
ELSEVIER Applied Mathematics and Computation 92 (1998) 9–27

APPLIED MATHEMATICS AND COMPUTATION

A unified analysis of a weighted least squares method for first-order systems [1]

Suh-Yuh Yang *, Jinn-Liang Liu [2]

Department of Applied Mathematics, National Chiao Tung University, Hsinchu 30050, Taiwan

Abstract

A unified analysis of a weighted least squares finite element method (WLSFEM) for approximating solutions of a large class of first-order differential systems is proposed. The method exhibits several advantageous features. For example, the trial and test functions are not required to satisfy the boundary conditions. Its discretization results in symmetric and positive definite algebraic systems with condition number $O(h^{-2} + w^2)$. And a single piecewise polynomial finite element space may be used for all test and trial functions. Asymptotic convergence of the least squares approximations with suitable weights is established in a natural norm without requiring extra smoothness of the solutions. If, instead, the solutions are sufficiently regular, a priori error estimates can be derived under two suitable assumptions which are related respectively to the symmetric positive systems of Friedrichs and first-order Agmon–Douglis–Nirenberg (ADN) elliptic systems. Numerous model problems fit into these two important systems. Some selective examples are examined and verified in the unified framework.

Keywords: Boundary value problems; First-order systems; Friedrichs' systems; ADN elliptic systems; Least squares methods; Convergence; Error estimates

* Corresponding author. E-mail: syyang@math.nctu.edu.tw.

[1] This work was supported by NSC-grant 85-2121-M-009-014, Taiwan, ROC.

[2] E-mail: jinnliu@math.nctu.edu.tw.

0096-3003/98/$19.00
PII: S0096-3003(97)10046-7

1. Introduction

The purpose of this paper is to give a unified analysis of a weighted least squares finite element method (WLSFEM) applied to a large class of first-order differential systems. The Friedrichs symmetric positive systems [24] and the first-order Agmon–Douglis–Nirenberg (ADN) elliptic systems [1] are of particular interest in this general framework.

Although there has been considerable attention to the use of least squares principles in connection with finite element applications during the last decade, the modern theory of least squares finite element methods (LSFEMs) for the approximate solution of elliptic boundary value problems dates back at least from the work of Bramble and Schatz [7,8] in 1970. In Refs. [7,8], the approximate solution is defined to be the minimizer of a least squares functional over a finite-dimensional approximating function (trial function) space. The functional consists of a weighted sum of the residuals occurring in the differential equation and the boundary condition. This method has the feature that the trial and test functions are not required to satisfy the boundary condition. On the other hand, it requires that the trial and test functions are smooth enough to lie in the domain of the elliptic operator. For example, they must be in the space $H^{2m}(\Omega)$ for a $2m$th-order problem. Thus, many seemingly natural finite elements are never admissible. However, this difficulty may be circumvented by introducing the derivatives of the unknown function as new dependent variables (in general, the combinations of these new dependent variables present certain physical meanings such as, fluxes, vorticity, and stresses, etc.), then the original higher order problem can be reformulated as a system of differential equations of first-order with possibly additional compatibility equations. Applying the least squares principles on this extended first-order system, the smoothness requirement on the trial and test function spaces can then be relaxed, which eliminates the main disadvantage of this approach.

The least squares approach to boundary value problems of first-order systems represents a fairly general methodology that can produce a variety of algorithms. Thus, various LSFEMs appeared in the literature. Roughly speaking, according to the boundary treatment, these methods can be classified into the following two categories.

- The least squares functional involves only the residuals in the differential equations. In this case, the trial and test functions are required to fulfill the homogeneous boundary conditions and thus more than L^2 regularity, say $H^{1/2}$, for the given boundary functions is necessary in the nonhomogeneous cases. See, e.g., Refs. [5,11,13,14,17–20,22,23,26,35].
- The least squares functional consists both of the residuals in the differential equations and the boundary conditions. The trial and test functions need not satisfy the boundary conditions. Hence, only L^2 boundary data is required whenever the problem is well posed. See, e.g., Refs. [2,4,12,15,16,25,37].

Both types of least squares functionals can be combined with the weighting techniques to enhance the stability and accuracy of the approximate solution, even allowing different equations and boundary conditions equipped with different weights [2,5].

Motivated by the WLSFEM of Aziz and Liu [4], we generalize the method (of the second category) in a unified framework for both Friedrichs' and ADN systems. More specifically, the method is applied to the boundary value problems of first-order systems in the general form:

$$\mathscr{L}\mathbf{u} := \sum_{i=1}^{d} A_i \frac{\partial \mathbf{u}}{\partial x_i} + A_0 \mathbf{u} = \mathbf{f} \quad \text{in } \Omega, \tag{1}$$

$$\mathscr{B}\mathbf{u} := B\mathbf{u} = \mathbf{g} \quad \text{on } \partial\Omega, \tag{2}$$

where $\Omega \subset \mathbb{R}^d, d \geqslant 2$, is a bounded domain with a smooth boundary $\partial\Omega$, and $\mathbf{u} = (u_1, \ldots, u_m)^t$, $\mathbf{f} = (f_1, \ldots, f_m)^t$, $\mathbf{g} = (g_1, \ldots, g_n)^t$. In the sequel, we shall always assume that the entries of $m \times m$ matrices $A_i \in [L^\infty(\Omega)]^{m\times m}$, $0 \leqslant i \leqslant d$, and of $n \times m$ boundary matrix $B \in [L^\infty(\partial\Omega)]^{n\times m}$ are regular enough on $\overline{\Omega}$ and $\partial\Omega$, respectively, such that problem (1), (2) has a unique solution $\mathbf{u} \in [H^1(\Omega)]^m$ with the given functions $\mathbf{f} \in [L^2(\Omega)]^m$, $\mathbf{g} \in [L^2(\partial\Omega)]^n$.

LSFEMs offer many attractive features in practice when applied to boundary value problems formulated in first-order systems, we refer to the references mentioned in the above. We summarize our results as follows.

- With a minimum regularity of the (known or unknown) functions as posed in Eqs. (1) and (2), asymptotic convergence of the approximate solutions obtained by the WLSFEM is given for the general problem (1) and (2).
- Under suitable assumptions ((19) and (20) in Section 4), an analysis of a priori estimates for both Friedrichs' and ADN systems is presented. In particular, the recent works on LSFEMs for the Stokes equations by Bochev and Gunzburger [5], Chang et al. [18,19], and Jiang and Chang [26] may be extended by using the WLSFEM. Consequently, the regularity requirement on the boundary conditions can be lessened and the trial and test functions are not required to satisfy the boundary conditions. However, it is not clear that the estimates are sharp under the general assumptions. If, in addition, stronger conditions such as, e.g., that of Ref. [37] are met, optimal convergence can be expected for certain systems.
- The condition number of the resulting system of linear equations is $O(h^{-2} + w^2)$, where h denotes the mesh parameter and w the weighting parameter.
- The framework is independent of the type of differential systems, i.e., it is for elliptic, parabolic, hyperbolic, or mixed type problems.

The remainder of the paper is organized as follows. Some notation and preliminary results will be introduced in Section 2. The WLSFEM is presented in Section 3 with its fundamental properties and asymptotic convergence result.

A priori estimates are derived in Section 4 under the two general assumptions. Two model problems, namely, the neutron transport equation and the Stokes equations cast in the framework of Eqs. (1) and (2) are examined in Section 5 to validate the assumptions. An estimate for the condition number of the resulting symmetric positive definite matrix is given in Section 6. Finally, some concluding remarks are drawn in Section 7.

2. Notation and preliminaries

Throughout this paper, we shall require some function spaces defined on Ω and $\partial\Omega$ [33]. The classical Sobolev spaces $H^s(\Omega), s \geqslant 0$ integer, and $L^2(\partial\Omega)$ with their associated inner products $(\cdot,\cdot)_{s,\Omega}, (\cdot,\cdot)_{0,\partial\Omega}$ and norms $\|\cdot\|_{s,\Omega}, \|\cdot\|_{0,\partial\Omega}$ are employed. As usual, $L^2(\Omega) := H^0(\Omega)$. For the Cartesian product spaces $[H^s(\Omega)]^m$ and $[L^2(\partial\Omega)]^n$, the corresponding inner products and norms are also denoted by $(\cdot,\cdot)_{s,\Omega}, (\cdot,\cdot)_{0,\partial\Omega}$ and $\|\cdot\|_{s,\Omega}, \|\cdot\|_{0,\partial\Omega}$, respectively, when there is no chance for confusion.

