Advances in Applied Mathematics and Mechanics
Adv. Appl. Math. Mech., Vol. **16**, No. 1, pp. 24-46

DOI: 10.4208/aamm.OA-2022-0019
February 2024

A Nitsche-Based Element-Free Galerkin Method for Semilinear Elliptic Problems

Tao Zhang[1] and Xiaolin Li[2,3,*]

[1] *School of Mathematics, Physics and Data Science, Chongqing University of Science and Technology, Chongqing 401331, China*
[2] *School of Mathematical Sciences, Chongqing Normal University, Chongqing 400047, China*
[3] *Key Laboratory for Optimization and Control Ministry of Education, Chongqing Normal University, Chongqing 400047, China*

Received 20 January 2022; Accepted (in revised version) 20 December 2022

Abstract. A Nitsche-based element-free Galerkin (EFG) method for solving semilinear elliptic problems is developed and analyzed in this paper. The existence and uniqueness of the weak solution for semilinear elliptic problems are proved based on a condition that the nonlinear term is an increasing Lipschitz continuous function of the unknown function. A simple iterative scheme is used to deal with the nonlinear integral term. We proved the existence, uniqueness and convergence of the weak solution sequence for continuous level of the simple iterative scheme. A commonly used assumption for approximate space, sometimes called inverse assumption, is proved. Optimal order error estimates in L^2 and H^1 norms are proved for the linear and semilinear elliptic problems. In the actual numerical calculation, the characteristic distance h does not appear explicitly in the parameter β introduced by the Nitsche method. The theoretical results are confirmed numerically.

AMS subject classifications: 65N15, 65N30

Key words: Meshless method, element-free Galerkin method, Nitsche method, semilinear elliptic problem, error estimate.

1 Introduction

Numerical methods are requisite and useful for the study of semilinear partial differential equations (PDEs) [1]. The nonlinearity of the semilinear problems only involves the unknown function, not its derivative. Many works have been devoted to the numerical

*Corresponding author.
Emails: zhangtao895701777@126.com (T. Zhang), lxlmath@163.com (X. Li)

solutions of semilinear elliptic problems such as finite element method (FEM) [2, 3], finite difference method [4], finite volume element method [5] and discontinuous Galerkin method [6]. Recently, some collocation meshless (or meshfree) methods [7, 8], Galerkin-type meshless method [8] and generalized finite difference method [9, 10] have been developed to solve the semilinear PDEs. Unlike mesh-based numerical methods, the shape functions used in the meshless methods [11–14] are linkage with nodes (or particles) scattered in the underlying computational domain, which reduces the dependence on the mesh. The meshless methods have greatly developed in the last three decades.

The element-free Galerkin (EFG) [14] method is a global Galerkin-type meshless discretization technique for PDEs. The EFG shape functions are derived from the moving least-squares (MLS) approximation [15]. The difficulty with the imposition of essential (or Dirichlet) boundary conditions stems from the fact that the MLS shape functions are not interpolating. That is, the shape function associated with a node does not vanish at other nodes. Recently, some variants of the MLS approximation have been developed to regain interpolating properties, e.g., interpolating MLS method [15], simplified interpolating MLS method [18, 19], and improved interpolating MLS method [20], and smoothed MLS approximation [21]. On the other hand, the EFG method have been developed for solving solute transport problems [22], tumor growth model [23] and heat transport equation [24], as well as some nonlinear models, such as magnetohydrodynamics(MHD) [25] and Korteweg-de Vries-Rosenau-regularized long-wave equations [26].

In addition to adopting the interpolating shape functions, some mandatory methods, such as the Lagrange multiplier method [12–14], the penalty method [12, 13, 16, 17, 27, 28] and Nitsche method [29–33, 35], can straightforwardly use the non-interpolating shape functions by modifying the original weak form. The Nitsche method was first introduced in early 70's in FEM context [29]. This approach seems to be more promising because of its ease of implementation, its smaller parameter-value compared with the penalty method, its maintenance in terms of the number of unknown variables and the symmetry positive definiteness of the resulting system. Therefore, the Nitsche method is seen as a consistent improvement of the penalty method [31], and these potential advantages bring some conveniences for numerical analysis.

There are a few theoretical results in the Nitsche-based meshless method. In Ref. [32], the approximation errors of the Galerkin meshless method for linear elliptic problem are analyzed based on the nonsymmetric Nitsche method and an inverse assumption, and the effect of the numerical integration are discussed. The error estimates combined with the effect of numerical integration are also developed in [33, 34] based on the reproducing kernel gradient smoothing integration method. Using the Nitsche method, a fast time discrete EFG method is analyzed for the fractional diffusion-wave equation [35]. In these currently reported works, however, the parameter β introduced by the Nitsche method is empirical rather than rational, meanwhile, an unproven inverse assumption is required in the Nitsche-based meshless numerical analysis. Moreover, the analysis presented in [32, 33, 35] addresses only the linear PDEs.

A Nitsche-based element-free Galerkin method is presented in this paper for semi-

linear elliptic problems. The modified weak form for semilinear elliptic problems is analyzed, and the presented simple iterative scheme is also interpreted. In addition, based on EFG discretization, a commonly-used assumption in the Nitsche method is proven, and optimal orders of convergence in H^1 and L^2 norms for the linear and semilinear elliptic problems are derived. Finally, the value of the Nitsche parameter and the convergence condition of the iterative scheme are discussed in detail in numerical examples.

The rest of the paper is organized as follows. First we introduce the Nitsche method for semilinear elliptic problem in Section 2, and give a simple iterative scheme and related convergence results in Section 3. We focus on the H^1 and L^2 error estimates of the EFG discretization in Section 4. Finally the theoretical results are tested by numerical examples in Section 5 and conclusions are summarized in Section 6.

Throughout this paper, we use C, with or without subscript, to denote a general positive constant which is independent of characteristic distance h and could take different values at different appearances.

2 The Nitsche method for semilinear elliptic problem

Consider the following semilinear elliptic problem

$$\begin{cases} -\nabla\cdot\mathbf{a}\nabla u+bu+c(u)=f & \text{in } \Omega, \\ u=g_1 & \text{on } \Gamma_1, \\ \mathbf{a}\nabla u\cdot\mathbf{n}=g_2 & \text{on } \Gamma_2, \end{cases} \tag{2.1}$$

in which $\Omega\subset\mathbb{R}^n(n\geq 1)$ be a nonempty bounded domain with a Lipschitz boundary $\Gamma=\Gamma_1\cup\Gamma_2$. $u=u(\mathbf{x})$, $g_1=g_1(\mathbf{x})\in H^{1/2}(\Gamma_1)$, $g_2=g_2(\mathbf{x})\in L^2(\Gamma_2)$, $f=f(\mathbf{x})\in L^2(\Omega)$ and $\mathbf{n}$ is the unit outward normal to Γ. Besides, the non-decreasing Lipschitz continuous function $c(u)=c(u(\mathbf{x}))$ depends on the unknown u and satisfies the following conditions [2,36]:

(A1) $(c(u_1)-c(u_2))(u_1-u_2)\geq 0$, $\forall u_1,u_2\in\mathbb{R}$.

(A2) There is a positive constant L such that

$$|c(u_1)-c(u_2)|\leq L|u_1-u_2|,\quad \forall u_1,u_2\in\mathbb{R}.$$

Moreover, $b=b(\mathbf{x})$ is a bounded function, i.e., $0\leq b_0\leq b\leq b_1$. Furthermore, $\mathbf{a}=\mathbf{a}(\mathbf{x})=\big(a_{ij}(\mathbf{x})\big)_{n\times n}$ is a symmetric matrix-valued function satisfying

$$a^0\mathbf{x}^{\mathrm{T}}\mathbf{x}\leq\mathbf{x}^{\mathrm{T}}\mathbf{a}\mathbf{x},\quad \forall\mathbf{x}\in\mathbb{R}^n,\quad \big\|a_{ij}\big\|_{L^\infty(\Omega)}\leq a^1,$$

for two positive constants a^0 and a^1.

A variational formulation of (2.1) is to find $u\in H^1(\Omega)$ such that

$$a_0(u,v)=F_0(v),\quad \forall v\in H^1(\Omega), \tag{2.2}$$

where

$$a_0(u,v)=(\mathbf{a}\nabla u,\nabla v)-(\mathbf{a}\nabla u\cdot\mathbf{n},v)_{\Gamma_1}+(bu,v)+(c(u),v),$$
$$F_0(v)=(f,v)+(g_2,v)_{\Gamma_2}.$$

The Nitsche method at the continuous level is based on a modification of (2.2) by adding some boundary integral terms at both sides of (2.2), that is

$$a_0(u,v)-(\mathbf{a}\nabla v\cdot\mathbf{n},u)_{\Gamma_1}+\beta(u,v)_{\Gamma_1}=F_0(v)+\beta(g_1,v)_{\Gamma_1}-(\mathbf{a}\nabla v\cdot\mathbf{n},g_1)_{\Gamma_1},\quad\forall v\in H^1(\Omega),\tag{2.3}$$

in which β is a large enough positive constant parameter in order to impose the essential boundary conditions and

$$(v_1,v_2)_{\Gamma_1}=\int_{\Gamma_1}v_1v_2d\mathbf{x}.$$

The second term on the left-hand side of the above equation ensures the symmetry and adjoint consistency of the formulation. The terms with β compensate for the departure of interpolation of approximate solution and render stability. For the sake of convenience, (2.3) can be abbreviated as

$$a(u,v)=F(v),$$

where

$$a(u,v)=(\mathbf{a}\nabla u,\nabla v)-(\mathbf{a}\nabla u\cdot\mathbf{n},v)_{\Gamma_1}+(bu,v)+(c(u),v)-(\mathbf{a}\nabla v\cdot\mathbf{n},u)_{\Gamma_1}+\beta(u,v)_{\Gamma_1}.$$

In [30,31], a necessary assumption is

$$\|\mathbf{a}\nabla v\cdot\mathbf{n}\|_{L^2(\Gamma_1)}\leq C\|\nabla v\|_{L^2(\Omega)}$$

for all v in the approximate space to ensure the coercivity of the bilinear form $a(\cdot,\cdot)$ at the discrete level. We put forward a same assumption in the continuous meaning based on a similar purpose, namely

(A3) $\|\mathbf{a}\nabla w\cdot\mathbf{n}\|_{L^2(\Gamma_1)}\leq C_0\|\nabla w\|_{L^2(\Omega)}$, $\forall w\in H^1(\Omega)$.

