrix} = -\begin{pmatrix} R_h\mathcal{L}^{-1}\mathcal{F}(\hat{u}) \\ (I-R_h)\mathcal{L}^{-1}\mathcal{F}(\hat{u}) \end{pmatrix} + \begin{pmatrix} R_h\mathcal{L}^{-1}\mathcal{G}(w_h+w_\perp) \\ (I-R_h)\mathcal{L}^{-1}\mathcal{G}(w_h+w_\perp) \end{pmatrix}. \tag{13}$$

However, we cannot directly calculate $\mathcal{L}^{-1}$ for the same reason as for the existing IN method. To overcome this difficulty, we introduce an operator matrix, whose existence will be proved later,

$$H = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} : V_h \times V_\perp \to V_h \times V_\perp \tag{14}$$

satisfying

$$\begin{pmatrix} R_h\mathcal{L}^{-1}g \\ (I-R_h)\mathcal{L}^{-1}g \end{pmatrix} = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}\begin{pmatrix} R_h\mathcal{A}^{-1}g \\ (I-R_h)\mathcal{A}^{-1}g \end{pmatrix}, \ \forall g \in H^{-1}(\Omega). \tag{15}$$

Here, we apply the fixed point theorem to the following equation:

$$\begin{pmatrix} w_h \\ w_\perp \end{pmatrix} = -H\begin{pmatrix} R_h\mathcal{A}^{-1}\mathcal{F}(\hat{u}) \\ (I-R_h)\mathcal{A}^{-1}\mathcal{F}(\hat{u}) \end{pmatrix} + H\begin{pmatrix} R_h\mathcal{A}^{-1}\mathcal{G}(w_h+w_\perp) \\ (I-R_h)\mathcal{A}^{-1}\mathcal{G}(w_h+w_\perp) \end{pmatrix} \tag{16}$$

with candidate sets (4) and (5). Note that the right-hand side of this fixed point equation no longer has linear terms in w_h nor $w_\perp$. Therefore, the proposed method overcomes the disadvantages of the FN method. Additionally, note that we can directly compute the finite-dimensional part (e.g., $H_{11}R_h\mathcal{A}^{-1}\mathcal{F}(\hat{u})$) by solving the linear system of equations, which is an advantage of the FN method. Therefore, the proposed method also overcomes the disadvantages of the IN method. Thus, our approach inherits the advantages of both Nakao's FN and IN methods using a finite-dimensional projection in [11] and [9, Part I], which also indicates that the disadvantages of both methods are resolved; that is, the disadvantage of the Newton method not working for the infinite-dimensional part of the FN method and the disadvantage of the poor computational efficiency of the finite-dimensional part in the IN method are removed. For the actual implementation of the verification procedure, it is necessary to obtain a more detailed form of the operator matrix H. Thus, we consider a specific construction of this matrix below.

The remainder of this paper is organized as follows: In Sect. 2, we describe how to construct the operator matrix H. In Sect. 3, we present the results of numerical experiments using the operator matrix H. We also describe why the proposed method offers an improvement over previous techniques based on some useful results in the “Appendix”.

2 Constitution of the inverse operator matrix *H*

In this section, we present a detailed description of the actual construction of the operator matrix H satisfying (14) and (15). The basic idea comes from the concept of block Gaussian elimination and its corresponding ‘Schur complement’ for matrix problems.

We consider a solution ϕ that satisfies the linear equation

$$\mathcal{L}\phi = g \tag{17}$$

for a given $g \in H^{-1}(\Omega)$. We denote

$$\begin{aligned} \phi_h &:= R_h\phi, & \phi_\perp &:= (I - R_h)\phi,\\ \mathcal{A}_h^{-1} &:= R_h\mathcal{A}^{-1}, & \mathcal{A}_\perp^{-1} &:= (I - R_h)\mathcal{A}^{-1}. \end{aligned}$$

We multiply $\mathcal{A}^{-1}$ from the left on both sides of (17), and decompose the result into the finite and infinite-dimensional parts using the Ritz projection R_h as follows:

$$\begin{aligned} &\begin{cases} R_h\mathcal{A}^{-1}\mathcal{L}(\phi_h + \phi_\perp) = \mathcal{A}_h^{-1}g\\ (I - R_h)\mathcal{A}^{-1}\mathcal{L}(\phi_h + \phi_\perp) = \mathcal{A}_\perp^{-1}g \end{cases}\\ \Leftrightarrow &\begin{cases} T\phi_h - R_h\mathcal{A}^{-1}f'[\hat{u}]\phi_\perp = \mathcal{A}_h^{-1}g\\ -(I - R_h)\mathcal{A}^{-1}f'[\hat{u}]\phi_h + (I_{V_\perp} - (I - R_h)\mathcal{A}^{-1}f'[\hat{u}])\phi_\perp = \mathcal{A}_\perp^{-1}g, \end{cases} \end{aligned}$$

where $I_{V_\perp}$ is the identity operator on $V_\perp$. Furthermore, transforming the above equation using the operator matrix yields

$$\begin{pmatrix} T & -\mathcal{A}_h^{-1}f'[\hat{u}]|_{V_\perp}\\ -\mathcal{A}_\perp^{-1}f'[\hat{u}]|_{V_h} & I_{V_\perp} - \mathcal{A}_\perp^{-1}f'[\hat{u}]|_{V_\perp} \end{pmatrix} \begin{pmatrix} \phi_h\\ \phi_\perp \end{pmatrix} = \begin{pmatrix} \mathcal{A}_h^{-1}g\\ \mathcal{A}_\perp^{-1}g \end{pmatrix}. \tag{18}$$

Let $Y : V_\perp \to V_h$, $Z : V_h \to V_\perp$, and $G : V_\perp \to V_\perp$ be bounded linear operators defined as

$$Y := -\mathcal{A}_h^{-1}f'[\hat{u}]|_{V_\perp},\ Z := -\mathcal{A}_\perp^{-1}f'[\hat{u}]|_{V_h} \text{ and } G := I_{V_\perp} - \mathcal{A}_\perp^{-1}f'[\hat{u}]|_{V_\perp}, \tag{19}$$

respectively. Additionally, we define the 2 × 2 block operator matrix D as

$$D := \begin{pmatrix} T & Y\\ Z & G \end{pmatrix} \equiv \begin{pmatrix} T & -\mathcal{A}_h^{-1}f'[\hat{u}]|_{V_\perp}\\ -\mathcal{A}_\perp^{-1}f'[\hat{u}]|_{V_h} & I_{V_\perp} - \mathcal{A}_\perp^{-1}f'[\hat{u}]|_{V_\perp} \end{pmatrix}. \tag{20}$$

Moreover, if the operators $\mathcal{L}$ and D are invertible, then we have

$$\begin{pmatrix} \phi_h \\ \phi_\perp \end{pmatrix} = \begin{pmatrix} R_h \mathcal{L}^{-1} g \\ (I - R_h)\mathcal{L}^{-1} g \end{pmatrix} = D^{-1} \begin{pmatrix} \mathcal{A}_h^{-1} g \\ \mathcal{A}_\perp^{-1} g \end{pmatrix}$$

and H is equal to D^{-1} from (15).