By $L^\infty(\Omega)$ and $L^\infty(\partial\Omega)$ we denote the usual Banach spaces of measurable and essentially bounded real-valued functions defined on Ω and $\partial\Omega$ with the norms $\|\cdot\|_{\infty,\Omega}$ and $\|\cdot\|_{\infty,\partial\Omega}$, respectively.

Since the boundary $\partial\Omega$ of the bounded domain Ω is smooth, there exists an operator $\gamma_0 : H^1(\Omega) \to L^2(\partial\Omega)$, linear and continuous, such that

$$\gamma_0 v = \text{restriction of } v \text{ on } \partial\Omega \text{ for every } v \in C^1(\overline{\Omega}).$$

The space $\gamma_0(H^1(\Omega))$ is not the whole space $L^2(\partial\Omega)$, it is denoted by $H^{1/2}(\partial\Omega)$ and define its norm by

$$\|\varphi\|_{1/2,\partial\Omega} = \inf\{\|v\|_{1,\Omega};\ v \in H^1(\Omega), \gamma_0 v = \varphi\},$$

which makes it a Hilbert space. Also, the associated norm of the product space $[H^{1/2}(\partial\Omega)]^n$ is still denoted by $\|\cdot\|_{1/2,\partial\Omega}$.

Define the following bilinear form and linear form: for any $\mathbf{v}, \mathbf{w} \in [H^1(\Omega)]^m$,

$$a_w(\mathbf{v},\mathbf{w}) = (\mathscr{L}\mathbf{v}, \mathscr{L}\mathbf{w})_{0,\Omega} + w(\mathscr{B}\mathbf{v}, \mathscr{B}\mathbf{w})_{0,\partial\Omega}, \tag{3}$$

$$\ell_w(\mathbf{v}) = (\mathbf{f}, \mathscr{L}\mathbf{v})_{0,\Omega} + w(\mathbf{g}, \mathscr{B}\mathbf{v})_{0,\partial\Omega}, \tag{4}$$

where w is a positive weight maybe depending on the mesh parameter h which will be introduced later. It is easily seen that, for each weight w, $a_w(\cdot,\cdot)$ defines an inner product on the space $[H^1(\Omega)]^m \times [H^1(\Omega)]^m$, and the reduced norm shall be given by

$$\|\mathbf{v}\|_{a_w}^2 = a_w(\mathbf{v},\mathbf{v}) \quad \forall \mathbf{v} \in [H^1(\Omega)]^m. \tag{5}$$

Note that the homogeneous property of $\|\cdot\|_{a_w}$ is ensured by the fact that problem (1) and (2) possesses a unique solution in $[H^1(\Omega)]^m$ for given functions $\mathbf{f} \in [L^2(\Omega)]^m$, $\mathbf{g} \in [L^2(\partial\Omega)]^n$.

To approximate (1) and (2), we consider a regular family [21] of triangulations $\{\mathscr{T}_h : 0 < h < 1\}$ of $\overline{\Omega}$, where the parameter h measures the mesh size of each discretization. For each h, define the finite element space $\mathscr{V}_{h,p} \subset [H^1(\Omega)]^m$, $p \geqslant 1$ integer, which is assumed to possess the following approximation property: for any $\mathbf{v} \in [H^{p+1}(\Omega)]^m$ there exists $\mathbf{v}_{h,p} \in \mathscr{V}_{h,p}$ such that

$$\|\mathbf{v} - \mathbf{v}_{h,p}\|_{0,\Omega} + h\|\mathbf{v} - \mathbf{v}_{h,p}\|_{1,\Omega} \leqslant C_1 h^{p+1} \|\mathbf{v}\|_{p+1,\Omega}, \tag{6}$$

where C_1 is a positive constant independent of $\mathbf{v}$ and h.

Using the following lemma, further approximation properties of the finite element space $\mathscr{V}_{h,p}$ can be deduced.

Lemma 2.1. *There is a positive constant C_2 such that, for any $\mathbf{v} \in [H^1(\Omega)]^m$ and any $\varepsilon > 0$,*

$$\|\mathbf{v}\|_{0,\partial\Omega} \leqslant C_2 \left(\varepsilon \|\mathbf{v}\|_{1,\Omega} + \frac{1}{\varepsilon} \|\mathbf{v}\|_{0,\Omega} \right). \tag{7}$$

A proof of Lemma 2.1 can be found in, for example, Refs. [9,33]. With (6) and (7), we immediately have:

Lemma 2.2. *Let $\mathbf{v} \in [H^{p+1}(\Omega)]^m$, $p \geqslant 1$ integer. Then there exists $\mathbf{v}_{h,p} \in \mathscr{V}_{h,p}$ such that, for any $\varepsilon > 0$,*

$$\|\mathbf{v} - \mathbf{v}_{h,p}\|_{0,\partial\Omega} \leqslant C_3 \left(\varepsilon h^p + \frac{1}{\varepsilon} h^{p+1} \right) \|\mathbf{v}\|_{p+1,\Omega}, \tag{8}$$

where C_3 is a positive constant independent of $\mathbf{v}$, ε, and h.

Lemma 2.3. *Let $\mathbf{v} \in [H^{p+1}(\Omega)]^m$, $p \geqslant 1$ integer. Then there exists $\mathbf{v}_{h,p} \in \mathscr{V}_{h,p}$ such that*

$$\|\mathbf{v} - \mathbf{v}_{h,p}\|_{a_w} \leqslant C_4 (h^p + wh^{p+1}) \|\mathbf{v}\|_{p+1,\Omega} \tag{9}$$

for some positive constant C_4 independent of $\mathbf{v}$, w, and h.

Proof. Let $\mathbf{v}_{h,p}$ be the same as in (8) with $\varepsilon = 1/\sqrt{w}$, by (5) and (3),

$$\begin{aligned} \|\mathbf{v} - \mathbf{v}_{h,p}\|_{a_w} &\leqslant \|\mathscr{L}(\mathbf{v} - \mathbf{v}_{h,p})\|_{0,\Omega} + \sqrt{w} \|\mathscr{B}(\mathbf{v} - \mathbf{v}_{h,p})\|_{0,\partial\Omega} \\ &\leqslant C_5 (\|\mathbf{v} - \mathbf{v}_{h,p}\|_{1,\Omega} + \sqrt{w} \|\mathbf{v} - \mathbf{v}_{h,p}\|_{0,\partial\Omega}) \\ &\leqslant C_6 \big(h^p + (h^p + wh^{p+1})\big) \|\mathbf{v}\|_{p+1,\Omega}, \end{aligned} \tag{10}$$

where the second inequality is ensured by the fact that $\mathscr{L}$ is a first-order differential operator and $A_i \in [L^\infty(\Omega)]^{m \times m}$, $0 \leqslant i \leqslant d$, $B \in [L^\infty(\partial\Omega)]^{n \times m}$. This completes the proof. □

3. Weighted least squares approximations

Define a weighted least squares functional $\mathscr{J}: [H^1(\Omega)]^m \to \mathbb{R}$ as

$$\mathscr{J}(\mathbf{v}) = \|\mathscr{L}\mathbf{v} - \mathbf{f}\|_{0,\Omega}^2 + w\|\mathscr{B}\mathbf{v} - \mathbf{g}\|_{0,\partial\Omega}^2, \tag{11}$$

where w is the same parameter of (3) and (4). Evidently, the exact solution $\mathbf{u} \in [H^1(\Omega)]^m$ of problem (1) and (2) minimizes the functional and vice versa, i.e.,

$$\mathscr{J}(\mathbf{u}) = \min_{\mathbf{v}\in[H^1(\Omega)]^m} \mathscr{J}(\mathbf{v}). \tag{12}$$

Taking the first variation, the solution equivalently satisfies the equation

$$a_w(\mathbf{u}, \mathbf{v}) = \ell_w(\mathbf{v}) \quad \forall \mathbf{v} \in [H^1(\Omega)]^m, \tag{13}$$

where the bilinear form and linear form are given in (3) and (4), respectively.