For fixed $u\in H^1(\Omega)$, $a(u,\cdot)\in H^{-1}(\Omega)$ (the dual space of $H^1(\Omega)$), we define an operator $A:H^1(\Omega)\to H^{-1}(\Omega)$ by $Au=a(u,\cdot)$. $\langle Au,v\rangle$ denotes the value of the continuous linear functional $Au\in H^{-1}(\Omega)$ applied to $v\in H^1(\Omega)$, i.e., $\langle Au,v\rangle=a(u,v)$.

From (A1) and (A3), we have

$$\begin{aligned}
&|\langle Au-Av,u-v\rangle|\\
=&|a(u,u)-a(u,v)-a(v,u)+a(v,v)|\\
=&\left|\begin{array}{c}(\mathbf{a}\nabla(u-v),\nabla(u-v))-2(\mathbf{a}\nabla(u-v)\cdot\mathbf{n},(u-v))_{\Gamma_1}\\+(b(u-v),(u-v))+(c(u)-c(v),(u-v))+\beta(u-v,u-v)_{\Gamma_1}\end{array}\right|\\
\geq&a^0\|\nabla(u-v)\|^2_{L^2(\Omega)}-\varepsilon\|u-v\|^2_{L^2(\Gamma_1)}-\frac{1}{\varepsilon}\|\mathbf{a}\nabla(u-v)\cdot\mathbf{n}\|^2_{L^2(\Gamma_1)}\\
&+b_0\|u-v\|^2_{L^2(\Omega)}+\beta\|u-v\|^2_{L^2(\Gamma_1)}\\
\geq&\left(a^0-\frac{C_0^2}{\varepsilon}\right)\|\nabla(u-v)\|^2_{L^2(\Omega)}+b_0\|u-v\|^2_{L^2(\Omega)}+(\beta-\varepsilon)\|u-v\|^2_{L^2(\Gamma_1)}\\
\geq&C_1\left(\|\nabla(u-v)\|^2_{L^2(\Omega)}+\|u-v\|^2_{L^2(\Omega)}+\beta\|u-v\|^2_{L^2(\Gamma_1)}\right),
\end{aligned}\tag{2.4}$$

for any $\varepsilon>0$, in which

$$C_1=\min\left(a^0-\frac{C_0^2}{\varepsilon},\frac{\beta-\varepsilon}{\beta},b_0\right)>0$$

implies

$$\beta>\varepsilon>\frac{C_0^2}{a^0}.$$

Now we define $\|\cdot\|_\beta$ as

$$\|w\|^2_\beta=\|\nabla w\|^2_{L^2(\Omega)}+\|w\|^2_{L^2(\Omega)}+\beta\|w\|^2_{L^2(\Gamma_1)}.\tag{2.5}$$

Clearly, $\|\cdot\|_\beta$ is a norm. The strong monotonicity of operator A is determined as,

$$|\langle Au-Av,u-v\rangle|\geq C_1\|u-v\|^2_\beta.\tag{2.6}$$

On the other hand,

$$\begin{aligned}
&|\langle Au_1,v\rangle-\langle Au_2,v\rangle|\\
=&\left|\begin{array}{c}(\mathbf{a}\nabla(u_1-u_2),\nabla v)-(\mathbf{a}\nabla(u_1-u_2)\cdot\mathbf{n},v)_{\Gamma_1}-(\mathbf{a}\nabla v\cdot\mathbf{n},u_1-u_2)_{\Gamma_1}\\+(b(u_1-u_2),v)+(c(u_1)-c(u_2),v)+\beta(u_1-u_2,v)_{\Gamma_1}\end{array}\right|\\
\leq&\|\nabla(u_1-u_2)\|_{L^2(\Omega)}\left(\frac{a^1}{2}\|\nabla v\|_{L^2(\Omega)}+C_0\|v\|_{L^2(\Gamma_1)}\right)+\beta\|u_1-u_2\|_{L^2(\Gamma_1)}\|v\|_{L^2(\Gamma_1)}\\
&+C_0\|\nabla v\|_{L^2(\Omega)}\|u_1-u_2\|_{L^2(\Gamma_1)}+(b_1+L)\|u_1-u_2\|_{L^2(\Omega)}\|v\|_{L^2(\Omega)}\\
&+\|\nabla v\|_{L^2(\Omega)}\frac{a^1}{2}\|\nabla(u_1-u_2)\|_{L^2(\Omega)}\\
\leq&C_2\|u_1-u_2\|_\beta\|v\|_\beta,
\end{aligned}\tag{2.7}$$

in which

$$C_2=\max\left(C_0,\frac{a^1}{2},L+b_1,1\right).$$

Hence the operator A is Lipschitz continuous. Furthermore, we define $\langle l,v\rangle=F(v)$,

$$|\langle l,v\rangle|\leq\left(\|f\|_{L^2(\Omega)}+\|g_2\|_{L^2(\Gamma_2)}+\left(\sqrt{\beta}+1\right)\|g_1\|_{H^{1/2}(\Gamma_1)}\right)\|v\|_\beta. \tag{2.8}$$

Then, $l\in H^{-1}(\Omega)$ and $\langle Au,v\rangle=\langle l,v\rangle$ has a unique solution $u\in H^1(\Omega)$ [36], namely, (2.3) has a unique solution u.

3 Simple iterative scheme

Due to the existence of the nonlinear integral term $(c(u),v)$, the weak solution u cannot be directly obtained from (2.3). We apply a simple iterative scheme to derive the weak solution and prove the convergence of this format.

Simple iterative scheme: choose an initial value $u^0\in H^1(\Omega)$ and acquire u^{k+1} by solving the linear system

$$a_1\left(u^{k+1},v\right)=F(v)-\left(c\left(u^k\right),v\right),\quad\forall v\in H^1(\Omega),\quad k=0,1,2,\cdots, \tag{3.1}$$

where

$$\begin{aligned}a_1\left(u^{k+1},v\right)=&\left(\mathbf{a}\nabla u^{k+1},\nabla v\right)-\left(\mathbf{a}\nabla u^{k+1}\cdot\mathbf{n},v\right)_{\Gamma_1}+\left(bu^{k+1},v\right)\\&-\left(\mathbf{a}\nabla v\cdot\mathbf{n},u^{k+1}\right)_{\Gamma_1}+\beta\left(u^{k+1},v\right)_{\Gamma_1}.\end{aligned}$$

Theorem 3.1. *For given initial value $u^0\in H^1(\Omega)$, there exists a unique weak solution sequence $\{u^k\}$ generated by* (3.1) *that converges to the weak solution u of* (2.3) *if $q=\frac{L}{C_1}\in(0,1)$, and*

$$\left\|u^k-u\right\|_j\leq\frac{q^k}{1-q}\left\|u^1-u^0\right\|_j, \tag{3.2}$$

where j can be β, $L^2(\Omega)$ or $H^1(\Omega)$ to represent different norms.

Proof. The coercivity of $a_1(\cdot,\cdot)$ comes from a derivation similar to (2.4),

$$a_1(v,v)\geq C_1\|v\|_\beta^2.$$

Similar to (2.7), the continuity of $a_1(\cdot,\cdot)$ follows

$$a_1(w,v)\leq C_2\|w\|_\beta\|v\|_\beta.$$

Since $c(u)$ is a continuous function, then $l_1(\cdot)=F(\cdot)-\left(c\left(u^k\right),\cdot\right):H^1(\Omega)\to\mathbb{R}$ is a bounded linear functional. According to Lax-Milgram theorem, there has a unique weak solution $u^{k+1}\in H^1(\Omega)$ for (3.1).

Let $u^k,u^{k+1}\in H^1(\Omega)$ be weak solutions derived from (3.1),

$$a_1\left(u^{k+1}-u^k,v\right)=-\left(c\left(u^k\right)-c\left(u^{k-1}\right),v\right),\quad \forall v\in H^1(\Omega). \tag{3.3}$$

Since $u^{k+1}-u^k\in H^1(\Omega)$, again by Lax-Milgram theorem, the above equality has a unique solution and

$$\left\|u^{k+1}-u^k\right\|_\beta\le\frac{1}{C_1}\sup_{v\in H^1(\Omega)}\frac{\left|\left(c\left(u^k\right)-c\left(u^{k-1}\right),v\right)\right|}{\|v\|_\beta}\le q\left\|u^k-u^{k-1}\right\|_{L^2(\Omega)}, \tag{3.4}$$

where $q=\frac{L}{C_1}$. In a general way, using

$$\left\|u^{k+1}-u^k\right\|_{L^2(\Omega)}\le\left\|u^{k+1}-u^k\right\|_{H^1(\Omega)}\le\left\|u^{k+1}-u^k\right\|_\beta$$

yields

$$\left\|u^m-u^k\right\|_j\le\left(\sum_{i=k}^{m-1}q^i\right)\left\|u^1-u^0\right\|_j,\quad m>k, \tag{3.5}$$

where j is β, $L^2(\Omega)$ or $H^1(\Omega)$. Since $\left\{u^k\right\}$ is a Cauchy sequence in $H^1(\Omega)$ when $q\in(0,1)$, by the completeness, there exists a unique $\bar{u}$ such that $u^k\to\bar{u}(k\to\infty)$. On the other hand, applying the continuity of $c(u^k)$ and $a_1\left(u^{k+1},v\right)$ obtains

$$\lim_{k\to\infty}\left(c\left(u^k\right),v\right)=(c(\bar{u}),v),\quad \lim_{k\to\infty}a_1\left(u^{k+1},v\right)=a_1(\bar{u},v).$$