We first present a sufficient condition for the invertibility of the operator D, which also provides a detailed expression for D^{-1}.

Lemma 1 *The finite-dimensional operator $T : V_h \to V_h$ defined as (7) is assumed to be invertible. Let $S : V_\perp \to V_\perp$ be a linear operator defined as*

$$S := G - ZT^{-1}Y, \tag{21}$$

which is the so-called Schur complement corresponding to the operator T of the block operator matrix D. If S is bijective, then the operator D defined by (20) is also bijective and we have

$$D^{-1} = M_3 M_2 M_1,$$

where

$$M_1 = \begin{pmatrix} I_{V_h} & 0 \\ -ZT^{-1} & I_{V_\perp} \end{pmatrix}, \quad M_2 = \begin{pmatrix} I_{V_h} & -YS^{-1} \\ 0 & I_{V_\perp} \end{pmatrix}, \quad M_3 = \begin{pmatrix} T^{-1} & 0 \\ 0 & S^{-1} \end{pmatrix},$$

where I_{V_h} is the identity operator on V_h.

Proof . We first apply block Gaussian elimination to the operator matrix D. We multiply M_1 from the left of D as follows:

$$M_1 D = \begin{pmatrix} I_{V_h} & 0 \\ -ZT^{-1} & I_{V_\perp} \end{pmatrix} \begin{pmatrix} T & Y \\ Z & G \end{pmatrix} = \begin{pmatrix} T & Y \\ 0 & S \end{pmatrix}.$$

This procedure is so-called block forward elimination. Next, we multiply M_2 from the left of $M_1 D$ as

$$M_2 M_1 D = \begin{pmatrix} I_{V_h} & -YS^{-1} \\ 0 & I_{V_\perp} \end{pmatrix} \begin{pmatrix} T & Y \\ 0 & S \end{pmatrix} = \begin{pmatrix} T & 0 \\ 0 & S \end{pmatrix}.$$

From the assumption that T and S are bijective, we have

$$M_3 M_2 M_1 D = \begin{pmatrix} I_{V_h} & 0 \\ 0 & I_{V_\perp} \end{pmatrix}.$$

Note that the multiplication of $M_3 M_2$ corresponds to so-called back substitution.

Next, to show that $M_3M_2M_1$ is D^{-1}, we multiply $M_3M_2M_1$ from the right of D as follows:

$$DM_3M_2M_1 = \begin{pmatrix} I_{V_h} & YS^{-1} \\ ZT^{-1} & GS^{-1} \end{pmatrix} M_2M_1 = \begin{pmatrix} I_{V_h} & 0 \\ ZT^{-1} & I_{V_\perp} \end{pmatrix} M_1 = \begin{pmatrix} I_{V_h} & 0 \\ 0 & I_{V_\perp} \end{pmatrix}.$$

Thus, $M_3M_2M_1 = D^{-1}$ is proved. □

We obtain the inverse operator of D using Lemma 1. However, the operator matrix H satisfying (15) cannot be introduced without the invertibility of $\mathcal{L}$. Therefore, the following theorem provides a sufficient condition for the invertibility of $\mathcal{L}$.

Theorem 1 *Under the same notation and assumptions as in Lemma 1, the linearized operator $\mathcal{L}$ defined by (6) is bijective, and the operator matrix H satisfying (15) is given by*

$$H = D^{-1} = M_3M_2M_1.$$

Proof We show that the linearized operator $\mathcal{L}$ is bijective. From the assumptions and Lemma 1, we have an inverse operator of D. Multiplying D^{-1} from the left of (18), we obtain

$$\begin{pmatrix} \phi_h \\ \phi_\perp \end{pmatrix} = D^{-1} \begin{pmatrix} \mathcal{A}_h^{-1} g \\ \mathcal{A}_\perp^{-1} g \end{pmatrix}. \tag{22}$$

To prove that $\mathcal{L}$ is injective, we show that $\phi = 0$ is the only solution of the equation $\mathcal{L}\phi = 0$. This can be readily seen from the fact that, using the bijectivity of operators D and $\mathcal{A}$, the solution of (18) with $g = 0$ implies that $\phi = 0$.

Finally, we prove that the linearized operator $\mathcal{L}$ is surjective. For this, it is sufficient to show that there exists a solution $\phi \in H_0^1(\Omega)$ satisfying $\mathcal{L}\phi = g$ for any $g \in H^{-1}(\Omega)$. Now, for $g \in H^{-1}(\Omega)$, define $(\phi_h, \phi_\perp)^T$ as the left-hand side of (22) and set $\phi := \phi_h + \phi_\perp$. Then, by the invertibility of the operator matrix H, the function ϕ satisfies (18), and therefore it also implies (17), which proves the desired surjectivity. Thus, the operator $\mathcal{L}$ is bijective. □

Theorem 1 provides a specific expression of the operator matrix H satisfying (14). Moreover, Theorem 1 provides a new expression for the solution ϕ of the linear noncoercive elliptic PDE $\mathcal{L}\phi = g$. Therefore, for example, the exact expression of the Ritz projection error for the solution of the linear noncoercive elliptic PDE $\mathcal{L}\phi = g$ is also derived from Theorem 1 (see Appendix C for details).

At the end of this section, we describe how to verify that T and S in the assumptions of Theorem 1 are invertible operators. Let $\mathbf{G} \in \mathbb{R}^{N\times N}$ be a real matrix defined as $(\mathbf{G})_{i,j} := (\nabla\psi_j, \nabla\psi_i)_{L^2} - (f'[\hat{u}]\psi_j, \psi_i)_{L^2}$. Then, the matrix $\mathbf{G}$ is invertible if and only if T is the invertible operator (e.g., [9, Lemma 2.1, p.44]). Therefore, by confirming the invertibility of the matrix $\mathbf{G}$ using numerical computation with guaranteed accuracy, we can easily check the invertibility of T.

To confirm the invertibility of the operator S, the following operator norm is used. For two Banach spaces X and Y, the set of bounded linear operators from X to Y is denoted by $L(X,Y)$ with the operator norm $\|\mathcal{B}\|_{L(X,Y)} := \sup\{\|\mathcal{B}u\|_Y/\|u\|_X : u \in X \setminus \{0\}\}$ for $\mathcal{B} \in L(X,Y)$. When $X = Y$, we simply use $L(X)$. Then, we can confirm the invertibility of S using the following well-known theorem.