The WLSFEM for problem (1) and (2) is then to find $\mathbf{u}_{h,p}^w \in \mathscr{V}_{h,p}$ such that

$$a_w(\mathbf{u}_{h,p}^w, \mathbf{v}_{h,p}) = \ell_w(\mathbf{v}_{h,p}) \quad \forall \mathbf{v}_{h,p} \in \mathscr{V}_{h,p}. \tag{14}$$

Note that the trial and test functions are not required to satisfy the boundary conditions in the approximation.

We first have the following results concerning existence, uniqueness, stability estimates, and some important properties of the approximate solution.

Theorem 3.1. *Let* $\mathbf{u}$ *be the exact solution of problem (1) and (2) with* $\mathbf{f} \in [L^2(\Omega)]^m$, $\mathbf{g} \in [L^2(\partial\Omega)]^n$.

(i) Problem (14) has a unique solution $\mathbf{u}_{h,p}^w \in \mathscr{V}_{h,p}$ *for each given positive weight* w, *and the solution satisfies the following stability estimate:*

$$\|\mathbf{u}_{h,p}^w\|_{a_w} \leqslant \|\mathbf{f}\|_{0,\Omega} + \sqrt{w}\|\mathbf{g}\|_{0,\partial\Omega}. \tag{15}$$

(ii) The matrix of the linear algebraic system associated with problem (14) is symmetric and positive definite.

(iii) The following orthogonality relation holds:

$$a_w(\mathbf{u} - \mathbf{u}_{h,p}^w, \mathbf{v}_{h,p}) = 0 \quad \forall \mathbf{v}_{h,p} \in \mathscr{V}_{h,p}. \tag{16}$$

(iv) The approximate solution $\mathbf{u}_{h,p}^w$ *is a best approximation of* $\mathbf{u}$ *in the* $\|\cdot\|_{a_w}$*-norm, that is,*

$$\|\mathbf{u} - \mathbf{u}_{h,p}^w\|_{a_w} = \inf_{\mathbf{v}_{h,p}\in\mathscr{V}_{h,p}} \|\mathbf{u} - \mathbf{v}_{h,p}\|_{a_w}. \tag{17}$$

Proof. To prove the unique solvability, it suffices to prove the uniqueness of solution since the finite dimensionality of $\mathscr{V}_{h,p}$. Let $\mathbf{u}_{h,p}^w$ be a solution of problem (14) then, by (5) and (4) and the Cauchy–Schwarz inequality,

$$\begin{aligned}\|\mathbf{u}_{h,p}^w\|_{a_w}^2 &= a_w(\mathbf{u}_{h,p}^w, \mathbf{u}_{h,p}^w)\\ &= (\mathbf{f}, \mathscr{L}\mathbf{u}_{h,p}^w)_{0,\Omega} + w(\mathbf{g}, \mathscr{B}\mathbf{u}_{h,p}^w)_{0,\partial\Omega}\\ &\leqslant \|\mathbf{f}\|_{0,\Omega}\|\mathscr{L}\mathbf{u}_{h,p}^w\|_{0,\Omega} + w\|\mathbf{g}\|_{0,\partial\Omega}\|\mathscr{B}\mathbf{u}_{h,p}^w\|_{0,\partial\Omega}\\ &\leqslant \|\mathbf{f}\|_{0,\Omega}\|\mathbf{u}_{h,p}^w\|_{a_w} + \sqrt{w}\|\mathbf{g}\|_{0,\partial\Omega}\|\mathbf{u}_{h,p}^w\|_{a_w}.\end{aligned}$$

Thus, we obtain (15). Consequently, the solution $\mathbf{u}_{h,p}^w$ of problem (14) is unique.

Part (ii) follows from the fact that the bilinear form $a_w(\cdot,\cdot)$ is symmetric and positive definite.

Part (iii) follows easily from Eqs. (13) and (14), since $\mathscr{V}_{h,p}$ is a subspace of $[H^1(\Omega)]^m$.

Finally, to prove (iv), by (16) and the Cauchy–Schwarz inequality,

$$\begin{aligned}\|\mathbf{u} - \mathbf{u}_{h,p}^w\|_{a_w}^2 &= a_w(\mathbf{u} - \mathbf{u}_{h,p}^w, \mathbf{u} - \mathbf{u}_{h,p}^w)\\ &= a_w(\mathbf{u} - \mathbf{u}_{h,p}^w, \mathbf{u} - \mathbf{v}_{h,p})\\ &\leqslant \|\mathbf{u} - \mathbf{u}_{h,p}^w\|_{a_w}\|\mathbf{u} - \mathbf{v}_{h,p}\|_{a_w}\end{aligned}$$

for all $\mathbf{v}_{h,p} \in \mathscr{V}_{h,p}$. This completes the proof. □

As a consequence of part (iv) in Theorem 3.1, we have the following asymptotic convergence.

Theorem 3.2. *Suppose that the positive weight w in the least squares approximation Eq. (14) is a constant or a bounded mesh-dependent function in $h \in (0,1)$. Then we have*

$$\lim_{h\to 0}\|\mathbf{u} - \mathbf{u}_{h,p}^w\|_{a_w} = 0. \tag{18}$$

Proof. Without loss of generality, let constant $C_7 > 0$ represent an upper bound of $\sqrt{w}$ on $(0,1)$. Let $\mathscr{D}(\overline{\Omega})$ denote the linear space of infinitely differentiable functions on Ω such that all the derivatives have continuous extensions to $\partial\Omega$. Since $[\mathscr{D}(\overline{\Omega})]^m$ is dense in $[H^1(\Omega)]^m$ with respect to the $\|\cdot\|_{1,\Omega}$-norm, for any $\epsilon > 0$, there exists $\mathbf{u}^* \in [\mathscr{D}(\overline{\Omega})]^m$ independent of h such that

$$\|\mathbf{u} - \mathbf{u}^*\|_{1,\Omega} < \frac{\varepsilon}{2C_5(1 + 2C_2C_7)},$$

which implies (cf. (10) and (7))

$$\|\mathbf{u} - \mathbf{u}^*\|_{a_w} \leqslant C_5(1 + 2C_2C_7)\|\mathbf{u} - \mathbf{u}^*\|_{1,\Omega} < \frac{\varepsilon}{2}.$$

For this fixed smooth function $\mathbf{u}^* \in [\mathscr{D}(\overline{\Omega})]^m$, by (6), we can find $\Pi_{h,p}\mathbf{u}^* \in \mathscr{V}_{h,p}$ so that

$$\|\mathbf{u}^* - \Pi_{h,p}\mathbf{u}^*\|_{1,\Omega} \leqslant C_1 h^p \|\mathbf{u}^*\|_{p+1,\Omega},$$

which implies, for sufficiently small h,

$$\|\mathbf{u}^* - \Pi_{h,p}\mathbf{u}^*\|_{a_w} \leqslant C_5(1 + 2C_2C_7)\|\mathbf{u}^* - \Pi_{h,p}\mathbf{u}^*\|_{1,\Omega} < \frac{\varepsilon}{2}.$$

By Eq. (17), we immediately obtain

$$0 \leqslant \|\mathbf{u} - \mathbf{u}_{h,p}^w\|_{a_w} \leqslant \|\mathbf{u} - \Pi_{h,p}\mathbf{u}^*\|_{a_w} \leqslant \|\mathbf{u} - \mathbf{u}^*\|_{a_w} + \|\mathbf{u}^* - \Pi_{h,p}\mathbf{u}^*\|_{a_w} < \epsilon.$$

This completes the proof. □

4. Error estimates

Theorem 3.2 indicates that, for example $w = \text{const}$, the approximate solution $\mathbf{u}_{h,p}^w$ satisfies the differential equations and the boundary conditions asymptotically in the $\|\cdot\|_{0,\Omega}$-norm and the $\|\cdot\|_{0,\partial\Omega}$-norm, respectively, without assuming additional regularity assumption on $\mathbf{u}$, that is,

$$\|\mathscr{L}\mathbf{u}_{h,p}^w - \mathbf{f}\|_{0,\Omega} \to 0 \quad \text{as } h \to 0,$$

$$\|\mathscr{B}\mathbf{u}_{h,p}^w - \mathbf{g}\|_{0,\partial\Omega} \to 0 \quad \text{as } h \to 0.$$

Of course, one may expect better convergence properties for the approximation provided that the exact solution is sufficiently regular and that the system (1) and (2) satisfies certain coercivity conditions. In fact, these conditions associated with some specific numerical methods are often circumstantial and hence somewhat restrictive to a wider class of problems. For example, the LSFEMs of the references cited in the second category in Section 1 are similar in principle and yet quite different in terms of the coercivity conditions or some related approximation assumptions. On the other hand, an attempt to create a universal conditions for the general system (1) and (2) in the context of LSFE approximation is very intractable if not impossible. Nevertheless, with WLSFEM (14), we classify the conditions for the Friedrichs and ADN systems by the following two respective assumptions.