Let $k\to\infty$ in (3.1), we get

$$a_1(\bar{u},v)=F(v)-(c(\bar{u}),v),\quad \forall v\in H^1(\Omega).$$

Therefore $\bar{u}=u$ a.e. in Ω and let $m\to\infty$ in (3.5) implies (3.2). At this point, we complete the proof. □

4 EFG error analysis

To approximate the solutions of the iterative formulation (3.1), we need to give the finite-dimensional subspace $\mathbf{V}_h\subset H^1$ as

$$\mathbf{V}_h=\text{span}\{\Phi_i(\mathbf{x}),\,1\le i\le N\},$$

in which $\{\mathbf{x}_i\}_{i=1}^N$ be a set of N nodes in $\bar{\Omega}=\Omega\cup\Gamma$ and the MLS shape functions $\Phi_i(\mathbf{x})$ are

$$\Phi_i(\mathbf{x})=\begin{cases}\sum_{j=1}^{m}p_j(\mathbf{x})\left[\mathbf{A}^{-1}(\mathbf{x})\mathbf{B}(\mathbf{x})\right]_{jI}, & i=\lambda_I\in\wedge(\mathbf{x}),\\ 0, & i\notin\wedge(\mathbf{x}),\end{cases}\qquad i=1,2,\cdots,N,\tag{4.1}$$

in which $p_j(\mathbf{x})$ denote the shifted and scaled monomial basis functions [17], $\wedge(\mathbf{x})=\{\lambda_1,\lambda_2,\cdots,\lambda_{n_\mathbf{x}}\}$ indicates the global sequence numbers of nodes whose support domains cover the point $\mathbf{x}$. The support domain of $\mathbf{x}$ is $\Re(\mathbf{x})$ with radius $r(\mathbf{x})$,

$$\Re(\mathbf{x})=\{\mathbf{y}\in\mathbb{R}^n:|\mathbf{y}-\mathbf{x}|\le r(\mathbf{x})\}.$$

Others,

$$\mathbf{A}(\mathbf{x})=\mathbf{P}^{\mathrm{T}}\mathbf{W}(\mathbf{x})\mathbf{P},\quad \mathbf{B}(\mathbf{x})=\mathbf{P}^{\mathrm{T}}\mathbf{W}(\mathbf{x}),$$

with

$$\begin{aligned}&\mathbf{P}=\left[\mathbf{p}(\mathbf{x}_{\lambda_1}),\cdots,\mathbf{p}\left(\mathbf{x}_{\lambda_{n_\mathbf{x}}}\right)\right]^{\mathrm{T}},\\ &\mathbf{p}(\mathbf{x}_i)=(p_1(\mathbf{x}_i),\cdots,p_m(\mathbf{x}_i))^{\mathrm{T}},\\ &\mathbf{W}(\mathbf{x})=\mathrm{diag}\left(w_{\lambda_1}(\mathbf{x}),\cdots,w_{\lambda_{n_\mathbf{x}}}(\mathbf{x})\right).\end{aligned}$$

Assume that the distribution of discrete nodes $\{\mathbf{x}_i\}_{i=1}^N$ meets the following conditions:

(B1) Define the characteristic distance h as

$$h\le h_i\le Ch,\quad h_i=\min_{1\le j\le N,\ j\ne i}\left|\mathbf{x}_i-\mathbf{x}_j\right|.$$

(B2) To ensure the invertibility of $\mathbf{A}(\mathbf{x})$,

$$\mathrm{card}\{\wedge(\mathbf{x})\}\ge\dim\{p_j(\mathbf{x})\}=\frac{(\hat{m}+n)!}{\hat{m}!n!},$$

where $\hat{m}$ represents the largest degree of the used monomial basis functions.

In addition, assume that the derivatives of weight function $w_i(\mathbf{x})$ up to order γ are bounded and continuous such that $w_i(\mathbf{x})\in C_0^\gamma(\Re(\mathbf{x}_i))$. Then, MLS shape functions $\Phi_i(\mathbf{x})$ are bounded and γ-times continuously differentiable, i.e., $\Phi_i(\mathbf{x})\in C_0^\gamma(\Re(\mathbf{x}_i))$.

Lemma 4.1 ([16,17]). *Suppose that $w\in H^{\hat{m}+1}(\Omega)$, conditions* (B1) *and* (B2) *are satisfied, $\mathcal{M}w$ denotes the MLS approximation of w, then*

$$\|w-\mathcal{M}w\|_{H^k(\Omega)}\le Ch^{\hat{m}+1-k}\|w\|_{H^{\hat{m}+1}(\Omega)},\quad 0\le k\le\min\{\hat{m}+1,\gamma\}.$$

The EFG method for the simple iterative scheme follows. Set $k=0$ and choose a tolerance τ. The initial value $u_h^0 \in H^1(\Omega)$ is the solution of the following linear system,

$$a_1\left(u_h^0, v\right) = F(v), \quad \forall v \in \mathbf{V}_h. \tag{4.2}$$

1. Obtain u_h^{k+1} by solving

$$a_1\left(u_h^{k+1}, v\right) = F(v) - \left(c\left(u_h^k\right), v\right), \quad \forall v \in \mathbf{V}_h. \tag{4.3}$$

2. Update k to $k+1$ and go to 1 if $|u_h^{k+1} - u_h^k| > \tau |u_h^k|$. Otherwise, stop the iterative algorithm.

Theorem 4.1. *The sequence of iterative solutions $\{u_h^k\}$ generated by* (4.3) *converges to u_h when $q = \frac{L}{C_1} \in (0,1)$, and*

$$\left\| u_h^k - u_h \right\|_j \leq \frac{q^k}{1-q} \left\| u_h^1 - u_h^0 \right\|_j, \tag{4.4}$$

where j means β, $L^2(\Omega)$ or $H^1(\Omega)$.

Proof. Let $u_h^k, u_h^{k+1} \in \mathbf{V}_h$ be solutions obtained from (4.3),

$$a_1\left(u_h^{k+1} - u_h^k, v\right) = -\left(c\left(u_h^k\right) - c\left(u_h^{k-1}\right), v\right), \quad \forall v \in \mathbf{V}_h. \tag{4.5}$$

Similar to (3.4),

$$\left\| u_h^{k+1} - u_h^k \right\|_j \leq q \left\| u_h^k - u_h^{k-1} \right\|_j, \tag{4.6}$$

where j can be β, $L^2(\Omega)$ or $H^1(\Omega)$ implies different norms. Therefore, $\{u_h^k\}$ is still a Cauchy sequence in $H^1(\Omega)$ because $\mathbf{V}_h \subset H^1(\Omega)$, so there exists a unique $u_h \in H^1(\Omega)$ such that $u_h^k \to u_h (k \to \infty)$. From the continuity of $c(u_h^k)$ and $a_1(u_h^{k+1}, v)$, let $k \to \infty$ in (4.3),

$$a_1(u_h, v) = F(v) - (c(u_h), v), \quad \forall v \in \mathbf{V}_h.$$

Finally, (4.6) implies (4.4). □

To estimate the errors, let $R_h : H^1 \to \mathbf{V}_h$ be the orthogonal projection of H^1 on $\mathbf{V}_h$ such that

$$a_1\left(R_h u^{k+1}, v\right) = a_1\left(u^{k+1}, v\right), \quad \forall v \in \mathbf{V}_h, \quad k = 0,1,2,\cdots. \tag{4.7}$$

Since $R_h u^{k+1}$ is the best approximation of u^{k+1} in $\mathbf{V}_h$ with respect to $\|\cdot\|_\beta$,

$$\left\| R_h u^{k+1} - u^{k+1} \right\|_\beta \leq \left\| \mathcal{M} u^{k+1} - u^{k+1} \right\|_\beta.$$

In order to derive the L^2 norm error of the projection R_h, we define the auxiliary problem

$$\begin{cases} -\nabla\cdot\mathbf{a}\nabla\psi+b\psi=u^{k+1}-R_hu^{k+1} & \text{in } \Omega,\\ \psi=0 & \text{on } \Gamma_1,\\ \mathbf{a}\nabla\psi\cdot\mathbf{n}=0 & \text{on } \Gamma_2, \end{cases}$$

the weak form of which is

$$a_1(\psi,v)=\left(u^{k+1}-R_hu^{k+1},v\right),\quad \forall v\in H^1.$$

Assume that Ω is a convex domain or the boundary Γ of Ω is smooth, then

$$\|\psi\|_{H^2(\Omega)}\leq C\left\|u^{k+1}-R_hu^{k+1}\right\|_{L^2(\Omega)}.$$

According to (4.7) and the fact that $\mathcal{M}\psi\in\mathbf{V}_h$, we have $a_1\left(\mathcal{M}\psi,R_hu^{k+1}\right)=a_1\left(\mathcal{M}\psi,u^{k+1}\right)$. Then, $a_1\left(\mathcal{M}\psi,u^{k+1}-R_hu^{k+1}\right)=0$. Hence,

$$\begin{aligned}&\left\|u^{k+1}-R_hu^{k+1}\right\|^2_{L^2(\Omega)}=\left(u^{k+1}-R_hu^{k+1},u^{k+1}-R_hu^{k+1}\right)\\ =&a_1\left(\psi,u^{k+1}-R_hu^{k+1}\right)=a_1\left(\psi-\mathcal{M}\psi+\mathcal{M}\psi,u^{k+1}-R_hu^{k+1}\right)\\ =&a_1\left(\psi-\mathcal{M}\psi,u^{k+1}-R_hu^{k+1}\right)\leq C_2\|\psi-\mathcal{M}\psi\|_\beta\left\|u^{k+1}-R_hu^{k+1}\right\|_\beta.\end{aligned}\tag{4.8}$$