Lemma 2 *[Well known] Let S be a bounded linear operator satisfying (21). Setting $\kappa := \|\mathcal{A}_\perp^{-1}f'[\hat{u}]|_{V_\perp} + \mathcal{A}_\perp^{-1}f'[\hat{u}]|_{V_h}T^{-1}\mathcal{A}_h^{-1}f'[\hat{u}]|_{V_\perp}\|_{L(V_\perp)}$, if $\kappa < 1$ holds, then S is bijective and its norm satisfies*

$$\|S^{-1}\|_{L(V_\perp)} \leq \frac{1}{1-\kappa}.$$

Proof The Ritz projection R_h is a continuous projection; therefore, V_h and $V_\perp$ are Hilbert spaces with the inner product $(\cdot,\cdot)_{H_0^1}$, respectively. The definition $S := I_{V_\perp} - \mathcal{A}_\perp^{-1}f'[\hat{u}]|_{V_\perp} - \mathcal{A}_\perp^{-1}f'[\hat{u}]|_{V_h}T^{-1}\mathcal{A}_h^{-1}f'[\hat{u}]|_{V_\perp}$ and the assumption $\kappa < 1$ yield the bijectively of S as a result of the well-known theory of the Neumann series. □

To compute the constant κ, we use the inequality (2) and the Aubin–Nitsche trick

$$\|u - R_h u\|_{L^2} \leq C(h)\|u - R_h u\|_{H_0^1}, \ \forall u \in H_0^1(\Omega).$$

There are several ways to compute the constant κ. For example, in the case of $f'[\hat{u}] \in L(L^2(\Omega))$, we define the constants C_{f,L^2} and τ_{L^2} as

$$\begin{aligned} \|f'[\hat{u}]v\|_{L^2} &\leq C_{f,L^2}\|v\|_{L^2}, \ v \in L^2(\Omega), \\ \|T^{-1}\mathcal{A}_h^{-1}\|_{L(L^2(\Omega))} &\leq \tau_{L^2}. \end{aligned}$$

Then, we can estimate κ as

$$\kappa = C(h)^2 C_{f,L^2}(1 + C_{f,L^2}\tau_{L^2}). \tag{23}$$

Note that τ_{L^2} can be estimated as the matrix norm $\|\mathbf{L}^{\frac{1}{2}}\mathbf{G}^{-1}\mathbf{L}^{\frac{T}{2}}\|_E$ induced by the Euclidean vector norm, where $\mathbf{L}^{\frac{1}{2}}$ is the Cholesky factorization of the matrix $\mathbf{L}$, which is defined as $(\mathbf{L})_{i,j} := (\psi_j, \psi_i)_{L^2}$.

If $f'[\hat{u}]$ is in $L(H_0^1(\Omega), L^2(\Omega))$, we define the constants C_f, $C_{f,\perp}$, and τ_{L^2,H_0^1} as

$$\begin{aligned} \|f'[\hat{u}]v\|_{L^2} &\leq C_f\|v\|_{H_0^1}, \ v \in H_0^1(\Omega), \\ \|f'[\hat{u}]v_\perp\|_{L^2} &\leq C_{f,\perp}\|v_\perp\|_{H_0^1}, \ v \in V_\perp, \\ \|T^{-1}\mathcal{A}_h^{-1}\|_{L(L^2(\Omega),H_0^1(\Omega))} &\leq \tau_{L^2,H_0^1}, \end{aligned}$$

where τ_{L^2,H_0^1} can be also estimated using verified computing (e.g., see [9, (3.63), p.93]). Then, we can estimate κ as

$$\kappa = C(h)C_{f,\perp}(1 + \tau_{L^2,H_0^1}C_f). \tag{24}$$

The estimation (24) is the same as [9, (3.64), p.93]. It is also possible to derive an evaluation similar to [9, (3.46), p.89] from Lemma 2. Moreover, since $C(h)$ has the property that $C(h) \to 0$ as $h \to 0$, it is expected to satisfy the sufficient condition $\kappa < 1$.

3 Numerical examples

3.1 Verification procedure

In this subsection, we describe a verification procedure to realize computer-assisted proofs using the fixed point formulation (16) with the operator matrix H. The proposed verification method is a combination of the FN and IN methods developed above.

Let $\hat{u} \in V_h \subset H_0^1(\Omega)$ be a solution that satisfies the following equation:

$$(\nabla\hat{u}, \nabla v_h)_{L^2} = (f(\hat{u}), v_h)_{L^2}, \ \forall v_h \in V_h. \tag{25}$$

For example, $\hat{u}$ may be obtained by numerical computations with guaranteed accuracy for finite-dimensional nonlinear equations, such as the Krawczyk method. Note that $R_h \mathcal{A}^{-1}\mathcal{F}(\hat{u})$ in (16) becomes zero.

Next, we verify that the matrix $\mathbf{G}$ defined as $(\mathbf{G})_{i,j} := (\nabla\psi_j, \nabla\psi_i)_{L^2} - (f'[\hat{u}]\psi_j, \psi_i)_{L^2}$ and the operator S are invertible. As Theorem 1 implies that the linearized operator $\mathcal{L}$ defined as (6) is invertible, we can transform problem (1) into the fixed point equation (16) using the operator matrix H.

Schauder or Banach's fixed point theorem may be applied to the equation (16) with the candidate sets (4) and (5), as in [4,5]. Here, the fixed point theorem may be selected as follows. If the nonlinear term $f(u)$ is similar to that in [15] (e.g., $f(u) \in L^2(\Omega)\ \forall u \in H_0^1(\Omega)$), then because $\mathcal{L}^{-1}\mathcal{G}$ is a compact operator, Schauder's fixed point theorem can be used. If this is not the case, or if we need to prove the local uniqueness of the solution, it is preferable to use Banach's fixed point theorem. A survey of the FN method [6] makes it easy to select the appropriate method.

3.2 Example

In this subsection, we present an example in which our method is used to verify a solution of the elliptic boundary value problem

$$\begin{cases} -\Delta u = u^2 & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases} \tag{26}$$

with $\Omega = (0, 1)^2$. This is Emden's equation, which is a good test problem for comparing the proposed method with other approaches. Additionally, because the results for ρ in (12) given by many existing IN methods [8,15,16,18] are almost the same, it is not necessary to compare numerous existing IN approaches. Emden's equation has also been discussed in terms of the FN method [23].

All computations were implemented on a computer with 2.20 GHz Intel Xeon E7-4830 v2 CPU × 4, 2 TB RAM, and CentOS 7.2 using C++11 with GCC version 4.8.5. All rounding errors were strictly estimated using kv library [1]. This guarantees the mathematical correctness of all the results.