(H1) There exists a constant $C_8 > 0$ such that:

$$\|\mathbf{v}\|_{0,\Omega} \leqslant C_8(\|\mathscr{L}\mathbf{v}\|_{0,\Omega} + \|\mathscr{B}\mathbf{v}\|_{0,\partial\Omega}) \quad \forall \mathbf{v} \in [H^1(\Omega)]^m. \tag{19}$$

(H2) There exists a constant $C_9 > 0$ such that:

$$\|\mathbf{v}\|_{1,\Omega} \leqslant C_9(\|\mathscr{L}\mathbf{v}\|_{0,\Omega} + \|\mathscr{B}\mathbf{v}\|_{1/2,\partial\Omega}) \quad \forall \mathbf{v} \in [H^1(\Omega)]^m. \tag{20}$$

Associated with the assumption (H2), we also need the following inverse assumption [21] on the finite element space $\mathscr{V}_{h,p}$: there exists a constant $C_{10} > 0$ such that

$$\|\mathscr{B}\mathbf{v}_{h,p}\|_{1/2,\partial\Omega} \leqslant C_{10}h^{-1/2}\|\mathscr{B}\mathbf{v}_{h,p}\|_{0,\partial\Omega} \quad \forall \mathbf{v}_{h,p} \in \mathscr{V}_{h,p}. \tag{21}$$

This type of assumption is commonly used in WLSFEMs (see, e.g., Refs. [2,15,25,36,37]) and holds for a large class of finite element spaces $\mathscr{V}_{h,p}$. More precisely, if the family $\{\mathscr{T}_h\}$ of triangulations of $\overline{\Omega}$ is quasi-uniform [21,27], i.e., there exists a positive constant v independent of h such that

$$h \leqslant v\,\mathrm{diam}(\Omega_i^h) \quad \forall \Omega_i^h \in \mathscr{T}_h, \ \ \mathscr{T}_h \in \{\mathscr{T}_h\}, \tag{22}$$

then the inverse estimates (21) are satisfied.

We now state the main results for the approximate solution $\mathbf{u}_{h,p}^w$.

Theorem 4.1. *Suppose that the exact solution* $\mathbf{u}$ *of problem (1) and (2) belongs to* $[H^{p+1}(\Omega)]^m$. *Then there exists a constant* $C > 0$ *independent of* $\mathbf{u}$, w, *and* h *such that*

$$\|\mathbf{u} - \mathbf{u}_{h,p}^w\|_{a_w} \leqslant C\{h^p + wh^{p+1}\}\|\mathbf{u}\|_{p+1,\Omega}, \tag{23}$$

$$\|\mathscr{L}\mathbf{u}_{h,p}^w - \mathbf{f}\|_{0,\Omega} \leqslant C\{h^p + wh^{p+1}\}\|\mathbf{u}\|_{p+1,\Omega}, \tag{24}$$

$$\|\mathscr{B}\mathbf{u}_{h,p}^w - \mathbf{g}\|_{0,\partial\Omega} \leqslant C\left\{\frac{h^p}{\sqrt{w}} + \sqrt{w}h^{p+1}\right\}\|\mathbf{u}\|_{p+1,\Omega}. \tag{25}$$

If, in addition, (H1) *holds then*

$$\|\mathbf{u} - \mathbf{u}_{h,p}^w\|_{0,\Omega} \leqslant C\left\{\left(1 + \frac{1}{\sqrt{w}}\right)h^p + (w + \sqrt{w})h^{p+1}\right\}\|\mathbf{u}\|_{p+1,\Omega}. \tag{26}$$

If (H2) *and (21) hold with* $1/w = \mathrm{O}(h)$, *then*

$$\|\mathbf{u} - \mathbf{u}_{h,p}^w\|_{1,\Omega} \leqslant C\{(1 + \sqrt{w})h^p + w^{3/2}h^{p+1}\}\|\mathbf{u}\|_{p+1,\Omega}. \tag{27}$$

Proof. Let $\mathbf{v}_{h,p} \in \mathscr{V}_{h,p}$ such that (9) hold with $\mathbf{v}$ replaced by $\mathbf{u}$. Then, by (17), we get (23) immediately. By the definitions (5) and (3), we obtain

$$\|\mathbf{u} - \mathbf{u}_{h,p}^w\|_{a_w}^2 = \|\mathscr{L}(\mathbf{u} - \mathbf{u}_{h,p}^w)\|_{0,\Omega}^2 + w\|\mathscr{B}(\mathbf{u} - \mathbf{u}_{h,p}^w)\|_{0,\partial\Omega}^2.$$

Then estimates (24) and (25) follow easily from (23).

The estimate (26) is an immediate consequence of (H1), (24) and (25).

To prove (27), assume that assumptions (H2) and (21) hold with $1/w = \mathrm{O}(h)$. Then we obtain, for any $\mathbf{v}_{h,p} \in \mathscr{V}_{h,p} \subset [H^1(\Omega)]^m$,

$$\begin{aligned}
\|\mathbf{v}_{h,p}\|_{1,\Omega}^2 &\leqslant \left\{C_9(\|\mathscr{L}\mathbf{v}_{h,p}\|_{0,\Omega} + \|\mathscr{B}\mathbf{v}_{h,p}\|_{1/2,\partial\Omega})\right\}^2 \\
&\leqslant C_{11}(\|\mathscr{L}\mathbf{v}_{h,p}\|_{0,\Omega}^2 + \|\mathscr{B}\mathbf{v}_{h,p}\|_{1/2,\partial\Omega}^2) \\
&\leqslant C_{12}(\|\mathscr{L}\mathbf{v}_{h,p}\|_{0,\Omega}^2 + h^{-1}\|\mathscr{B}\mathbf{v}_{h,p}\|_{0,\partial\Omega}^2) \\
&\leqslant C_{13}a_w(\mathbf{v}_{h,p}, \mathbf{v}_{h,p}).
\end{aligned} \tag{28}$$

Applying inequality (28) to $\mathbf{u}_{h,p}^w - \mathbf{v}_{h,p} \in \mathscr{V}_{h,p}$ and using (16), we have

$$\begin{aligned}
\|\mathbf{u}_{h,p}^w - \mathbf{v}_{h,p}\|_{1,\Omega}^2 &\leqslant C_{13} a_w\left(\mathbf{u}_{h,p}^w - \mathbf{v}_{h,p}, \mathbf{u}_{h,p}^w - \mathbf{v}_{h,p}\right) \\
&= C_{13}\left\{a_w\left(\mathbf{u}_{h,p}^w - \mathbf{u}, \mathbf{u}_{h,p}^w - \mathbf{v}_{h,p}\right) + a_w\left(\mathbf{u} - \mathbf{v}_{h,p}, \mathbf{u}_{h,p}^w - \mathbf{v}_{h,p}\right)\right\} \\
&= C_{13} a_w\left(\mathbf{u} - \mathbf{v}_{h,p}, \mathbf{u}_{h,p}^w - \mathbf{v}_{h,p}\right) \\
&\leqslant C_{14}\Big\{\|\mathbf{u} - \mathbf{v}_{h,p}\|_{1,\Omega}\|\mathbf{u}_{h,p}^w - \mathbf{v}_{h,p}\|_{1,\Omega} \\
&\qquad + w\|\mathbf{u} - \mathbf{v}_{h,p}\|_{0,\partial\Omega}\|\mathbf{u}_{h,p}^w - \mathbf{v}_{h,p}\|_{0,\partial\Omega}\Big\}.
\end{aligned}$$