From the trace inequality and Lemma 4.1, we have

$$\|\psi-\mathcal{M}\psi\|^2_{L^2(\Gamma_1)}\leq C\|\psi-\mathcal{M}\psi\|_{L^2(\Omega)}\|\psi-\mathcal{M}\psi\|_{H^1(\Omega)}\leq Ch^3\|\psi\|^2_{H^2(\Omega)},\tag{4.9a}$$

$$\|\nabla(\psi-\mathcal{M}\psi)\cdot\mathbf{n}\|^2_{L^2(\Gamma_1)}\leq Ch\|\psi\|^2_{H^2(\Omega)}.\tag{4.9b}$$

Then

$$\begin{aligned}\|\psi-\mathcal{M}\psi\|_\beta\leq&\|\psi-\mathcal{M}\psi\|_{H^1(\Omega)}+\|\nabla(\psi-\mathcal{M}\psi)\cdot\mathbf{n}\|^{1/2}_{L^2(\Gamma_1)}\|\psi-\mathcal{M}\psi\|^{1/2}_{L^2(\Gamma_1)}\\ &+\sqrt{\beta}\|\psi-\mathcal{M}\psi\|_{L^2(\Gamma_1)}\\ \leq&C\left(h\|\psi\|_{H^2(\Omega)}+\sqrt{\beta}h^{1+1/2}\|\psi\|_{H^2(\Omega)}\right).\end{aligned}\tag{4.10}$$

Combining (4.8) and (4.10) gains

$$\left\|u^{k+1}-R_hu^{k+1}\right\|_{L^2(\Omega)}\leq C\left(h+\sqrt{\beta}h^{1+1/2}\right)\left\|u^{k+1}-R_hu^{k+1}\right\|_\beta.\tag{4.11}$$

Lemma 4.2. *For $q=\frac{L}{C_1}\in(0,1)$, we have*

$$\left\|R_hu^{k+1}-u_h^{k+1}\right\|_{H^1(\Omega)}\leq\left\|R_hu^{k+1}-u_h^{k+1}\right\|_\beta\leq q\left\|u^k-u_h^k\right\|_{H^1(\Omega)},\tag{4.12a}$$

$$\left\|R_hu^{k+1}-u_h^{k+1}\right\|_{L^2(\Omega)}\leq q\left\|R_hu^k-u_h^k\right\|_{L^2(\Omega)}.\tag{4.12b}$$

Proof. Subtracting (3.1) from (4.3) yields

$$a_1\left(u^{k+1}-u_h^{k+1},v\right)=-\left(c\left(u^k\right)-c\left(u_h^k\right),v\right),\quad \forall v\in\mathbf{V}_h. \tag{4.13}$$

Since $R_hu^{k+1}-u_h^{k+1}\in\mathbf{V}_h$,

$$\begin{aligned}\left\|R_hu^{k+1}-u_h^{k+1}\right\|_\beta^2&\leq\frac{1}{C_1}a_1\left(R_hu^{k+1}-u_h^{k+1},R_hu^{k+1}-u_h^{k+1}\right)\\&=\frac{1}{C_1}a_1\left(R_hu^{k+1}-u^{k+1}+u^{k+1}-u_h^{k+1},R_hu^{k+1}-u_h^{k+1}\right)\\&=\frac{1}{C_1}a_1\left(u^{k+1}-u_h^{k+1},R_hu^{k+1}-u_h^{k+1}\right)\\&=\frac{1}{C_1}\left(c\left(u_h^k\right)-c\left(u^k\right),R_hu^{k+1}-u_h^{k+1}\right)\\&\leq q\left\|R_hu^{k+1}-u_h^{k+1}\right\|_{L^2(\Omega)}\left\|u^k-u_h^k\right\|_{L^2(\Omega)}.\end{aligned} \tag{4.14}$$

The inequality relation

$$\|R_hu^{k+1}-u_h^{k+1}\|_{L^2(\Omega)}\leq\|R_hu^{k+1}-u_h^{k+1}\|_{H^1(\Omega)}\leq\|R_hu^{k+1}-u_h^{k+1}\|_\beta$$

implies (4.12a) and (4.12b). □

Theorem 4.2. *Let u^{k+1} and u_h^{k+1} be the solutions of* (3.1) *and* (4.3)*, respectively. Then*

$$\left\|u^{k+1}-u_h^{k+1}\right\|_{H^1(\Omega)}\leq C\left(h^{\hat{m}}+\sqrt{\beta}h^{\hat{m}+1/2}\right)A+q^{k+1}\left\|u^0-u_h^0\right\|_{H^1(\Omega)}, \tag{4.15a}$$

$$\left\|u^{k+1}-u_h^{k+1}\right\|_{L^2(\Omega)}\leq C\left(h^{\hat{m}+1}+\sqrt{\beta}h^{\hat{m}+3/2}+\beta h^{\hat{m}+2}\right)A+q^{k+1}\left\|u^0-u_h^0\right\|_{L^2(\Omega)}, \tag{4.15b}$$

where

$$A=\sum_{i=1}^{k+1}q^{i-1}\left\|u^{k+2-i}\right\|_{H^{\hat{m}+1}(\Omega)}.$$

Proof. From the triangle inequality, we have

$$\begin{aligned}\left\|u^{k+1}-u_h^{k+1}\right\|_{H^1(\Omega)}&\leq\left\|u^{k+1}-R_hu^{k+1}\right\|_{H^1(\Omega)}+\left\|R_hu^{k+1}-u_h^{k+1}\right\|_{H^1(\Omega)},\\\left\|u^{k+1}-u_h^{k+1}\right\|_{L^2(\Omega)}&\leq\left\|u^{k+1}-R_hu^{k+1}\right\|_{L^2(\Omega)}+\left\|R_hu^{k+1}-u_h^{k+1}\right\|_{L^2(\Omega)}.\end{aligned}$$

Combining the errors of the projection operator R_h and Lemma 4.2, summing k implies (4.15a) and (4.15b). □

Since u_h^0 is the numerical solution for linear system (4.2), it seems that the most reasonable and straightforward option is to choose u^0 as the weak solution for the following linear system

$$a_1\left(u^0,v\right)=F(v),\quad \forall v\in H^1(\Omega). \tag{4.16}$$

Lemma 4.3. *Let u^0 and u_h^0 be the solutions of* (4.2) *and* (4.16)*, respectively. Then*

$$\left\|u^0-u_h^0\right\|_{H^1(\Omega)}\leq C\left(h^{\hat{m}}\left\|u^0\right\|_{H^{\hat{m}+1}(\Omega)}+\sqrt{\beta}h^{\hat{m}+1/2}\left\|u^0\right\|_{H^{\hat{m}+1}(\Omega)}\right). \tag{4.17}$$

In particular, when $\beta=Ch^{-1}>\frac{C_0^2}{a^0}$, an optimal error estimate in H^1 norm can be derived,

$$\left\|u^0-u_h^0\right\|_{H^1(\Omega)}\leq Ch^{\hat{m}}\left\|u^0\right\|_{H^{\hat{m}+1}(\Omega)}.$$

Proof. Using (4.2) and (4.16) and $\mathcal{M}u^0-u_h^0\in\mathbf{V}_h$

$$\begin{aligned}\left\|u^0-u_h^0\right\|_\beta^2&\leq\frac{1}{C_1}a\left(u^0-u_h^0,u-u_h\right)=\frac{1}{C_1}a\left(u^0-u_h^0,u^0-\mathcal{M}u^0+\mathcal{M}u^0-u_h^0\right)\\&=\frac{1}{C_1}a\left(u^0-u_h^0,u-\mathcal{M}u^0\right)\leq\frac{C_2}{C_1}\left\|u^0-u_h^0\right\|_\beta\left\|u^0-\mathcal{M}u^0\right\|_\beta.\end{aligned} \tag{4.18}$$

Similar to (4.9a) and (4.9b), we have

$$\begin{aligned}\left\|u^0-\mathcal{M}u^0\right\|_{L^2(\Gamma_1)}^2\leq&Ch^{2\hat{m}+1}\left\|u^0\right\|_{H^{\hat{m}+1}(\Omega)}^2,\left\|\nabla\left(u^0-\mathcal{M}u^0\right)\cdot\mathbf{n}\right\|_{L^2(\Gamma_1)}^2\\\leq&Ch^{2\hat{m}-1}\left\|u^0\right\|_{H^{\hat{m}+1}(\Omega)}^2.\end{aligned} \tag{4.19}$$

Substituting (4.18) and (4.19) into the following formula,

$$\left\|u^0-u_h^0\right\|_{H^1(\Omega)}\leq\left\|u^0-u_h^0\right\|_\beta\leq\frac{C_2}{C_1}\left\|u^0-\mathcal{M}u^0\right\|_\beta,$$

the proof is finished. □

The following estimate, sometimes regarded as an inverse assumption, plays an important role to establish the error of u^0 in L^2 norm in the analysis of the Nitsche method [29,30,32].

Lemma 4.4. *For any $v\in\mathbf{V}_h$,*

$$h\|\mathbf{a}\nabla v\|_{L^2(\Gamma_1)}^2\leq C\|\nabla v\|_{L^2(\Omega)}^2. \tag{4.20}$$

Proof. Since $v\in\mathbf{V}_h$,

$$v=\sum_{i=1}^N\Phi_i(\mathbf{x})v_i,$$

where $v_i=v(\mathbf{x}_i)$. Then

$$\begin{aligned}&\int_{\Gamma_1}(\mathbf{a}\nabla v)^2d\mathbf{x}\leq\int_{\Gamma_1}\left(a^1\sum_{i=1}^Nv_i\nabla\Phi_i(\mathbf{x})\right)^2d\mathbf{x}\\\leq&C\sum_{i=1}^Nv_i^2\int_{(\Re(\mathbf{x}_i)\cap\Omega)\cap\Gamma_1}(\nabla\Phi_i(\mathbf{x}))^2d\mathbf{x}\leq C\sum_{i=1}^Nv_i^2h^{n-3}.\end{aligned} \tag{4.21}$$

On the other hand,

$$\int_{(\Re(\mathbf{x}_i)\cap\Omega)\cap\Gamma_2} d\mathbf{x}=Ch^{n-1},\quad \int_{\Re(\mathbf{x}_i)\cap\Omega} d\mathbf{x}=Ch^n \quad \text{and} \quad C_4h^{-1}\leq\nabla\Phi_i\leq C_5h^{-1},$$

then

$$\begin{aligned}\int_\Omega(\nabla v)^2d\mathbf{x}&=\int_\Omega\left(\sum_{i=1}^N v_i\nabla\Phi_i(\mathbf{x})\right)^2d\mathbf{x}\\ &\geq C\sum_{i=1}^N v_i^2\int_{\Re(\mathbf{x}_i)\cap\Omega}(\nabla\Phi_i(\mathbf{x}))^2d\mathbf{x}\geq C\sum_{i=1}^N v_i^2h^{n-2}.\end{aligned}\tag{4.22}$$

Combining (4.21) and (4.22) yields (4.20). □

An error bound of u^0 in terms of the L^2 norm can be obtained using a duality argument.