We constructed approximate solutions for (26) using a Legendre polynomial basis. Specifically, we defined the set $\{\psi_1, \psi_2, \dots \psi_N\}$ of Legendre polynomials as

$$\psi_i(x) := \frac{1}{i(i+1)} x(1-x) \frac{dP_i}{dx}(x),\ i = 1, 2, \dots,$$

with

$$P_i = \frac{(-1)^i}{i!} \left(\frac{d}{dx}\right)^i x^i (1-x)^i$$

and defined the finite-dimensional subspace as a tensor product

$$V_h^N := \mathrm{span}(\psi_1, \dots \psi_N) \otimes \mathrm{span}(\psi_1, \dots \psi_N).$$

See [2] for information on how to compute $C(h)$ in (2) corresponding to the Legendre polynomial basis. Then, $\hat{u} \in V_h^N$ satisfying the finite-dimensional nonlinear problem (25) can be written as

$$\hat{u}(x, y) = \sum_{i,j=1}^{N} \hat{u}_{i,j} \psi_i(x) \psi_j(y),$$

where $\hat{u}_{i,j}$ are real numbers. Note that, in a real computation, $\hat{u}_{i,j}$ are obtained as intervals that include the exact solution of (25) using the Krawczyk method. For example, when $N = 10$, a solution $\hat{u}$ satisfying (25) can be obtained (see Table 1). Here, 1.23_{456}^{789} denotes the interval [1.23456, 1.23789]. As the solution has symmetry, note that $\hat{u}_{i,j}$ is zero when i and j are even. Under this setting, exact solutions $\hat{u} \in V_h^{10}$ satisfying (25) were computed numerically. Their graphs are displayed in Fig. 1. The computational results of the constants $C(h)$ and κ using (23) were 0.037268163011998514 and 0.31088290279527165, respectively. Therefore, because $\kappa < 1$, the operator S was bijective. Thus, taking the candidate set as

$$W_h := \left\{ \sum_{i,j=1}^{N} W_{i,j} \psi_i \subset V_h^N \mid W_{ij} \text{ is a closed interval in } \mathbb{R} \right\}, \tag{27}$$

$$W_\perp := \{ w_\perp \in V_\perp \mid \|w_\perp\|_{H_0^1} \le \alpha \}, \tag{28}$$

the method proposed in this paper succeeded in the numerical verification of problem (26). The verified W_h results for the finite-dimensional parts are presented in Tables 1 and 2, and the α results for the infinite-dimensional parts are given in Table 2.

Table 1 For the case $N = 10$, the coefficient $\hat{u}_{i,j}$ of $\hat{u}$ and the coefficient $W_{i,j}$ of the guaranteed result W_h of the finite-dimensional part

i	j	$\hat{u}_{i,j}$	$W_{i,j}$
1	1	$366.41347081895 18^{5}_{4}$	$[-0.59855339300191369, 0.59855339300191369]$
1	3	$-152.7013688554553^{3}_{4}$	$[-1.2253950367566178, 1.2253950367566178]$
1	5	$51.38270599821782^{2}_{1}$	$[-0.98711934879483799, 0.98711934879483799]$
1	7	$-10.39254971836046^{6}_{7}$	$[-0.80157453121573663, 0.80157453121573663]$
1	9	$1.938028134034345^{5}_{4}$	$[-0.6402005288216297, 0.6402005288216297]$
3	1	$-152.7013688554553^{3}_{4}$	$[-1.2253950367566338, 1.2253950367566338]$
3	3	$106.206570760356^{50}_{49}$	$[-3.4911377868360356, 3.4911377868360356]$
3	5	$-47.02334109624576^{1}_{2}$	$[-5.2643807800044531, 5.2643807800044531]$
3	7	$11.16273975521142^{3}_{2}$	$[-5.4645410263719932, 5.4645410263719932]$
3	9	$-2.60410486533461^{89}_{90}$	$[-4.6121337856900269, 4.6121337856900269]$
5	1	$51.38270599821782^{2}_{1}$	$[-0.98711934879489661, 0.98711934879489661]$
5	3	$-47.02334109624576^{1}_{2}$	$[-5.2643807800046299, 5.2643807800046299]$
5	5	$23.64100645585704^{9}_{8}$	$[-10.278811561149042, 10.278811561149042]$
5	7	$-6.393838247547973^{8}_{9}$	$[-12.749567305746329, 12.749567305746329]$
5	9	$1.615528823767520^{6}_{5}$	$[-10.862495129381467, 10.862495129381467]$
7	1	$-10.39254971836046^{6}_{7}$	$[-0.80157453121577616, 0.80157453121577616]$
7	3	$11.16273975521142^{3}_{2}$	$[-5.4645410263722019, 5.4645410263722019]$
7	5	$-6.393838247547973^{8}_{9}$	$[-12.749567305746491, 12.749567305746491]$
7	7	$2.118831906690744^{2}_{1}$	$[-17.471111729696386, 17.471111729696386]$
7	9	$-0.6015754707587541^{4}_{5}$	$[-15.400941972506744, 15.400941972506744]$
9	1	$1.938028134034345^{5}_{4}$	$[-0.64020052882164458, 0.64020052882164458]$
9	3	$-2.60410486533461^{89}_{90}$	$[-4.6121337856902125, 4.6121337856902125]$
9	5	$1.615528823767520^{6}_{5}$	$[-10.862495129381687, 10.862495129381687]$
9	7	$-0.6015754707587541^{5}_{4}$	$[-15.400941972506861, 15.400941972506861]$
9	9	$0.1963785013053931^{8}_{7}$	$[-13.81780174057843, 13.81780174057843]$

Furthermore, we can prove that the exact solution u^* of (26) exists in $\hat{u} + W_h + W_\perp$, and we can estimate

$$\begin{aligned}\|u^* - \hat{u}\|_{H_0^1} &= \sqrt{\|R_h(u^* - \hat{u})\|^2_{H_0^1} + \|(I - R_h)(u^* - \hat{u})\|^2_{H_0^1}} \\ &\leq \sqrt{\sup \|W_h\|^2_{H_0^1} + \alpha^2} =: \rho.\end{aligned}$$

FN-Int (e.g., [5,6,11]) was also applied using similar candidate sets (27) and (28) and performing a detailed evaluation of the finite-dimensional and infinite-dimensional parts. However, as FN-Int only applies Newton's method to the finite-dimensional

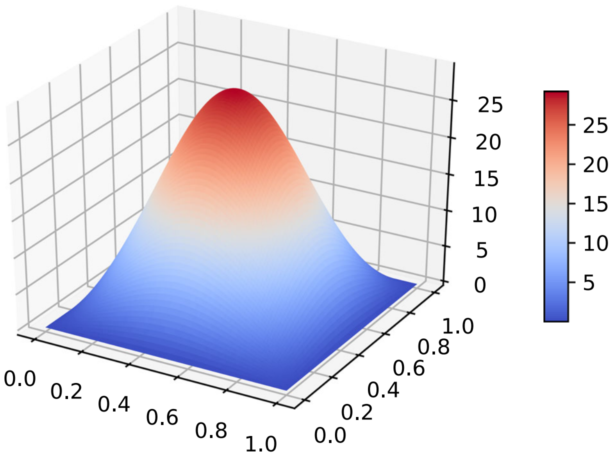

Fig. 1 Approximate solution of (26) ($N = 10$)

Table 2 Results of the norm evaluation for $N = 10$

method	$\sup \|W_h\|_{H_0^1}$	$\sup \|W_\perp\|_{H_0^1} \le \alpha$	$\|u^* - \hat{u}\|_{H_0^1} \le \rho$
Proposed	0.41226282803760456	0.14598888170328537	0.43734813702877418
[8,18]	–	–	1.4392104268509974

parts, the verification failed for $N = 10$, which implies that $N = 10$ was too small to achieve a successful verification.