Applying Lemma 2.1 to $\|\mathbf{u} - \mathbf{v}_{h,p}\|_{0,\partial\Omega}$ and $\|\mathbf{u}_{h,p}^w - \mathbf{v}_{h,p}\|_{0,\partial\Omega}$ with $\varepsilon = 1/\sqrt{w}$ and $\varepsilon = 1$, respectively, we get

$$\begin{aligned}
\|\mathbf{u}_{h,p}^w - \mathbf{v}_{h,p}\|_{1,\Omega}^2 \leqslant C_{14}\Big\{&\|\mathbf{u} - \mathbf{v}_{h,p}\|_{1,\Omega}\|\mathbf{u}_{h,p}^w - \mathbf{v}_{h,p}\|_{1,\Omega} \\
&+ C_2\left(\sqrt{w}\|\mathbf{u} - \mathbf{v}_{h,p}\|_{1,\Omega} + w^{3/2}\|\mathbf{u} - \mathbf{v}_{h,p}\|_{0,\Omega}\right) \\
&\times\left(\|\mathbf{u}_{h,p}^w - \mathbf{v}_{h,p}\|_{1,\Omega} + \|\mathbf{u}_{h,p}^w - \mathbf{v}_{h,p}\|_{0,\Omega}\right)\Big\}.
\end{aligned}$$

Hence,

$$\|\mathbf{u}_{h,p}^w - \mathbf{v}_{h,p}\|_{1,\Omega} \leqslant C_{15}\left\{\|\mathbf{u} - \mathbf{v}_{h,p}\|_{1,\Omega} + \sqrt{w}\|\mathbf{u} - \mathbf{v}_{h,p}\|_{1,\Omega} + w^{3/2}\|\mathbf{u} - \mathbf{v}_{h,p}\|_{0,\Omega}\right\}.$$

Using the approximation property (6), we can choose $\mathbf{v}_{h,p} \in \mathscr{V}_{h,p}$ so that

$$\|\mathbf{u} - \mathbf{v}_{h,p}\|_{0,\Omega} + h\|\mathbf{u} - \mathbf{v}_{h,p}\|_{1,\Omega} \leqslant C_1 h^{p+1}\|\mathbf{u}\|_{p+1,\Omega}.$$

Then, by the triangle inequality,

$$\begin{aligned}
\|\mathbf{u} - \mathbf{u}_{h,p}^w\|_{1,\Omega} &\leqslant \|\mathbf{u} - \mathbf{v}_{h,p}\|_{1,\Omega} + \|\mathbf{v}_{h,p} - \mathbf{u}_{h,p}^w\|_{1,\Omega} \\
&\leqslant C_1 h^p\|\mathbf{u}\|_{p+1,\Omega} + C_{15}\Big\{C_1 h^p\|\mathbf{u}\|_{p+1,\Omega} + C_1\sqrt{w}h^p\|\mathbf{u}\|_{p+1,\Omega} \\
&\qquad + C_1 w^{3/2}h^{p+1}\|\mathbf{u}\|_{p+1,\Omega}\Big\} \\
&\leqslant C\left\{(1 + \sqrt{w})h^p + w^{3/2}h^{p+1}\right\}\|\mathbf{u}\|_{p+1,\Omega}.
\end{aligned}$$

The proof is complete. □

Corollary 4.2. *Under the same assumptions as in Theorem* 4.1, *if we take* $w = const$ *or* $w = h^{-1}$, *then the error estimates* (23)–(26) *become respectively as*

$$\|\mathbf{u} - \mathbf{u}_{h,p}^{w}\|_{a_w} \leqslant Ch^{p}\|\mathbf{u}\|_{p+1,\Omega}, \tag{23'}$$

$$\|\mathscr{L}\mathbf{u}_{h,p}^{w} - \mathbf{f}\|_{0,\Omega} \leqslant Ch^{p}\|\mathbf{u}\|_{p+1,\Omega}, \tag{24'}$$

$$\|\mathscr{B}\mathbf{u}_{h,p}^{w} - \mathbf{g}\|_{0,\partial\Omega} \leqslant Ch^{p+k}\|\mathbf{u}\|_{p+1,\Omega}, \tag{25'}$$

$$\|\mathbf{u} - \mathbf{u}_{h,p}^{w}\|_{0,\Omega} \leqslant Ch^{p}\|\mathbf{u}\|_{p+1,\Omega}, \tag{26'}$$

where $k = 0$ *if* $w =$ *constant, and* $k = \frac{1}{2}$ *if* $w = h^{-1}$. *Moreover, if* $w = h^{-1}$ *then (27) becomes as*

$$\|\mathbf{u} - \mathbf{u}_{h,p}^{w}\|_{1,\Omega} \leqslant Ch^{p-(1/2)}\|\mathbf{u}\|_{p+1,\Omega}. \tag{27'}$$

Remark 4.3. Evidently, the error estimates (26) and (27) are not optimal. It is unclear if these estimates are sharp under the above assumptions. For certain systems, it is possible to achieve optimal convergence with stronger conditions (see, e.g., Section 8.4 in Ref. [37]).

5. First-order systems

There are many important first-order problems such as the Friedrichs and ADN systems that satisfy (H1) or (H2).

5.1. Friedrichs' symmetric positive systems

In Ref. [24], Friedrichs introduced the notion of symmetric positive linear differential equations independent of type. The criterion of symmetric- positiveness has many advantageous features. For example, suitable boundary conditions can always be determined and equations of different types are treated in a unified way. In particular, the Friedrichs theory has been shown to be a very useful tool in the theoretical analysis for mixed type PDEs such as the Tricomi equation and the forward–backward heat equation that are cast into equivalent first-order systems. For the details, we refer to Refs. [3,4,24,28–31].

Consider the following system of differential equations of first-order which is a special form of problem (1) and (2),

$$\mathscr{L}\mathbf{u} := \sum_{i=1}^{d} A_i \frac{\partial \mathbf{u}}{\partial x_i} + A_0\mathbf{u} = \mathbf{f} \quad \text{in } \Omega, \tag{29}$$

$$\mathscr{B}\mathbf{u} := (\mu - \beta)\mathbf{u} = \mathbf{0} \quad \text{on } \partial\Omega, \tag{30}$$

where $\beta = \sum_{i=1}^{d} n_i A_i$, the n_i, $1 \leqslant i \leqslant d$, being the components of the unit outer normal vector $\mathbf{n}$ on $\partial\Omega$, μ is a given continuous $m \times m$ matrix defined along $\partial\Omega$. The differential operator $\mathscr{L}$ in (29) and the boundary conditions (30) are

symmetric positive and admissible, respectively, in the following sense. The operator $\mathscr{L}$ is symmetric positive if

1. the $m \times m$ matrices A_i, $1 \leqslant i \leqslant d$, are symmetric on $\overline{\Omega}$,
2. $M = A_0 + A_0^{\mathrm{t}} - \sum_{i=1}^{d} \partial A_i / \partial x_i \geqslant c_1 I$ in Ω,

where c_1 is a positive constant and I denotes the $m \times m$ identity matrix. The boundary conditions Eq. (30) is admissible if

3. $\mu + \mu^{\mathrm{t}} \geqslant \mathbf{0}_{m \times m}$ on $\partial\Omega$,
4. $\mathrm{Ker}(\mu - \beta) \oplus \mathrm{Ker}(\mu + \beta) = \mathbb{R}^m$ on $\partial\Omega$.