Lemma 4.5. *Let u^0 and u_h^0 be the solutions of* (4.2) *and* (4.16)*, respectively. Assume that Ω is a convex domain or the boundary Γ of Ω is smooth, meanwhile, $Cb_1h^2<1$ and $\beta=Ch^{-1}$ based on Lemma 4.3, then*

$$\left\|u^0-u_h^0\right\|_{L^2(\Omega)}\leq Ch^{\hat{m}+1}\left\|u^0\right\|_{H^{\hat{m}+1}(\Omega)}.\tag{4.23}$$

Proof. Define the error $e=u^0-u_h^0$, the dual problem of (4.16)

$$\begin{cases}-\nabla\cdot\mathbf{a}\nabla w+bw=e & \text{in } \Omega,\\ w=0 & \text{on } \Gamma_1,\\ \mathbf{a}\nabla w\cdot\mathbf{n}=0 & \text{on } \Gamma_2.\end{cases}\tag{4.24}$$

Assume that Ω is a convex polygon or convex polyhedron, or the boundary Γ of Ω is a smooth curve, then the solution of (4.24) satisfies

$$\|w\|_{H^2(\Omega)}\leq C\|e\|_{L^2(\Omega)}.\tag{4.25}$$

We arrive at

$$(e,e)=(\mathbf{a}\nabla e,\nabla w)+(bw,e)-(\mathbf{a}\nabla w\cdot\mathbf{n},e)_{\Gamma_1}.\tag{4.26}$$

Subtracting (4.2) from (4.16) gives

$$(\mathbf{a}\nabla e,\nabla v)+(be,v)-(\mathbf{a}\nabla e\cdot\mathbf{n},v)_{\Gamma_1}-(\mathbf{a}\nabla v\cdot\mathbf{n},e)_{\Gamma_1}+\beta(e,v)_{\Gamma_1}=0,\quad\forall v\in\mathbf{V}_h,\tag{4.27}$$

Choosing $v=\mathcal{M}w$ in (4.27) and inserting (4.26),

$$\begin{aligned}(e,e)=&(\mathbf{a}\nabla e,\nabla(w-\mathcal{M}w))+(be,w-\mathcal{M}w)-(\mathbf{a}\nabla(w-\mathcal{M}w)\cdot\mathbf{n},e)_{\Gamma_1}\\ &-(\mathbf{a}\nabla e\cdot\mathbf{n},w-\mathcal{M}w)_{\Gamma_1}+\beta(e,w-\mathcal{M}w)_{\Gamma_1}.\end{aligned}\tag{4.28}$$

Applying Lemma 4.4,

$$\begin{aligned}\|\mathbf{a}\nabla e\cdot\mathbf{n}\|_{L^2(\Gamma_1)}\leq&\left\|\mathbf{a}\nabla\left(u^0-\mathcal{M}u^0\right)\cdot\mathbf{n}\right\|_{L^2(\Gamma_1)}+\left\|\mathbf{a}\nabla\left(\mathcal{M}u^0-u_h^0\right)\cdot\mathbf{n}\right\|_{L^2(\Gamma_1)}\\ \leq&Ch^{\hat{m}-1/2}\left\|u^0\right\|_{H^{\hat{m}+1}}+Ch^{-1/2}\left(\left\|\nabla\left(\mathcal{M}u^0-u^0\right)\right\|_{L^2(\Omega)}+\|\nabla e\|_{L^2(\Omega)}\right)\\ \leq&Ch^{\hat{m}-1/2}\left\|u^0\right\|_{H^{\hat{m}+1}}+Ch^{-1/2}\|\nabla e\|_{L^2(\Omega)}.\end{aligned}\tag{4.29}$$

Hence, for any $\varepsilon_1>0$

$$\begin{aligned}(\mathbf{a}\nabla e\cdot\mathbf{n},e)_{\Gamma_1}\leq&\|e\|_{L^2(\Gamma_1)}\left(Ch^{\hat{m}-1/2}\left\|u^0\right\|_{H^{\hat{m}+1}}+Ch^{-1/2}\|\nabla e\|_{L^2(\Omega)}\right)\\ \leq&\varepsilon_1h^{-1}\|e\|_{L^2(\Gamma_1)}^2+Ch^{2\hat{m}}\left\|u^0\right\|_{H^{\hat{m}+1}}^2+\frac{C}{2\varepsilon_1}\|\nabla e\|_{L^2(\Omega)}^2\\ \leq&C\left(h^{2\hat{m}}\left\|u^0\right\|_{H^{\hat{m}+1}}^2+\|\nabla e\|_{L^2(\Omega)}^2\right)+\varepsilon_1h^{-1}\|e\|_{L^2(\Gamma_1)}^2.\end{aligned}\tag{4.30}$$

Inserting the above formula into the following formula,

$$\begin{aligned}a_1(e,e)=&(\mathbf{a}\nabla e,\nabla e)-2(\mathbf{a}\nabla e\cdot\mathbf{n},e)_{\Gamma_1}+(be,e)+\beta(e,e)_{\Gamma_1}\\ \geq&C\left(\|\nabla e\|_{L^2(\Omega)}^2+\|e\|_{L^2(\Omega)}^2-h^{2\hat{m}}\left\|u^0\right\|_{H^{\hat{m}+1}}^2\right)+\left(\beta-2\varepsilon_1h^{-1}\right)\|e\|_{L^2(\Gamma_1)}^2.\end{aligned}\tag{4.31}$$

From (4.18),

$$\|e\|_\beta^2\leq\left(\frac{C_2}{C_1}\right)^2\left\|u^0-\mathcal{M}u^0\right\|_\beta^2.\tag{4.32}$$

Combining (4.31) and (4.32),

$$\begin{aligned}&C\left(\|\nabla e\|_{L^2(\Omega)}^2+\|e\|_{L^2(\Omega)}^2-h^{2\hat{m}}\left\|u^0\right\|_{H^{\hat{m}+1}}^2\right)+\left(\beta-2\varepsilon_1h^{-1}\right)\|e\|_{L^2(\Gamma_1)}^2\\ \leq&C_2\left(\frac{C_2}{C_1}\right)^2\left\|u^0-\mathcal{M}u^0\right\|_\beta^2.\end{aligned}$$

Moreover,

$$\begin{aligned}&C\left(\|\nabla e\|_{L^2(\Omega)}^2+\|e\|_{L^2(\Omega)}^2-h^{2\hat{m}}\left\|u^0\right\|_{H^{\hat{m}+1}}^2\right)+\left(\beta-2\varepsilon_1h^{-1}\right)\|e\|_{L^2(\Gamma_1)}^2\\ \leq&C\left(h^{2\hat{m}}\left\|u^0\right\|_{H^{\hat{m}+1}(\Omega)}^2+\beta h^{2\hat{m}+1}\left\|u^0\right\|_{H^{\hat{m}+1}(\Omega)}^2\right).\end{aligned}$$

Choosing an appropriate ε_1 ensures $\beta-2\varepsilon_1h^{-1}>0$ when $\beta=Ch^{-1}$ comes from Lemma 4.3,

$$\left(\beta-2\varepsilon_1h^{-1}\right)\|e\|_{L^2(\Gamma_1)}^2\leq C\left(h^{2\hat{m}}\left\|u^0\right\|_{H^{\hat{m}+1}(\Omega)}^2+\beta h^{2\hat{m}+1}\left\|u^0\right\|_{H^{\hat{m}+1}(\Omega)}^2\right).$$

Then,

$$\|e\|_{L^2(\Gamma_1)}^2\leq Ch^{2\hat{m}+1}\left\|u^0\right\|_{H^{\hat{m}+1}(\Omega)}^2.\tag{4.33}$$

Therefore,

$$\begin{aligned}(e,e)=&(\mathbf{a}\nabla e,\nabla(w-\mathcal{M}w))+(be,w-\mathcal{M}w)-(\mathbf{a}\nabla(w-\mathcal{M}w)\cdot\mathbf{n},e)_{\Gamma_1}\\&-(\mathbf{a}\nabla e\cdot\mathbf{n},w-\mathcal{M}w)_{\Gamma_1}+Ch^{-1}(e,w-\mathcal{M}w)_{\Gamma_1}\\=&I_1+I_2+I_3+I_4+I_5,\end{aligned}\tag{4.34}$$

where

$$\begin{aligned}&I_1=(\mathbf{a}\nabla e,\nabla(w-\mathcal{M}w))\leq a^1\|\nabla e\|_{L^2(\Omega)}\|\nabla(w-\mathcal{M}w)\|_{L^2(\Omega)}\\&\quad\leq Ca^1h\left(h^{\hat{m}}\left\|u^0\right\|_{H^{\hat{m}+1}(\Omega)}+C\sqrt{h^{-1}}h^{\hat{m}+1/2}\left\|u^0\right\|_{H^{\hat{m}+1}(\Omega)}\right)\|w\|_{H^2(\Omega)},\\&I_2=(be,w-\mathcal{M}w)\leq Cb_1h^2\|e\|_{L^2(\Omega)}\|w\|_{H^2(\Omega)},\\&I_3=-(\mathbf{a}\nabla(w-\mathcal{M}w)\cdot\mathbf{n},e)_{\Gamma_1}\leq a^1Ch^{\hat{m}+1/2}\left\|u^0\right\|_{H^{\hat{m}+1}(\Omega)}h^{1/2}\|w\|_{H^2(\Omega)},\\&I_4=-(\mathbf{a}\nabla e\cdot\mathbf{n},w-\mathcal{M}w)_{\Gamma_1}\leq Ca^1h^{3/2}\left(h^{\hat{m}-1/2}\|u\|_{H^{\hat{m}+1}}+h^{-1/2}\|\nabla e\|_{L^2(\Omega)}\right)\|w\|_{H^2(\Omega)},\\&I_5=\beta(e,w-\mathcal{M}w)_{\Gamma_1}\leq Ch^{-1}h^{\hat{m}+1/2}\left\|u^0\right\|_{H^{\hat{m}+1}(\Omega)}h^{3/2}\|w\|_{H^2(\Omega)},\end{aligned}$$

which together with (4.25) and (4.33) imply that (4.23) holds. □

The following theorem is a direct consequence of Lemmas 4.3, 4.5 and Theorem 4.2.