As the error bound of the form $\|u^* - \hat{u}\|_{H_0^1} \le \rho$ was obtained in the course of a successful verification by the IN method for $N = 10$, we compared the proposed method with the IN method [8,18] from this viewpoint in Table 2. It is clear from the table that the value of ρ given by the proposed method was smaller than that produced by the existing IN method [8,18]. Additionally, as described in Sect. 3.1, because we partially incorporated the detailed calculations of the IN method into the proposed method, enhancing the IN method could lead to improvements to the results of the proposed method. The reason why the proposed method produced smaller values of ρ than [8,18] is discussed in Appendix B.

From Table 1, which presents results for $N = 10$, the error interval of the finite-dimensional part appears to be rather large. This is because N was too small. In fact, very sharp results were obtained in the case of $N = 40$ (see Tables 3 and 4). Additionally, we obtained $C(h) = 0.011501093662913 57$ and $\kappa = 0.029612643887446451$ using (23). Because $C(h)$ was smaller than the case of $N = 10$, the constant κ was also small . Here, $\pm 1.23^{789}_{456} \times 10^{-11}$ denotes the interval $[-1.23456 \times 10^{-11}, 1.23789 \times 10^{-11}]$.

Table 3 For the case $N = 40$, the coefficient $\hat{u}_{i,j}$ of $\hat{u}$ and the coefficient $W_{i,j}$ of the guaranteed result W_h of the finite-dimensional part

i	j	$\hat{u}_{i,j}$	$W_{i,j}$
1	1	$366.4132119391286_{8}^{9}$	$\pm 2.890339699_{1423603}^{0254412} \times 10^{-11}$
1	3	$-152.7015836846106_{1}^{0}$	$\pm 5.953039492_{5126237}^{6304853} \times 10^{-11}$
1	5	$51.38361560840872_{6}^{7}$	$\pm 4.9329670131_{634702}^{167493} \times 10^{-11}$
1	7	$-10.39601151027711_{1}^{0}$	$\pm 4.3133385563_{658583}^{703399} \times 10^{-11}$
1	9	$1.946100573383924_{1}^{2}$	$\pm 4.05963569250_{10538}^{54126} \times 10^{-11}$
...	...	...	...
1	39	$-4.66924461_{568399560}^{123737562} \times 10^{-12}$	$\pm 2.40673056579_{81076}^{49860} \times 10^{-11}$
3	1	$-152.7015836846106_{1}^{0}$	$\pm 5.953039492_{5723655}^{6902272} \times 10^{-11}$
3	3	$106.2071425465360_{7}^{8}$	$\pm 1.7449437714_{236332}^{139130} \times 10^{-10}$
3	5	$-47.02462830095045_{7}^{6}$	$\pm 2.780252688_{5991257}^{6047521} \times 10^{-10}$
3	7	$11.16900188749037_{0}^{1}$	$\pm 3.25583866986_{64090}^{45918} \times 10^{-10}$
3	9	$-2.604322866563870_{7}^{6}$	$\pm 3.59972699307_{48283}^{54461} \times 10^{-10}$
...	...	...	...
3	39	$-6.18701438_{85354111}^{75591113} \times 10^{-11}$	$\pm 2.432811717189_{8549}^{6228} \times 10^{-10}$
5	1	$51.38361560840872_{6}^{7}$	$\pm 4.932967013_{3060612}^{2593404} \times 10^{-11}$
5	3	$-47.02462830095045_{7}^{6}$	$\pm 2.7802526886_{393443}^{449707} \times 10^{-10}$
5	5	$23.6420677490914_{49}^{50}$	$\pm 5.90180296672_{40844}^{13061} \times 10^{-10}$
5	7	$-6.399132323486540_{8}^{7}$	$\pm 8.5522929274_{298881}^{306522} \times 10^{-10}$
5	9	$1.610834571271026_{0}^{1}$	$\pm 1.03788894888648_{46}^{85} \times 10^{-09}$
...	...	...	...
5	39	$-2.4529205624_{387362}^{231429} \times 10^{-10}$	$\pm 7.790762135816_{5518}^{2716} \times 10^{-10}$
7	1	$-10.39601151027711_{1}^{0}$	$\pm 4.3133385564_{792260}^{837076} \times 10^{-11}$
7	3	$11.16900188749037_{0}^{1}$	$\pm 3.25583866991_{72636}^{54463} \times 10^{-10}$
7	5	$-6.399132323486540_{8}^{7}$	$\pm 8.55229292747_{66569}^{74210} \times 10^{-10}$
7	7	$2.124267226042428_{2}^{3}$	$\pm 1.43656500261616_{32}^{22} \times 10^{-09}$
7	9	$-0.5986559904211578_{9}^{8}$	$\pm 1.936424374565_{5089}^{4920} \times 10^{-09}$
...	...	...	...
7	39	$-6.160592931_{6028183}^{4758797} \times 10^{-10}$	$\pm 1.6509622637659_{917}^{898} \times 10^{-09}$
9	1	$1.946100573383924_{1}^{2}$	$\pm 4.0596356925_{874838}^{918427} \times 10^{-11}$
9	3	$-2.604322866563870_{7}^{6}$	$\pm 3.59972699312_{58183}^{64361} \times 10^{-10}$
9	5	$1.610834571271026_{0}^{1}$	$\pm 1.03788894889364_{30}^{69} \times 10^{-09}$
9	7	$-0.5986559904211578_{9}^{8}$	$\pm 1.9364243745704_{203}^{034} \times 10^{-09}$
9	9	$0.1907886733421144_{7}^{8}$	$\pm 2.8485078315651_{604}^{930} \times 10^{-09}$
...	...	...	...
9	39	$-1.2284714549_{968888}^{703340} \times 10^{-09}$	$\pm 2.8184100199765_{508}^{181} \times 10^{-09}$
...	...	...	...
39	39	$9.32486073_{24817149}^{46069927} \times 10^{-10}$	$\pm 5.810714309286_{8021}^{9468} \times 10^{-09}$