To verify the assumption (H1), we use the well-known (second) identity of Friedrichs [24]

$$2(\mathbf{v}, \mathscr{L}\mathbf{v})_{0,\Omega} + (\mathbf{v}, \mathscr{B}\mathbf{v})_{0,\partial\Omega} = (\mathbf{v}, M\mathbf{v})_{0,\Omega} + (\mathbf{v}, \mu\mathbf{v})_{0,\partial\Omega} \quad \forall \mathbf{v} \in [H^1(\Omega)]^m. \tag{31}$$

Assume further that $\mu + \mu^{\mathrm{t}} \geqslant c_2 I$ on $\partial\Omega$, $c_2 > 0$ constant (see also Theorem 2.1 in Ref. [28]). Then, by the Cauchy–Schwarz inequality and the basic inequality, $ab \leqslant (\varepsilon^2 a^2/2) + (b^2/2\varepsilon^2)$, for all real numbers a, b, and $\varepsilon > 0$, we get

$$\begin{aligned} c_1 \|\mathbf{v}\|_{0,\Omega}^2 + \frac{1}{2} c_2 \|\mathbf{v}\|_{0,\partial\Omega}^2 &\leqslant 2\|\mathbf{v}\|_{0,\Omega} \|\mathscr{L}\mathbf{v}\|_{0,\Omega} + \|\mathbf{v}\|_{0,\partial\Omega} \|\mathscr{B}\mathbf{v}\|_{0,\partial\Omega} \\ &\leqslant \varepsilon_1^2 \|\mathscr{L}\mathbf{v}\|_{0,\Omega}^2 + \frac{1}{\varepsilon_1^2} \|\mathbf{v}\|_{0,\Omega}^2 + \frac{1}{2}\varepsilon_2^2 \|\mathscr{B}\mathbf{v}\|_{0,\partial\Omega}^2 + \frac{1}{2\varepsilon_2^2} \|\mathbf{v}\|_{0,\partial\Omega}^2 \end{aligned}$$

for any positive constants $\varepsilon_1, \varepsilon_2$. Choosing $\varepsilon_1, \varepsilon_2$ such that $C_{18} := c_1 - 1/\varepsilon_1^2 > 0$ and $1/2\varepsilon_2^2 = c_2/2$, we thus have

$$C_{18} \|\mathbf{v}\|_{0,\Omega}^2 \leqslant C_{19} \left(\|\mathscr{L}\mathbf{v}\|_{0,\Omega}^2 + \|\mathscr{B}\mathbf{v}\|_{0,\partial\Omega}^2 \right)$$

which illustrates (H1).

Example 5.1 (*The neutron transport equation*). Let $\mathbf{d} = (1,1)^{\mathrm{t}} \in \mathbb{R}^2$. Consider the following neutron transport equation in plane that no neutrons are entering the system from outside,

$$\begin{aligned} &\nabla u \cdot \mathbf{d} + u = f \quad \text{in } \Omega := (0,1) \times (0,1), \\ &u = 0 \quad \text{on } \partial\Omega_-, \end{aligned} \tag{32}$$

where $\partial\Omega_-$ is the inflow boundary defined by

$$\begin{aligned} \partial\Omega_- &= \{\mathbf{x} \in \partial\Omega : \mathbf{n}(\mathbf{x}) \cdot \mathbf{d} < 0\} \\ &= \{(0,y)^{\mathrm{t}} : y \in (0,1)\} \cup \{(x,0)^{\mathrm{t}} : x \in (0,1)\}, \end{aligned}$$

$\mathbf{n}(\mathbf{x})$ being the outward unit normal vector to $\partial\Omega$ at the point $\mathbf{x} \in \partial\Omega$. Then problem (32) is a simple symmetric positive system with $m = 1$, $A_1 = A_2 = A_0 = 1$, $\beta = \mathbf{n}(\mathbf{x}) \cdot \mathbf{d}$ and $\mu = |\beta| = 1 > 0$. Thus, the assumption (H1) is fulfilled.

5.2. First-order ADN elliptic systems

Another interesting class of differential systems are the first-order ADN elliptic systems. In Ref. [1], the ellipticity of the general system of partial differential equations is determined by three ordered sets of integral indices $\{s_i\} = (s_1, \ldots, s_m)$, $s_i \leqslant 0$, $\{t_j\} = (t_1, \ldots, t_m)$, $t_j \geqslant 0$, and $\{r_k\} = (r_1, \ldots, r_n)$ corresponding respectively to differential equations, unknown functions, and boundary conditions. Based on the ADN theory, the operators $\mathscr{L}$, $\mathscr{B}$ appearing in (1) and (2) must satisfy the so-called uniform ellipticity condition, supplementary condition and the complementing boundary condition in order to have the following coercive type estimates.

For each $l \geqslant 0$, there exists a constant $C_{20} > 0$ such that if $\mathbf{v} = (v_1, \ldots, v_m)^t$, $v_j \in H^{l+t_j}(\Omega)$, $j = 1, \ldots, m$, then

$$\sum_{j=1}^{m} \|v_j\|_{l+t_j,\Omega} \leqslant C_{20} \left(\sum_{i=1}^{m} \|(\mathscr{L}\mathbf{v})_i\|_{l-s_i,\Omega} + \sum_{k=1}^{n} \|(\mathscr{B}\mathbf{v})_k\|_{l-r_k-1/2,\partial\Omega} \right), \tag{33}$$

where $\mathscr{L}\mathbf{v} = ((\mathscr{L}\mathbf{v})_1, \ldots, (\mathscr{L}\mathbf{v})_m)^t$, $\mathscr{B}\mathbf{v} = ((\mathscr{B}\mathbf{v})_1, \ldots, (\mathscr{B}\mathbf{v})_n)^t$.

We shall not state these conditions here. To fulfill these three conditions in turn leads to a rather complicated algebraic checking on the three ordered sets $\{s_i\}$, $\{t_j\}$, and $\{r_k\}$.

Recently, Bochev and Gunzburger [5], Chang et al. [18,19], and Jiang and Chang [26] have successfully formulated the Stokes equations into first-order systems in two- or three-dimensional bounded regions and then proved that, under appropriate formulation (see the next two examples), (33) is satisfied with

$$\begin{aligned} &\{s_i\} = (0, \ldots, 0), \\ &\{t_j\} = (1, \ldots, 1), \\ &\{r_k\} = (-1, \ldots, -1). \end{aligned} \tag{34}$$

Consequently, we have (H2) by taking $l = 0$.

Example 5.2 (*The Stokes equations in the velocity–vorticity–pressure formulation*). Let $\Omega \subset \mathbb{R}^2$ be a bounded domain with smooth boundary $\partial\Omega$. The Stokes equations for incompressible flow can be expressed as

$$\begin{aligned} -\Delta\mathbf{u} + \operatorname{grad} p &= \mathbf{f} \quad \text{in } \Omega, \\ \operatorname{div} \mathbf{u} &= 0 \quad \text{in } \Omega, \end{aligned} \tag{35}$$

where $\mathbf{u} = (u_1, u_2)^t$ denotes the velocity, p the pressure, and $\mathbf{f} = (f_1, f_2)^t$ the body force. By introducing the vorticity $\omega := \operatorname{curl}\mathbf{u} = \partial u_2/\partial x - \partial u_1/\partial y$ as an auxiliary variable and utilizing another two-dimensional curl operator $\operatorname{curl}\omega = (\omega_y, -\omega_x)^t$, (35) can be transformed into the following first-order system in velocity–vorticity–pressure form

$$\begin{aligned} \operatorname{curl} \omega + \operatorname{grad} p &= \mathbf{f} \quad \text{in } \Omega, \\ -\omega + \operatorname{curl} \mathbf{u} &= 0 \quad \text{in } \Omega, \\ \operatorname{div} \mathbf{u} &= 0 \quad \text{in } \Omega. \end{aligned} \tag{36}$$

If system (36) is supplemented with the following boundary conditions,

$$\begin{aligned} p &= 0 \quad \text{on } \partial\Omega, \\ u_1 n_1 + u_2 n_2 &= 0 \quad \text{on } \partial\Omega, \end{aligned} \tag{37}$$

then Eqs. (36) and (37) is an ADN elliptic system and (33) holds with (34) (cf. Ref. [5]). Unfortunately, if system (36) is imposed by the homogeneous velocity boundary conditions,

$$\begin{aligned} u_1 &= 0 \quad \text{on } \partial\Omega, \\ u_2 &= 0 \quad \text{on } \partial\Omega \end{aligned} \tag{38}$$

with $(p, 1)_{0,\Omega} = 0$, (34) will not hold, i.e., the assumption (H2) fails to hold for this problem. However, the following formulation works well for this type of boundary conditions which are more useful.