Theorem 4.3. *Let u^{k+1} and u_h^{k+1} be the solutions of* (3.1) *and* (4.3)*, respectively. Then for $q=\frac{L}{C_1}\in(0,1)$*

$$\left\|u^{k+1}-u_h^{k+1}\right\|_{H^1(\Omega)}\leq C\left(h^{\hat{m}}+\sqrt{\beta}h^{\hat{m}+1/2}\right)A_0,\tag{4.35}$$

where

$$A_0=\sum_{i=0}^{k+1}q^i\left\|u^{k+1-i}\right\|_{H^{\hat{m}+1}(\Omega)}.$$

Particularly, when $\beta=Ch^{-1}>\frac{C_0^2}{a^0}$, we obtain the optimal error in H^1 norm

$$\left\|u^{k+1}-u_h^{k+1}\right\|_{H^1(\Omega)}\leq Ch^{\hat{m}}A_0.\tag{4.36}$$

Furthermore, when $Cb_1h^2<1$ and Ω is a convex domain or the boundary Γ of Ω is smooth, the optimal error in L^2 norm can be derived

$$\left\|u^{k+1}-u_h^{k+1}\right\|_{L^2(\Omega)}\leq Ch^{\hat{m}+1}A_1.\tag{4.37}$$

The following theorem is the limiting case of Theorem 4.3.

Theorem 4.4. *Let u^{k+1} and u_h^{k+1} be the solutions of* (3.1) *and* (4.3)*, respectively. Then, u^{k+1} and u_h^{k+1} converge to u and u_h, respectively. Besides, under the conditions of Theorem* 4.3*, the optimal error in H^1 and L^2 norms are*

$$\|u-u_h\|_{H^1(\Omega)} \leq Ch^{\hat{m}}, \quad \|u-u_h\|_{L^2(\Omega)} \leq Ch^{\hat{m}+1}. \tag{4.38}$$

In proving that the operator A (or the coercivity of the $a_1(\cdot,\cdot)$) is strongly monotonic, an indispensable condition is $\beta > \frac{C_0^2}{a^0}$. According to Theorems 4.3 and 4.4, the choice of $\beta = Ch^{-1} > \frac{C_0^2}{a^0}$ can obtain the optimal convergence order. However, the exact value of C_0 is difficult to obtain. According to the assumption (A3), a feasible choice may be to consider the approximation of C_0 from the maximum eigenvalue of the following eigenvalue matrix equation [30,31]

$$\mathbf{A}\mathbf{x} = \mu \mathbf{B}\mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^N, \tag{4.39}$$

in which

$$\mathbf{A} = \left(A_{ij}\right)_{N\times N}, \quad A_{ij} = \int_{\Gamma_1} \left(\mathbf{a}\nabla\Phi_i \cdot \mathbf{n}\right)\left(\mathbf{a}\nabla\Phi_j \cdot \mathbf{n}\right) d\mathbf{x},$$
$$\mathbf{B} = \left(B_{ij}\right)_{N\times N}, \quad B_{ij} = \int_{\Omega} \nabla\Phi_i \nabla\Phi_j d\mathbf{x}.$$

Clearly,

$$C_0^2 \approx \mu_{\max} \quad \text{implies} \quad \beta = Ch^{-1} > \frac{\mu_{\max}}{a^0},$$

where $\mu_{\max}$ is the maximal (by moduli) eigenvalue of (4.39). Therefore, if taking

$$C = \theta h \frac{\mu_{\max}}{a^0}, \quad (\theta > 1),$$

in β, then

$$\beta = \theta \frac{\mu_{\max}}{a^0}, \quad \theta > 1. \tag{4.40}$$

For the semilinear elliptic problem, the condition $q = \frac{L}{C_1} \in (0,1)$ in Theorems 4.3 and 4.4 can be obtained as follows. The Lipschitz constant L can be taken as the maximum of the absolute value of the first derivative of $c(u)$ with respect to u, and

$$C_1 = \min\left(a^0 \frac{\theta-1}{\theta+1}, \frac{\theta-1}{2\theta}, b_0\right)$$

comes from choosing

$$\varepsilon = \frac{\beta + \frac{C_0^2}{a^0}}{2}$$

in the expression of

$$C_1 = \min\left(a^0 - \frac{C_0^2}{\varepsilon}, \frac{\beta-\varepsilon}{\beta}, b_0\right).$$

5 Numerical examples

In this section, we present numerical examples to investigate the influence of the value of β on precision and the order of convergence, and to illustrate the performance of the error estimates that have been proposed earlier. For semilinear elliptic problem, we set tolerance $\tau=1\text{E}-12$ as the termination condition of iteration. When the iteration is terminated at the kth step, we use u_h to represent u_h^k if no ambiguity happens.

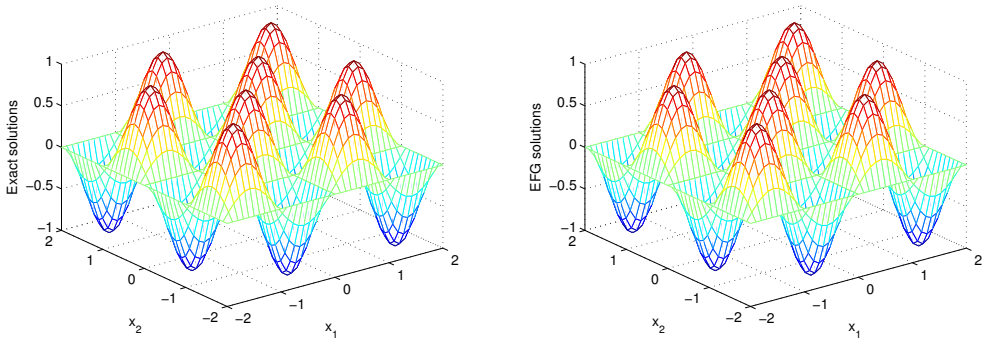
5.1 Linear elliptic case

For the first example, we consider linear elliptic case (2.1) with $c(u)=0$. Besides, $\mathbf{a}$ is a constant identity matrix, $b=1$ and $\Gamma_1=\Gamma$. The exact solution is

$$u=\sin(\pi x_1)\sin(\pi x_2),\quad (x_1,x_2)\in\Omega=[-2,2]\times[-2,2].$$

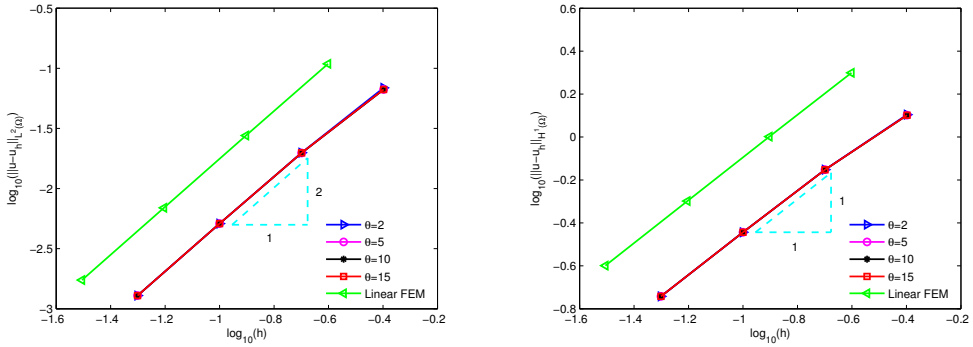
Fig. 1 displays the exact and the EFG solutions. The solutions are obtained using 41×41 equidistant nodal arrangement and the linear basis function is chosen in the MLS approximation (i.e., $\hat{m}=1$). In computation, the radius of support domain is $r(\mathbf{x})=1.5h$, and $\beta=\theta\frac{\mu_{\max}}{a^0}=2\mu_{\max}$ by $a^0=1$. Clearly, it can be found that the EFG numerical solutions with the Nitsche method are in good agreement with the exact solutions.

The log-log plots of the errors $\|u-u_h\|_{L^2(\Omega)}$ and $\|u-u_h\|_{H^1(\Omega)}$ with respect to $\theta=2,5,10,15$ are depicted in Figs. 2 and 3. In these figures, linear basis and quadratic basis are used respectively. The radius of support domain are $2.5h$ for quadratic basis. Obviously, the optimal convergence order can be obtained in the H^1 and L^2 norms, which matches the theoretical error estimates in Lemmas 4.3 and 4.5, and the value of θ in β does not have a significant impact on the accuracy and the order of convergence. For comparison, the errors of linear FEM and quadratic FEM using triangular elements are also shown in these figures. Clearly, the errors of the EFG method are much less than those of the FEM.