Table 4 Results of the norm evaluation for $N = 40$

method	$\sup \|W_h\|_{H_0^1}$	$\sup \|W_\perp\|_{H_0^1} \le \alpha$	$\|u^* - \hat{u}\|_{H_0^1} \le \rho$
Proposed	4.43666×10^{-10}	4.56210×10^{-11}	4.46007×10^{-10}
[8,18]	–	–	9.35991×10^{-10}

4 Conclusion

In this paper, we described a new formulation (16) for the numerical proof of the existence of solutions for elliptic problems. The proposed approach has advantages over both Nakao's FN and IN methods using the Ritz projection R_h in [11] and [9, Part I]. In particular, we derived a specific formula for the operator matrix H, which is needed to compute (16), in Theorem 1. As a result, while using the infinite-dimensional Newton method, the error evaluation for each coefficient interval of the finite basis is also enclosed, as demonstrated by the results in Tables 1 and 3. This is considered an advantage of the FN method. Furthermore, the proposed method produced better results than the IN method [8,18], even for the norm estimation, as demonstrated in Tables 2 and 4.

Acknowledgements This work was supported by JST CREST Grant Number JPMJCR14D4 and MEXT under the "Exploratory Issue on Post-K computer" project (Development of verified numerical computations and super high-performance computing environment for extreme research). The second author was supported by JSPS KAKENHI Grant Number 18K03434. We thank the editors and reviewers for providing useful comments that helped to improve the content of this manuscript.

Appendix A Another formula for the operator matrix *H*

In Lemma 1 and Theorem 1, the Schur complement S (defined in (21)) for the $(1, 1)$ element T of the operator matrix D is created and some properties are proved. We present another form of the operator matrix H using the Schur complement $S_h : V_h \to V_h$ for the $(2, 2)$ element $I_{V_\perp} - \mathcal{A}_\perp^{-1} f'[\hat{u}]|_{V_\perp} (=: G)$ of the operator matrix D.

Lemma 3 *The infinite-dimensional operator $G : V_\perp \to V_\perp$ defined in (19) is assumed to be nonsingular. Let $S_h : V_h \to V_h$ be a linear operator defined as*

$$S_h := T - \mathcal{A}_h^{-1} f'[\hat{u}]|_{V_\perp} G^{-1} \mathcal{A}_\perp^{-1} f'[\hat{u}]|_{V_h}, \tag{29}$$

which is the so-called Schur complement corresponding to the operator G of the block operator matrix D. If S_h is bijective, then the operator D defined by (20) *is also bijective and we have*

$$D^{-1} = M_3' M_2' M_1',$$

where

$$M_1' = \begin{pmatrix} I_{V_h} & -YG^{-1} \\ 0 & I_{V_\perp} \end{pmatrix}, \ M_2' = \begin{pmatrix} I_{V_h} & 0 \\ -ZS_h^{-1} & I_{V_\perp} \end{pmatrix}, \ M_3' = \begin{pmatrix} S_h^{-1} & 0 \\ 0 & G^{-1} \end{pmatrix}.$$

Proof We prove Lemma 3 in the same way as Lemma 1. We multiply M_1' from the left of D as follows:

$$M_1' D = \begin{pmatrix} I_{V_h} & -YG^{-1} \\ 0 & I_{V_\perp} \end{pmatrix} \begin{pmatrix} T & Y \\ Z & G \end{pmatrix} = \begin{pmatrix} S_h & 0 \\ Z & G \end{pmatrix}.$$

Next, we multiply M_2' from the left of $M_1' D$ as follows:

$$M_2' M_1' D = \begin{pmatrix} I_{V_h} & 0 \\ -ZS_h^{-1} & I_{V_\perp} \end{pmatrix} \begin{pmatrix} S_h & 0 \\ Z & G \end{pmatrix} = \begin{pmatrix} S_h & 0 \\ 0 & G \end{pmatrix}.$$

From the assumption that S_h and G are bijective, we have

$$M_3' M_2' M_1' D = \begin{pmatrix} I_{V_h} & 0 \\ 0 & I_{V_\perp} \end{pmatrix}.$$

Next, to show that $M_3' M_2' M_1'$ is D^{-1}, we multiply $M_3' M_2' M_1'$ from the right of D as follows:

$$D M_3' M_2' M_1' = \begin{pmatrix} TS_h^{-1} & YG^{-1} \\ ZS_h^{-1} & I_{V_\perp} \end{pmatrix} M_2' M_1' = \begin{pmatrix} I_{V_h} & YG^{-1} \\ 0 & I_{V_\perp} \end{pmatrix} M_1' = \begin{pmatrix} I_{V_h} & 0 \\ 0 & I_{V_\perp} \end{pmatrix}.$$

Therefore, $M_3' M_2' M_1'$ is the bounded inverse operator of D. □

Theorem 2 *Under the same notation and assumptions as in Lemma 3, the linearized operator $\mathcal{L}$ defined by* (6) *is bijective, and the operator matrix H satisfying* (15) *is*

$$H = D^{-1} = M_3' M_2' M_1'.$$

Proof This assertion can be proved in a similar manner to Theorem 1 using Lemma 3 instead of Lemma 1. □

As we can verify the invertibility of the operator G in a similar manner to that used in Lemma 2, it is also possible to derive a verification procedure using Theorem 2 instead of Theorem 1.

Moreover, it is possible to derive, from the operator matrix D in Theorem 1, the results in previous papers [8,12,21] for the norm evaluation $\|\mathcal{L}^{-1}\|_{L(H^{-1},H_0^1)}$ of the inverse operator $\mathcal{L}^{-1}$ (see Appendix B). However, there has been no previous discussion (e.g., [9]) of the norm evaluation of the operator $\mathcal{L}^{-1} : H^{-1}(\Omega) \to H_0^1(\Omega)$ using the operator matrix D in Theorem 2.

Appendix B Relation between the norm $\|\mathcal{L}^{-1}\|_{L(H^{-1},H_0^1)}$ and Theorem 1

There have been many studies on the estimation of the norm $\|\mathcal{L}^{-1}\|_{L(H^{-1},H_0^1)}$ because it plays an important role in the existing IN method (e.g., [8,9,12–15,19,21]). In the following, we show that it is also possible to obtain the upper bound of the inverse operator norm $\|\mathcal{L}^{-1}\|_{L(H^{-1},H_0^1)}$ using Theorem 1.