Example 5.3 (*The Stokes equations in the velocity–stress–pressure formulation*). In Ref. [19], the velocity–stress–pressure formulation for the two-dimensional Stokes equations is proposed as follows:

$$\begin{aligned} -\frac{\partial \varphi_1}{\partial x} - \frac{\partial \varphi_2}{\partial y} + \frac{\partial p}{\partial x} &= f_1 \quad \text{in } \Omega, \\ \frac{\partial \varphi_1}{\partial y} - \frac{\partial \varphi_3}{\partial x} + \frac{\partial p}{\partial y} &= f_2 \quad \text{in } \Omega, \\ \frac{\partial \varphi_1}{\partial x} + \frac{\partial \varphi_3}{\partial y} &= 0 \quad \text{in } \Omega, \\ \frac{\partial \varphi_1}{\partial y} - \frac{\partial \varphi_2}{\partial x} &= 0 \quad \text{in } \Omega, \\ \operatorname{div} \mathbf{u} &= 0 \quad \text{in } \Omega, \\ \operatorname{curl} \mathbf{u} - \varphi_3 + \varphi_2 &= 0 \quad \text{in } \Omega \end{aligned} \tag{39}$$

with $(p, 1)_{0,\Omega} = 0$, where the auxiliary variables φ_1, φ_2, and φ_3 are introduced as

$$\begin{aligned} \varphi_1 &= \frac{\partial u_1}{\partial x} \quad \text{in } \Omega, \\ \varphi_2 &= \frac{\partial u_1}{\partial y} \quad \text{in } \Omega, \\ \varphi_3 &= \frac{\partial u_2}{\partial x} \quad \text{in } \Omega, \end{aligned} \tag{40}$$

and their combinations represent the usual stresses. If system (39) is supplemented with the boundary conditions

$$\begin{aligned} n_1\varphi_2 - n_2\varphi_1 &= 0 \quad \text{on } \partial\Omega, \\ n_1\varphi_1 + n_2\varphi_3 &= 0 \quad \text{on } \partial\Omega, \\ n_1 u_1 + n_2 u_2 &= 0 \quad \text{on } \partial\Omega, \end{aligned} \tag{41}$$

which are equivalent to (38), then it is an ADN elliptic system and (33) holds with (34).

Many other boundary value problems can also be proved to have the estimates (33) with (34) by using the ADN theory. For the details, we refer to Refs. [1,2,13,14,16].

6. Condition number

In this section, we analyze the asymptotic conditioning of the linear system arising from problem (14). Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_K\}$ be a set of basis functions for the finite element space $\mathscr{V}_{h,p}$ and we assume the basis is chosen so that the following two conditions hold [2,6,25].

There exist positive constants Λ_1 and Λ_2 such that for all $\xi_1, \ldots, \xi_K \in \mathbb{R}$,

$$\Lambda_1 h^d \sum_{i=1}^{K} \xi_i^2 \leqslant \left(\sum_{i=1}^{K} \xi_i \mathbf{u}_i, \sum_{j=1}^{K} \xi_j \mathbf{u}_j \right)_{0,\Omega} \leqslant \Lambda_2 h^d \sum_{i=1}^{K} \xi_i^2, \tag{42}$$

$$\left(\sum_{i=1}^{K} \xi_i \mathbf{u}_i, \sum_{j=1}^{K} \xi_j \mathbf{u}_j \right)_{1,\Omega} \leqslant \Lambda_2 h^{d-2} \sum_{i=1}^{K} \xi_i^2. \tag{43}$$

Note that the above inequalities hold for most finite element spaces $\mathscr{V}_{h,p}$ if condition (22) is satisfied.

Theorem 6.1. *Suppose that the basis conditions (42) and (43) are satisfied. If (H1) holds with $w \geqslant 1$, or (H2) and (21) hold with $1/w = \mathrm{O}(h)$, then the condition number of the resulting linear system of (14) is $\mathrm{O}(h^{-2} + w^2)$.*

Proof. Since the matrix

$$\mathbf{M} := (\mathbf{M}_{i,j})_{K\times K} = \left(a_w(\mathbf{u}_i, \mathbf{u}_j)\right)_{K\times K}$$

is symmetric and positive definite, we find that

$$\text{condition number of } \mathbf{M} = \frac{\lambda_{\max}}{\lambda_{\min}} = \frac{\max \rho(\Xi)}{\min \rho(\Xi)}, \tag{44}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the largest and smallest eigenvalues of $\mathbf{M}$, $\rho(\Xi)$ is the Rayleigh quotient,

$$\rho(\Xi) := \frac{\Xi^{\mathrm{t}}\mathbf{M}\Xi}{\Xi^{\mathrm{t}}\Xi} = \frac{a_w\left(\sum_{i=1}^K \xi_i\mathbf{u}_i, \sum_{j=1}^K \xi_j\mathbf{u}_j\right)}{\Xi^{\mathrm{t}}\Xi} \tag{45}$$

for any $\Xi = (\xi_1, \ldots, \xi_K)^{\mathrm{t}} \in \mathbb{R}^K$, $\Xi \neq \mathbf{0}$.

Let $\mathbf{v}_{h,p} = \sum_{i=1}^K \xi_i\mathbf{u}_i$. Suppose (H1) holds with $w \geqslant 1$, then, by (42),

$$\begin{aligned}
\Lambda_1 h^d \sum_{i=1}^K \xi_i^2 &\leqslant \left(\sum_{i=1}^K \xi_i\mathbf{u}_i, \sum_{j=1}^K \xi_j\mathbf{u}_j\right)_{0,\Omega} = \|\mathbf{v}_{h,p}\|_{0,\Omega}^2 \\
&\leqslant C_8^2\left(\|\mathscr{L}\mathbf{v}_{h,p}\|_{0,\Omega} + \|\mathscr{B}\mathbf{v}_{h,p}\|_{0,\partial\Omega}\right)^2 \\
&\leqslant C_{21}\left(\|\mathscr{L}\mathbf{v}_{h,p}\|_{0,\Omega}^2 + \|\mathscr{B}\mathbf{v}_{h,p}\|_{0,\partial\Omega}^2\right) \\
&\leqslant C_{21}\left(\|\mathscr{L}\mathbf{v}_{h,p}\|_{0,\Omega}^2 + w\|\mathscr{B}\mathbf{v}_{h,p}\|_{0,\partial\Omega}^2\right) \\
&= C_{21}a_w(\mathbf{v}_{h,p}, \mathbf{v}_{h,p}).
\end{aligned} \tag{46}$$

If (H2) and (21) hold with $1/w = \mathrm{O}(h)$, by (42) and (28), we have

$$\begin{aligned}
\Lambda_1 h^d \sum_{i=1}^K \xi_i^2 &\leqslant \left(\sum_{i=1}^K \xi_i\mathbf{u}_i, \sum_{j=1}^K \xi_j\mathbf{u}_j\right)_{0,\Omega} = \|\mathbf{v}_{h,p}\|_{0,\Omega}^2 \leqslant \|\mathbf{v}_{h,p}\|_{1,\Omega}^2 \\
&\leqslant C_{13}a_w(\mathbf{v}_{h,p}, \mathbf{v}_{h,p}).
\end{aligned} \tag{47}$$

On the other hand, by (3), we obtain (cf. (10))

$$\begin{aligned}
a_w(\mathbf{v}_{h,p}, \mathbf{v}_{h,p}) &= (\mathscr{L}\mathbf{v}_{h,p}, \mathscr{L}\mathbf{v}_{h,p})_{0,\Omega} + w(\mathscr{B}\mathbf{v}_{h,p}, \mathscr{B}\mathbf{v}_{h,p})_{0,\partial\Omega} \\
&\leqslant C_{22}\left(\|\mathbf{v}_{h,p}\|_{1,\Omega}^2 + w\|\mathbf{v}_{h,p}\|_{0,\partial\Omega}^2\right).
\end{aligned} \tag{48}$$

By Lemma 2.1, we get

$$\|\mathbf{v}_{h,p}\|_{0,\partial\Omega}^2 \leqslant 2C_2^2\left(\varepsilon^2\|\mathbf{v}_{h,p}\|_{1,\Omega}^2 + \frac{1}{\varepsilon^2}\|\mathbf{v}_{h,p}\|_{0,\Omega}^2\right). \tag{49}$$

Taking $\varepsilon^2 = 1/w$ in (49), then (48) becomes

$$\begin{aligned}
a_w(\mathbf{v}_{h,p}, \mathbf{v}_{h,p}) &\leqslant C_{23}\left(\|\mathbf{v}_{h,p}\|_{1,\Omega}^2 + w\left(\frac{1}{w}\|\mathbf{v}_{h,p}\|_{1,\Omega}^2 + w\|\mathbf{v}_{h,p}\|_{0,\Omega}^2\right)\right) \\
&= C_{24}\left(\|\mathbf{v}_{h,p}\|_{1,\Omega}^2 + w^2\|\mathbf{v}_{h,p}\|_{0,\Omega}^2\right) \\
&\leqslant C_{24}\Lambda_2\left(h^{d-2} + w^2h^d\right)\sum_{i=1}^K \xi_i^2.
\end{aligned} \tag{50}$$

The proof is completed by (44)–(47), and (50). □

7. Concluding remarks

A unified analysis of a weighted least squares finite element method applied to a general class of first-order differential systems is presented. The method is based on the minimization of a least squares functional that is a sum of the residuals in the differential equations and the residuals with the same weight in the boundary conditions. Compared with other LSFEMs, the most significant feature of the method is that the trial and test functions need not satisfy the boundary conditions. Consequently, it applies to a broad scope of problems with only L^2 regularity required on the boundary data.