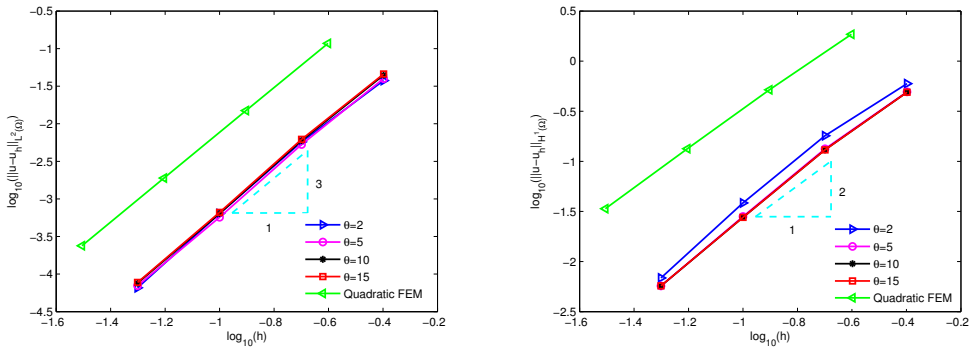
(a) (b)

Figure 1: Graphs of (a) exact solutions and (b) EFG solutions with $\beta=2\mu_{\max}$ for linear elliptic case.



(a) (b)

Figure 2: Errors of (a) $\|u-u_h\|_{L^2(\Omega)}$ and (b) $\|u-u_h\|_{H^1(\Omega)}$ for different θ with the linear basis ($\hat{m}=1$) for linear elliptic case.



(a) (b)

Figure 3: Errors of (a) $\|u-u_h\|_{L^2(\Omega)}$ and (b) $\|u-u_h\|_{H^1(\Omega)}$ for different θ with the quadratic basis ($\hat{m}=2$) for linear elliptic case.

On the other hand, Fig. 4 exhibits the graph of the condition numbers of the coefficient matrix for the variable $\theta=2,5,10,15,20,\cdots,45,50$ with the quadratic basis. The condition numbers increase with the increase of the θ, hence, a smaller θ, such as $\theta=2$, is a better

Table 1: Errors and convergence orders with $\beta=2\mu_{\max}$ for linear elliptic case.

h	Linear basis ($\hat{m}=1$)				Quadratic basis ($\hat{m}=2$)			
	$\|u-u_h\|_{L^2(\Omega)}$	Order	$\|u-u_h\|_{H^1(\Omega)}$	Order	$\|u-u_h\|_{L^2(\Omega)}$	Order	$\|u-u_h\|_{H^1(\Omega)}$	Order
4/10	6.876E-2		1.271		3.760E-2		5.958E-1	
4/20	1.986E-2	1.79	7.048E-1	0.85	5.735E-3	2.71	1.801E-1	1.72
4/40	5.114E-3	1.95	3.602E-1	0.97	6.391E-4	3.16	3.842E-2	2.22
4/80	1.288E-3	1.99	1.810E-1	0.99	6.561E-5	3.28	6.859E-3	2.48

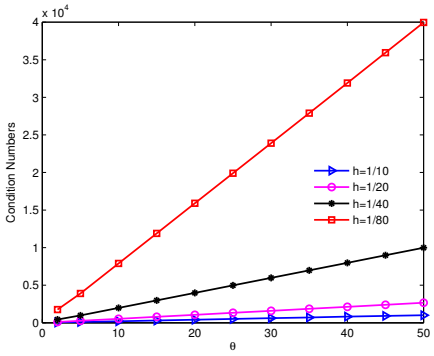


Figure 4: Condition numbers of the coefficient matrix for the variable θ with the quadratic basis ($\hat{m}=2$) for linear elliptic case.

choice for linear elliptic problem. Additionally, Table 1 gives the errors for linear basis ($\hat{m}=1$) and quadratic basis ($\hat{m}=2$). Conspicuously, the numerical results agree well with the derived theoretical analysis.

5.2 Semilinear elliptic case

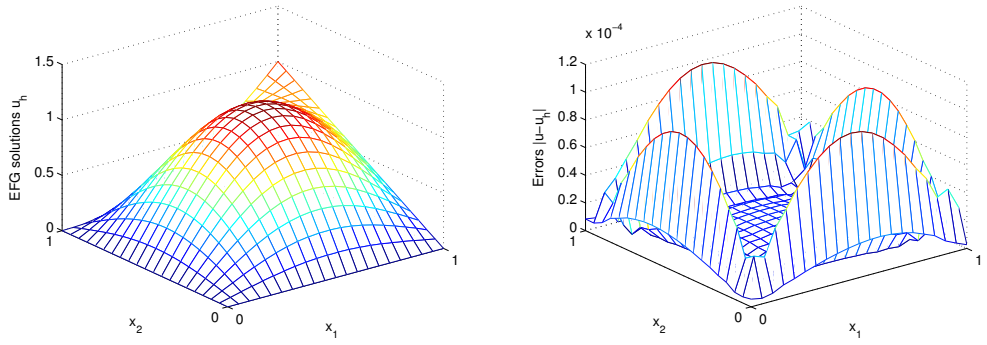
For the second example, we consider the semilinear elliptic case with $c(u)=\frac{1}{4}\sin u$ and the exact solution

$$u=x_1^2x_2+\sin(\pi x_1)\sin(\pi x_2),\quad (x_1,x_2)\in\Omega=[0,1]\times[0,1],$$

and

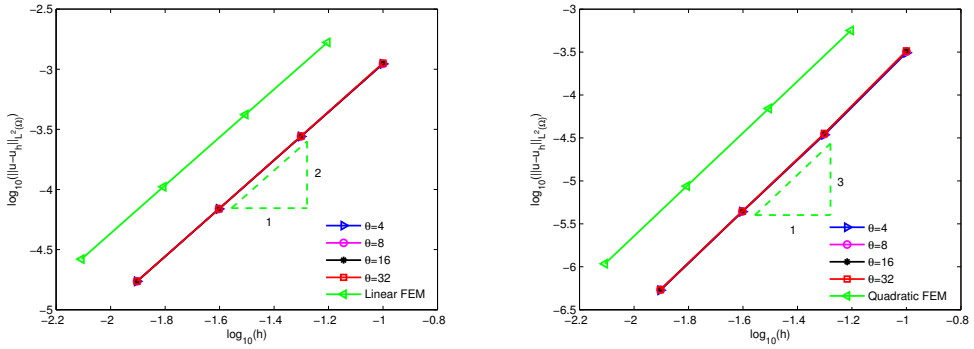
$$\mathbf{a}=\begin{bmatrix} a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix}=\begin{bmatrix} x_2^2+1 & -x_1x_2\\ -x_1x_2 & x_1^2+1\end{bmatrix},\quad b=x_1^2+x_2^3+2,$$
$$\Gamma_1=\{(x_1,x_2)\in\Omega: x_1=0\}\cup\{(x_1,x_2)\in\Omega: x_2=0\}.$$

In this case, $a^0=1$ and $b_0=2$. The Lipschitz constant $L=\frac{1}{4}$ is the maximum of the absolute value of the first derivative of $c(u)$ with respect to u, and $C_1=\frac{\theta-1}{2\theta}$, so $q=\frac{L}{C_1}\in(0,1)$ when $\theta>2$. Fig. 5 depicts the EFG solutions and the absolute errors between the exact and numerical solutions. The obtained numerical solution is based on 21×21 uniformly distributed nodes, quadratic basis ($\hat{m}=2$) and $\beta=4\mu_{\max}$. Evidently, the Nitsche method can effectively impose essential boundary conditions in the EFG method. The log-log plots of the errors $\|u-u_h\|_{L^2(\Omega)}$ with linear basis ($\hat{m}=1$) and quadratic basis ($\hat{m}=2$) for various θ are shown in Fig. 6. At the same time, the log-log plots of the errors $\|u-u_h\|_{H^1(\Omega)}$ are drawn in Fig. 7 based on the same configuration. These convergence trends keep in line with theoretical analysis, and again demonstrate that a smaller parameter θ can obtain more satisfactory numerical solutions. The numerical results of linear FEM and quadratic FEM are also given in these figures. It can be found that the EFG method has



(a) (b)

Figure 5: Graphs of (a) EFG solutions u_h and (c) errors $|u-u_h|$ with quadratic basis ($\hat{m}=2$) and $\beta=4\mu_{\max}$ for semilinear elliptic case.



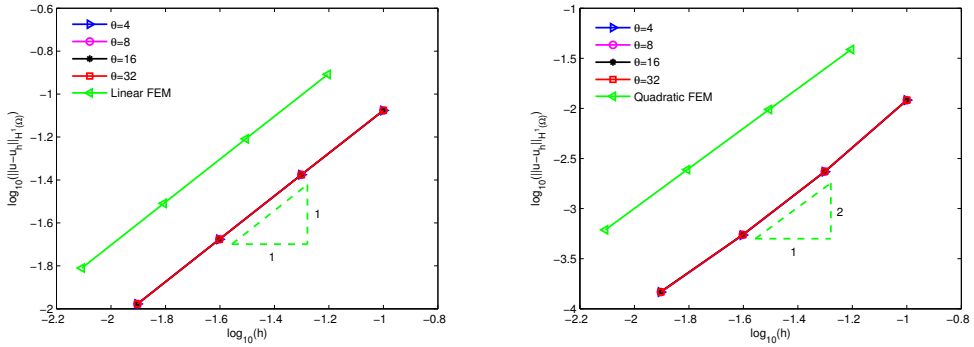
(a) (b)

Figure 6: Errors $\|u-u_h\|_{L^2(\Omega)}$ with (a) $\hat{m}=1$ and (b) $\hat{m}=2$ for semilinear elliptic case.

much better accuracy than the FEM. In addition, the numerical errors and convergence orders for linear basis and quadratic basis have been tabulated and revealed in Table 2. Apparently, these numerical results are still in good accordance with theoretical results.