Corollary 1 *(of Theorem 1) Under the same assumptions as in Theorem 1, $\mathcal{L}$ is invertible, and it follows that*

$$\|\mathcal{L}^{-1}\|_{L(H^{-1},H_0^1)} \le \left\| \begin{pmatrix} \|T^{-1}+T^{-1}\mathcal{A}_h^{-1}f'[\hat{u}]|_{V_\perp}S^{-1}\mathcal{A}_\perp^{-1}f'[\hat{u}]|_{V_h}T^{-1}\|_{L(H_0^1)} & \|T^{-1}\mathcal{A}_h^{-1}f'[\hat{u}]|_{V_\perp}S^{-1}\|_{L(H_0^1)} \\ \|S^{-1}\mathcal{A}_\perp^{-1}f'[\hat{u}]|_{V_h}T^{-1}\|_{L(H_0^1)} & \|S^{-1}\|_{L(H_0^1)} \end{pmatrix} \right\|_E,$$

where $\|\cdot\|_E$ denotes a matrix norm induced by the Euclidean vector norm $|\cdot|_E$.

Proof We define the norm of the direct product space $V_h \times V_\perp$ as

$$\left\| \begin{pmatrix} \phi_h \\ \phi_\perp \end{pmatrix} \right\|_{V_h\times V_\perp} := \left| \begin{pmatrix} \|\phi_h\|_{H_0^1} \\ \|\phi_\perp\|_{H_0^1} \end{pmatrix} \right|_E = \sqrt{\|\phi_h\|_{H_0^1}^2 + \|\phi_\perp\|_{H_0^1}^2},\ (\phi_h,\phi_\perp)^T \in V_h \times V_\perp.$$

Then, from Theorem 1, the solution ϕ of the linear equation (17) can be evaluated as

$$\begin{aligned} \|\phi\|_{H_0^1} &= \left| \begin{pmatrix} \|R_h\phi\|_{H_0^1} \\ \|(I-R_h)\phi\|_{H_0^1} \end{pmatrix} \right|_E = \left\| H \begin{pmatrix} \mathcal{A}_h^{-1}g \\ \mathcal{A}_\perp^{-1}g \end{pmatrix} \right\|_{V_h\times V_\perp} \\ &\le \|H\|_{L(V_h\times V_\perp)} \left\| \begin{pmatrix} \mathcal{A}_h^{-1}g \\ \mathcal{A}_\perp^{-1}g \end{pmatrix} \right\|_{V_h\times V_\perp} = \|H\|_{L(V_h\times V_\perp)} \|\mathcal{L}\phi\|_{H^{-1}}. \end{aligned}$$

Therefore, we have

$$\|\mathcal{L}^{-1}\|_{L(H^{-1},H_0^1)} \le \|H\|_{L(V_h\times V_\perp)}.$$

Moreover, using the structure of the operator matrix H satisfying (14), we have

$$\|H\|_{L(V_h\times V_\perp)} = \sup_{z=(z_h,z_\perp)^T\in V_h\times V_\perp} \frac{\left\| \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \begin{pmatrix} z_h \\ z_\perp \end{pmatrix} \right\|_{V_h\times V_\perp}}{\|z\|_{V_h\times V_\perp}}$$

$$
\begin{aligned}
&= \sup_{z=(z_h,z_\perp)^T\in V_h\times V_\perp} \frac{\left|\begin{pmatrix} \|H_{11}z_h + H_{12}z_\perp\|_{H_0^1} \\ \|H_{21}z_h + H_{22}z_\perp\|_{H_0^1} \end{pmatrix}\right|_E}{\|z\|_{V_h\times V_\perp}} \\
&\le \left\| \begin{pmatrix} \|H_{11}\|_{L(H_0^1)} & \|H_{12}\|_{L(H_0^1)} \\ \|H_{21}\|_{L(H_0^1)} & \|H_{22}\|_{L(H_0^1)} \end{pmatrix} \right\|_E .
\end{aligned}
$$

□

Remark 1 The estimation of the inverse operator norm in Corollary 1 is closely related to the method used in previous studies [8,9,12,21]. However, it is noted that Theorem 1 can be applied directly without using Corollary 1. For example, for the first term on the right-hand side of (9), the existing IN method evaluates

$$\|\mathcal{L}^{-1}\mathcal{F}(\hat{u})\|_{H_0^1} \le \|\mathcal{L}^{-1}\|_{L(H^{-1},H_0^1)}\|\mathcal{F}(\hat{u})\|_{H^{-1}}$$

to apply Corollary 1. If $\hat{u}$ is a solution satisfying (25), then $R_h\mathcal{A}^{-1}\mathcal{F}(\hat{u})$ vanishes, but this is not reflected in the estimation of the norm $\|\mathcal{L}^{-1}\|_{L(H^{-1},H_0^1)}$. By contrast, using Theorem 1 directly, we can estimate

$$\begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}\begin{pmatrix} R_h\mathcal{A}^{-1}\mathcal{F}(\hat{u}) \\ (I-R_h)\mathcal{A}^{-1}\mathcal{F}(\hat{u}) \end{pmatrix} = \begin{pmatrix} 0 & H_{12} \\ 0 & H_{22} \end{pmatrix}\begin{pmatrix} 0 \\ (I-R_h)\mathcal{A}^{-1}\mathcal{F}(\hat{u}) \end{pmatrix}.$$

Therefore, the estimations of $\|H_{11}\|_{L(H_0^1)}$ and $\|H_{21}\|_{L(H_0^1)}$ are useless in the evaluation for $\|\mathcal{L}^{-1}\|_{L(H^{-1},H_0^1)}$; that is, the proposed method is better than existing IN methods for evaluating $\|\mathcal{L}^{-1}\|_{L(H^{-1},H_0^1)}$.

The inverse operator norm $\|\mathcal{L}^{-1}\|_{L(H^{-1},H_0^1)}$ can be evaluated from Theorem 2 using the same procedure as in Corollary 1.

Corollary 2 (of Theorem 2) *Under the same assumptions as in Theorem 2, $\mathcal{L}$ is invertible and*

$$
\begin{aligned}
&\|\mathcal{L}^{-1}\|_{L(H^{-1},H_0^1)} \\
&\le \left\| \begin{pmatrix} \|S_h^{-1}\|_{L(H_0^1)} & \|S_h^{-1}\mathcal{A}_h^{-1}f'[\hat{u}]|_{V_\perp}G^{-1}\|_{L(H_0^1)} \\ \|G^{-1}\mathcal{A}_\perp^{-1}f'[\hat{u}]|_{V_h}S_h^{-1}\|_{L(H_0^1)} & \|G^{-1}+G^{-1}\mathcal{A}_\perp^{-1}f'[\hat{u}]|_{V_h}S_h^{-1}\mathcal{A}_h^{-1}f'[\hat{u}]|_{V_\perp}G^{-1}\|_{L(H_0^1)} \end{pmatrix} \right\|_E .
\end{aligned}
$$

Appendix C An exact expression for the Ritz projection error in the solution of the noncoercive equation (17)

In this appendix, we derive an expression for the Ritz projection error $\phi - R_h\phi$ in the exact solution ϕ of the linear noncoercive elliptic PDE (17). Using this result, we can, for example, determine a constant C that satisfies the inequality $\|\phi - R_h\phi\|_{H_0^1} \le$

$C\|\mathcal{L}\phi\|_{L^2}$, $\phi \in \{\phi \in H_0^1 | \Delta\phi \in L^2(\Omega)\}$, even though the elliptic operator $\mathcal{L}$ is noncoercive (cf. [7]).