Asymptotic convergence is established in a natural norm without any extra regularity conditions on the exact solution. Many mathematical model problems fit into this general framework. In particular, we present two types of assumptions which are respectively suitable for Friedrichs' symmetric positive systems and for first-order Agmon–Douglis–Nirenberg elliptic systems. Under these assumptions, more specific convergence properties can be analyzed. The resulting linear system is symmetric positive definite with condition number $O(h^{-2} + w^2)$. Three examples, namely, the neutron transport equation and two first-order formulations for the Stokes equations with various boundary conditions are examined.

It is evident that the least squares approximation involves more degrees of freedom in the solution procedure since there are more unknowns to be determined at each nodal point and more equations to be approximated under the reduced first-order system. Nevertheless, with its advantageous properties such as symmetric positive definiteness and uniform finite element spectral order, this drawback may be alleviated via effective and efficient adaptive process [32,34] and/or parallel implementation.

Acknowledgements

We would like to thank Professor C.L. Chang (Department of Mathematics, Cleveland State University, Cleveland, Ohio) for his helpful comments and suggestions on the paper when he was visiting our department in September 1996. The second author would like to express his gratitude to the Department of Mathematics, Texas A & M University for a stimulating and enjoyable visit during which part of this work was undertaken.

References

[1] S. Agmon, A. Douglis, L. Nirenberg, Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions II, Commun. Pure Appl. Math. 17 (1964) 35–92.

[2] A.K. Aziz, R.B. Kellogg, A.B. Stephens, Least squares methods for elliptic systems, Math. Comp. 44 (1985) 53–70.
[3] A.K. Aziz, S.H. Leventhal, Numerical solution of linear partial differential equations of elliptic-hyperbolic type, in: B.E. Hubbard (Ed.), Numerical Solution of Partial Differential Equations, III, Academic Press, New York, 1976, pp. 55–88.
[4] A.K. Aziz, J.-L. Liu, A weighted least squares method for the backward-forward heat equation, SIAM J. Numer. Anal. 28 (1991) 156–167.
[5] P.B. Bochev, M.D. Gunzburger, Analysis of least squares finite element methods for the Stokes equations, Math. Comp. 63 (1994) 479–506.
[6] J.H. Bramble, J.A. Nitsche, A generalized Ritz-least-squares method for Dirichlet problems, SIAM J. Numer. Anal. 10 (1973) 81–93.
[7] J.H. Bramble, A.H. Schatz, Rayleigh-Ritz-Galerkin methods for Dirichlet's problem using subspaces without boundary conditions, Commun. Pure Appl. Math. 23 (1970) 653–675.
[8] J.H. Bramble, A.H. Schatz, Least squares methods for $2m$th order elliptic boundary-value problems, Math. Comp. 25 (1971) 1–32.
[9] J.H. Bramble, V. Thomée, Semidiscrete-least squares methods for a parabolic boundary value problem, Math. Comp. 26 (1972) 633–648.
[10] F. Brezzi, M. Fortin, Mixed and Hybrid Finite Element Methods, Springer, New York, 1991.
[11] Z. Cai, R. Lazarov, T.A. Manteuffel, S.F. McCormick, First-order system least squares for second-order partial differential equations: Part I, SIAM J. Numer. Anal. 31 (1994) 1785–1799.
[12] G.F. Carey, J.T. Oden, Finite Elements: A Second Course, Prentice-Hall, Englewood Cliffs, NJ, 1983.
[13] C.L. Chang, A least squares finite element method for the Helmholtz equation, Comput. Methods Appl. Mech. Eng. 83 (1990) 1–7.
[14] C.L. Chang, Finite element approximation for grad-div type systems in the plane, SIAM J. Numer. Anal. 29 (1992) 452–461.
[15] C.L. Chang, An error estimate of the least squares finite element method for the Stokes problem in three dimensions, Math. Comp. 63 (1994) 41–50.
[16] C.L. Chang, Least-squares finite elements for second-order boundary value problems with optimal rates of convergence, Appl. Math. Comput. 76 (1996) 267–284.
[17] C.L. Chang, M.D. Gunzburger, A finite element method for first order elliptic systems in three dimensions, Appl. Math. Comput. 23 (1987) 171–184.
[18] C.L. Chang, B.-N. Jiang, An error analysis of least squares finite element method of velocity–pressure–vorticity formulation for Stokes problem, Comput. Methods Appl. Mech. Eng. 84 (1990) 247–255.
[19] C.L. Chang, S.-Y. Yang, C.-H. Hsu, A least-squares finite element method for incompressible flow in stress–velocity–pressure version, Comput. Methods Appl. Mech. Eng. 128 (1995) 1–9.
[20] T.-F. Chen, On least squares approximations to compressible flow problems, Numer. Methods Partial Differential Equations 12 (1986) 207–228.
[21] P.G. Ciarlet, The Finite Element Method for Elliptic Problems, North-Holland, Amsterdam, 1978.
[22] G.J. Fix, M.D. Gunzburger, R.A. Nicolaides, On finite element methods of the least squares type, Comp. Math. Appl. 5 (1979) 87–98.
[23] G.J. Fix, M.E. Rose, A comparative study of finite element and finite difference methods for Cauchy-Riemann type equations, SIAM J. Numer. Anal. 22 (1985) 250–261.
[24] K.O. Friedrichs, Symmetric positive differential equations, Comm. Pure Appl. Math. 11 (1958) 333–418.
[25] D.C. Jespersen, A least squares decomposition method for solving elliptic equations, Math. Comp. 31 (1977) 873–880.
[26] B.-N. Jiang, C.L. Chang, Least-squares finite elements for the Stokes problem, Comput. Methods Appl. Mech. Eng. 78 (1990) 297–311.

[27] C. Johnson, Numerical Solution of Partial Differential Equations by the Finite Element Method, Cambridge University Press, Cambridge, 1990.
[28] T. Katsanis, Numerical solution of symmetric positive differential equations, Math. Comp. 22 (1968) 763–783.
[29] T. Katsanis, Numerical solution of Tricomi equation using theory of symmetric positive differential equations, SIAM J. Numer. Anal. 6 (1969) 236–253.
[30] P. Lesaint, Finite element methods for symmetric hyperbolic equations, Numer. Math. 21 (1973) 244–255.
[31] P. Lesaint, Continuous and discontinuous finite element methods for solving the transport equation, in: J.R. Whiteman (Ed.), The Mathematics of Finite Elements and Applications II, MAFELAP 1975, Academic Press, New York, 1976, 151–161.
[32] I.-J. Lin, D.-P. Chen, J.-L. Liu, Adaptive least squares finite element methods for the Stokes problem (submitted).
[33] J.L. Lions, E. Magenes, Nonhomogeneous Elliptic Boundary Value Problems and Applications, vol. I, Springer, Berlin, 1972.
[34] J.-L. Liu, I.-J. Lin, M.-Z. Shih, R.-C. Chen, M.-C. Hsieh, Object oriented programming of adaptive finite element and finite volume methods, to appear in Applied Numerical Mathematics.
[35] A.I. Pehlivanov, G.F. Carey, R.D. Lazarov, Least-squares mixed finite elements for second-order elliptic problems, SIAM J. Numer. Anal. 31 (1994) 1368–1377.
[36] P. Sermer, R. Mathon, Least-squares methods for mixed type equations, SIAM J. Numer. Anal. 18 (1981) 705–723.
[37] W.L. Wendland, Elliptic Systems in the Plane, Pitman, London, 1979.