Table 2: Errors and convergence orders with $\beta=4\mu_{\max}$ for semilinear elliptic case.

h	Linear basis ($\hat{m}=1$)				Quadratic basis ($\hat{m}=2$)			
	$\|u-u_h\|_{L^2(\Omega)}$	Order	$\|u-u_h\|_{H^1(\Omega)}$	Order	$\|u-u_h\|_{L^2(\Omega)}$	Order	$\|u-u_h\|_{H^1(\Omega)}$	Order
1/10	1.108E-3		8.392E-2		3.101E-4		1.209E-2	
1/20	2.761E-4	2.00	4.208E-2	1.00	3.440E-5	3.17	2.333E-3	2.37
1/40	6.894E-5	2.00	2.106E-2	1.00	4.387E-6	2.97	5.439E-4	2.10
1/80	1.722E-5	2.00	1.053E-2	1.00	5.926E-7	2.89	1.467E-4	1.89



(a) (b)

Figure 7: Errors $\|u-u_h\|_{H^1(\Omega)}$ with (a) $\hat{m}=1$ and (b) $\hat{m}=2$ for semilinear elliptic case.

6 Conclusions

In this paper, the stability and error estimates of the Nitsche-based EFG method are derived for linear and semilinear elliptic problems. The Nitsche method allows the non-interpolating shape functions to be directly used for trial and test functions by adding some boundary integral terms in the original weak form. The convergence of a simple iterative scheme for semilinear elliptic problem is analyzed, and a frequently-used inverse assumption for the EFG numerical discretization space is proved. The optimal convergence orders in the H^1 and L^2 norms are obtained for linear elliptic problem and semilinear elliptic problem with the condition $q=\frac{L}{C_1}\in(0,1)$. In numerical calculations, by solving an eigenvalue matrix equation, the Nitsche parameter β can be taken as $\beta=\theta\frac{\mu_{\max}}{a^0}$ $(\theta>1)$, which shows that β is not explicitly dependent on h. Meanwhile, the convergence condition q is also provided and verified for semilinear problem. Numerical experiments confirm the theoretical results, and show that better numerical solutions can be obtained with a small parameter θ.

Acknowledgements

This work was supported by the Innovation Research Group Project in Universities of Chongqing of China (No. CXQT19018), the National Natural Science Foundation of China (Grant No. 11971085), the Natural Science Foundation of Chongqing (Grant Nos. cstc2021jcyj-jqX0011 and cstc2020jcyj-msxm0777) and an open project of Key Laboratory for Optimization and Control Ministry of Education, Chongqing Normal University (Grant No. CSSXKFKTM202006).

References

[1] M. BADIALE, AND E. SERRA, Semilinear Elliptic Equations for Beginners, Springer-Verlag, London, 2011.

[2] Y. Z. CAO, H. T. YANG, AND L. YIN, *Finite element methods for semilinear elliptic stochastic partial differential equations*, Numer. Math., 106 (2007), pp. 181–198.

[3] J. P. ZENG, AND H. X. YU, *Error estimates of the lumped mass finite element method for semilinear elliptic problems*, J. Comput. Appl. Math., 236(7) (2012), pp. 1993–2004.

[4] Y. M. WANG, B. Y. GUO, AND W. J. WU, *Fourth-order compact finite difference methods and monotone iterative algorithms for semilinear elliptic boundary value problems*, Comput. Math. Appl., 68(12) (2014), pp. 1671–1688.

[5] Z. G. XIONG, AND Y. P. CHEN, *A rectangular finite volume element method for a semilinear elliptic equation*, J. Sci. Comput., 36(2) (2008), pp. 177–191.

[6] P. HOUSTON, AND T. WIHLER, *An hp-adaptive Newton-discontinuous-Galerkin finite element approach for semilinear elliptic boundary value problems*, Math. Comput., 87(314) (2018), pp. 2641–2674.

[7] H. Y. HU, AND J. S. CHEN, *Radial basis collocation method and quasi-Newton iteration for nonlinear elliptic problems*, Numer. Meth. Partial Differential Equations, 24(3) (2010), pp. 991–1017.

[8] H. WENDLAND, AND J. KÜNEMUND, *Solving partial differential equations on (evolving) surfaces with radial basis functions*, Adv. Comput. Math., 46(4) (2020), p. 64.

[9] F. UREÑA, L. GAVETEB, A. GARCÍA, J. J. BENITOA, AND A. M. VARGAS, *Solving second order non-linear hyperbolic PDEs using generalized finite difference method (GFDM)*, J. Comput. Appl. Math., 363 (2020), pp. 1–21.

[10] Z. Y. ZHENG, AND X. L. LI, *Theoretical analysis of the generalized finite difference method*, Comput. Math. Appl., 120 (2022), pp. 1–14.

[11] S. H. LI, AND W. K. LIU, Meshfree Particle Methods, Springer, Berlin, 2004.

[12] G. R. LIU, Meshfree Methods: Moving beyond the Finite Element Method, 2nd edn., CRC, Boca Raton, 2009.

[13] Y. M. CHENG, Meshless Methods, Science Press, Beijing, 2015.

[14] T. BELYTSCHKO, Y. Y. LU, AND L. GU, *Element-free Galerkin methods*, Int. J. Numer. Methods Eng., 37(2) (1994), pp. 229–256.

[15] P. LANCASTER, AND K. SALKAUSKAS, *Surfaces generated by moving least squares methods*, Math. Comput., 37(155) (1981), pp. 141–158.

[16] X. L. LI, *Error estimates for the moving least-square approximation and the element-free Galerkin method in n-dimensional spaces*, Appl. Numer. Math., 99 (2016), pp. 77–97.

[17] X. L. LI, AND S. L. LI, *On the stability of the moving least squares approximation and the element-free Galerkin method*, Comput. Math. Appl., 72(6) (2016), pp. 1515–1531.

[18] F. X. SUN, J. F. WANG, Y. M. CHENG, AND A. X. HUANG, *Error estimates for the interpolating moving least-squares method in n-dimensional space*, Appl. Numer. Math., 98 (2015), pp. 79–105.

[19] T. ZHANG, AND X. L. LI, *Variational multiscale interpolating element-free Galerkin method for the nonlinear Darcy-Forchheimer model*, Comput. Math. Appl., 72 (2019), pp. 363–377.

[20] J. F. WANG, F. X. SUN, AND Y. M. CHENG, *An improved interpolating element-free Galerkin method with a nonsingular weight function for two-dimensional potential problems*, Chin. Phys. B, 21 (2012), 090204.

[21] J. S. WAN, AND X. L. LI, *Analysis of a superconvergent recursive moving least squares approximation*, Appl. Math. Lett., 133 (2022), 108223.

[22] M. DEHGHAN, AND M. ABBASZADEH, *A reduced proper orthogonal decomposition (POD) ele-*

ment free Galerkin (POD-EFG) method to simulate two-dimensional solute transport problems and error estimate, Appl. Numer. Math., 126 (2018), pp. 92–112.

[23] V. MOHAMMADI, AND M. DEHGHAN, *Simulation of the phase field Cahn-Hilliard and tumor growth models via a numerical scheme: element-free Galerkin method*, Comput. Methods Appl. Mech. Eng., 345 (2019), pp. 919–950.

[24] M. ABBASZADEH, AND M. DEHGHAN, *Investigation of heat transport equation at the microscale via interpolating element-free Galerkin method*, Eng. Comput-Germany, (2021), pp. 1–17.

[25] M. DEHGHAN, AND M. ABBASZADEH, *Error analysis and numerical simulation of magneto-hydrodynamics (MHD) equation based on the interpolating element free Galerkin (IEFG) method*, Appl. Numer. Math., 137 (2019), pp. 252–273.

[26] M. ABBASZADEH, AND M. DEHGHAN, *The interpolating element-free Galerkin method for solving Korteweg-de Vries-Rosenau-regularized long-wave equation with error analysis*, Nonlinear Dyn., 96(2) (2019), pp. 1345–1365.

[27] T. ZHANG, AND X. L. LI, *Analysis of the element-free Galerkin method with penalty for general second-order elliptic problems*, Appl. Math. Comput., 380 (2020), 125306.

[28] T. ZHANG, X. L. LI, AND L. W. XU, *Error analysis of an implicit Galerkin meshfree scheme for general second-order parabolic problems*, Appl. Numer. Math., 177 (2022), pp. 58–78.

[29] J. NITSCHE, *über ein variations zur löung von dirichlet-problemen bei verwendung von teilräumen die keinen randbedingungen unterworfen sind*, Abh. Math. Se. Univ., 36 (1970), pp. 9–15.

[30] M. GRIEBEL, AND M. A. SCHWEITZER, A particle-partition of unity method. Part V: Boundary conditions, in: S. Hildebrandt, H. Karcher (Eds.), Geometric Analysis and Nonlinear Partial Differential Equations, Springer, Berlin, 2002, pp. 517–540.

[31] S. FERNÁNDEZ-MÉNDEZ, AND A. HUERTA, *Imposing essential boundary conditions in mesh-free methods*, Comput. Methods Appl. Mech. Eng., 193(12-14) (2004), pp. 1257–1275.

[32] Q. H. ZHANG, *Quadrature for meshless Nitsche's method*, Numer. Meth. Partial Differential Equations, 30(1) (2014), pp. 265–288.

[33] J. C. WU, AND D. D. WANG, *An accuracy analysis of Galerkin meshfree methods accounting for numerical integration*, Comput. Methods Appl. Mech. Eng., 375 (2021), 113631.

[34] X. L. LI, *Theoretical analysis of the reproducing kernel gradient smoothing integration technique in Galerkin meshless methods*, J. Comput. Math., 41(3) (2023), pp. 502–524.

[35] X. L. LI, AND S. L. LI, *A fast element-free Galerkin method for the fractional diffusion-wave equation*, Appl. Math. Lett., 112 (2021), 106724.

[36] C. GROSSMANN, H. G. ROOS, AND M. STYNES, Numerical Treatment of Partial Differential Equations, Springer, Berlin, 2007.