Corollary 3 (of Theorem 1) *Under the same assumptions as in Theorem 1, the Ritz projection error of the solution $\phi \in H_0^1(\Omega)$ of the linear noncoercive elliptic PDE (17) is*

$$\phi - R_h\phi = S^{-1}(I - R_h)\mathcal{A}^{-1}\left(I + f'[\hat{u}]|_{V_h} T^{-1} R_h \mathcal{A}^{-1}\right)\mathcal{L}\phi.$$

Proof The proof is obtained immediately from $\phi_\perp$ in the expression of the proof for Theorem 1. □

Using Corollary 3 as $g \in L^2(\Omega)$, it is easy to derive the constant C satisfying $\|\phi - R_h\phi\|_{H_0^1} \leq C\|\mathcal{L}\phi\|_{L^2}$ in [7].

Similarly, we can derive the following proposition from Theorem 2.

Corollary 4 *(of Theorem 2) Under the same assumptions as in Theorem 2, the Ritz projection error in the solution $\phi \in H_0^1(\Omega)$ of the linear noncoercive elliptic PDE (17) is represented as follows:*

$$\begin{aligned}&\phi - R_h\phi\\&= G^{-1}(I - R_h)\mathcal{A}^{-1}\left(I + f'[\hat{u}]|_{V_h} S_h^{-1} R_h \mathcal{A}^{-1}\left(I + f'[\hat{u}]|_{V_\perp} G^{-1}(I - R_h)\mathcal{A}^{-1}\right)\right)\mathcal{L}\phi.\end{aligned}$$

References

1. Kashiwagi, M.: kv library (2016). http://verifiedby.me/kv/index-e.html
2. Kimura, S., Yamamoto, N.: On explicit bounds in the error for the H_0^1-projection into piecewise polynomial spaces. Bull. Inf. Cybern. **31**(2), 109–115 (1999)
3. Kinoshita, T., Watanabe, Y., Nakao, M.T.: An alternative approach to norm bound computation for inverses of linear operators in Hilbert spaces. J. Differ. Equ. **266**(9), 5431–5447 (2019)
4. Nakao, M.T.: A numerical approach to the proof of existence of solutions for elliptic problems. Jpn. J. Appl. Math. **5**(2), 313–332 (1988)
5. Nakao, M.T.: A numerical approach to the proof of existence of solutions for elliptic problems ii. Jpn. J. Appl. Math. **7**(3), 477 (1990)
6. Nakao, M.T.: Numerical verification methods for solutions of ordinary and partial differential equations. Numer. Funct. Anal. Optim. **22**(3–4), 321–356 (2001)
7. Nakao, M.T., Hashimoto, K.: Guaranteed error bounds for finite element approximations of noncoercive elliptic problems and their applications. J. Comput. Appl. Math. **218**(1), 106–115 (2008)
8. Nakao, M.T., Hashimoto, K., Watanabe, Y.: A numerical method to verify the invertibility of linear elliptic operators with applications to nonlinear problems. Computing **75**(1), 1–14 (2005)
9. Nakao, M.T., Plum, M., Watanabe, Y.: Numerical Verification Methods and Computer-Assisted Proofs for Partial Differential Equations. Springer, Berlin (2019)
10. Nakao, M.T., Watanabe, Y.: An efficient approach to the numerical verification for solutions of elliptic differential equations. Numer. Algorithms **37**(1–4), 311–323 (2004)
11. Nakao, M.T., Watanabe, Y.: Numerical verification methods for solutions of semilinear elliptic boundary value problems. Nonlinear Theory Appl. IEICE **2**(1), 2–31 (2011)
12. Nakao, M.T., Watanabe, Y., Kinoshita, T., Kimura, T., Yamamoto, N.: Some considerations of the invertibility verifications for linear elliptic operators. Jpn. J. Ind. Appl. Math. **32**(1), 19–31 (2015)
13. Oishi, S.: Numerical verification of existence and inclusion of solutions for nonlinear operator equations. J. Comput. Appl. Math. **60**(1), 171–185 (1995)

14. Plum, M.: Bounds for eigenvalues of second-order elliptic differential operators. Zeitschrift für angewandte Mathematik und Physik ZAMP **42**(6), 848–863 (1991)
15. Plum, M.: Enclosures for weak solutions of nonlinear elliptic boundary value problems. In: Agarwal, R.P. (ed.) Inequalities and Applications, pp. 505–521. World Scientific, Singapore (1994)
16. Plum, M.: Existence and multiplicity proofs for semilinear elliptic boundary value problems by computer assistance. Jahresbericht der Deutschen Mathematiker Vereinigung **110**(1), 19–54 (2008)
17. Plum, M.: Computer-assisted proofs for semilinear elliptic boundary value problems. Jpn. J. Ind. Appl. Math. **26**(2–3), 419–442 (2009)
18. Takayasu, A., Liu, X., Oishi, S.: Verified computations to semilinear elliptic boundary value problems on arbitrary polygonal domains. Nonlinear Theory Appl. IEICE **4**(1), 34–61 (2013)
19. Tanaka, K., Takayasu, A., Liu, X., Oishi, S.: Verified norm estimation for the inverse of linear elliptic operators using eigenvalue evaluation. Jpn. J. Ind. Appl. Math. **31**(3), 665–679 (2014)
20. Watanabe, Y., Kinoshita, T., Nakao, M.: A posteriori estimates of inverse operators for boundary value problems in linear elliptic partial differential equations. Math. Comput. **82**(283), 1543–1557 (2013)
21. Watanabe, Y., Kinoshita, T., Nakao, M.T.: An improved method for verifying the existence and bounds of the inverse of second-order linear elliptic operators mapping to dual space. Jpn. J. Ind. Appl. Math. **36**(2), 1–14 (2019)
22. Watanabe, Y., Nagatou, K., Plum, M., Nakao, M.T.: Norm bound computation for inverses of linear operators in Hilbert spaces. J. Differ. Equ. **260**(7), 6363–6374 (2016)
23. Watanabe, Y., Nakao, M.T.: Numerical verifications of solutions for nonlinear elliptic equations. Jpn. J. Ind. Appl. Math. **10**(1), 165–178 (1993)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.