SIAM J. NUMER. ANAL.
Vol. 58, No. 6, pp. 3286–3308
© 2020 Society for Industrial and Applied Mathematics

Downloaded 01/02/21 to 59.64.129.2. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

ADAPTIVE FIRST-ORDER SYSTEM LEAST-SQUARES FINITE ELEMENT METHODS FOR SECOND-ORDER ELLIPTIC EQUATIONS IN NONDIVERGENCE FORM*

WEIFENG QIU† AND SHUN ZHANG†

Abstract. This paper studies adaptive first-order system least-squares finite element methods (LSFEMs) for second-order elliptic partial differential equations in nondivergence form. Unlike the classical finite element methods which use weak formulations of PDEs that are not applicable for the nondivergence equation, the first-order least-squares formulations naturally have stable weak forms without using integration by parts, allow simple finite element approximation spaces, and have built-in a posteriori error estimators for adaptive mesh refinements. The nondivergence equation is first written as a system of first-order equations by introducing the gradient as a new variable. Then, two versions of least-squares finite element methods using simple C^0 finite elements are developed in the paper. One is the L^2-LSFEM which uses linear elements, and the other is the W-LSFEM with a mesh-dependent weight to ensure the optimal convergence. Under the assumption that the PDE has a unique solution, a priori and a posteriori error estimates are presented. With an extra assumption on the operator regularity, convergence in standard norms for the W-LSFEM is also discussed. L^2-error estimates are derived for both formulations. Extensive numerical experiments for continuous, discontinuous, and even degenerate coefficients on smooth and singular solutions are performed to test the accuracy and efficiency of the proposed methods.

Key words. nondivergence elliptic equation, first-order system least-squares, finite element method, a priori error estimate, a posteriori error estimate, adaptive method

AMS subject classifications. 65N12, 65N15, 65N30, 65N50

DOI. 10.1137/19M1271099

1. Introduction.

In this paper we consider finite element approximations of the following elliptic PDE in nondivergence form:

$$
\begin{aligned} -A : D^2 u &= f \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{1.1}
$$

Here, the domain $\Omega \in \mathbb{R}^d$ is an open and bounded polytope for $d = 2$ or 3, and the coefficient matrix $A = A(x) \in L^\infty(\Omega)^{d\times d}$ is a symmetric positive definite matrix with eigenvalues bounded by $\lambda > 0$ below and $\Lambda > 0$ above on Ω, but not necessarily differentiable. The right-hand side f is assumed in $L^2(\Omega)$.

Note that when $A \in [C^1(\Omega)]^{n\times n}$, we have the following equation in divergence form, where $\nabla \cdot A$ is taken rowwise:

$$
\begin{aligned} -\nabla\cdot(A\nabla u) + (\nabla\cdot A)\cdot\nabla u &= f \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{1.2}
$$

*Received by the editors June 27, 2019; accepted for publication (in revised form) September 29, 2020; published electronically November 12, 2020.
https://doi.org/10.1137/19M1271099
Funding: The work of the first author was supported by the Research Grants Council of the Hong Kong Special Administrative Region, China, project CityU 11302219. The work of the second author was supported by the Research Grants Council of the Hong Kong Special Administrative Region, China, project CityU 11300517.
†Department of Mathematics, City University of Hong Kong, Kowloon Tong, Hong Kong SAR, China (weifeqiu@cityu.edu.hk, shun.zhang@cityu.edu.hk).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

The elliptic PDE in nondivergence form arises in the linearization of fully nonlinear PDEs, for example, stochastic control problems, nonlinear elasticity, and mathematical finance. The matrix $A(x)$ is not smooth nor even continuous in many such cases. For example, for fully nonlinear PDEs solved by C^0 finite element methods, the coefficient matrix of its linearization is possibly only elementwise smooth if the coefficients contain derivatives of the numerical solution.

Since the matrix A is not differentiable, the standard weak notion of elliptic equation is not applicable. The existence and uniqueness of equations of nondivergence form are often based on the classical or strong senses of the solutions; see discussions in [40, 39]. These PDE theories often assume that the domain Ω is convex, the boundary is sufficiently smooth, or some other restrictive conditions on the smoothness of A. For a discontinuous A, there is the possibility that the solution is nonunique; see an example given in Remark 2.3. It is worthwhile to mention that these available theoretical PDE results are all sufficient theories. For example, the Poisson equation ($A = I$ in (1.1)) on a nonconvex polygonal domain (e.g., an L-shaped domain) does not satisfy any sufficient conditions listed in Assumption 2.1.

Several numerical methods are available for the problem in nondivergence form. Based on discrete Calderon–Zygmond estimates, Feng, Neilan, and coauthors developed finite element methods for problems with a continuous coefficient matrix in [26, 27, 38]. For equations with discontinuous coefficients satisfying the Cordes condition, a discontinuous Galerkin method [41], a mixed method [28], and a nonsymmetric method [39] are developed. A primal-dual method is developed in [43]. The analyses of these papers mostly assumes the full H^2 regularity of the operator and studies the H^2-error estimates of the approximations. In some sense, these methods keep the nondivergence operator second order and borrow techniques from variational fourth-order problems. Nochetto and Zhang [40] studied a two-scale method, which is based on the integro-differential approach and focuses on L^∞-error estimates.

Traditionally, the finite element method is based on the variational formulation of an elliptic equation, where the integration by parts plays an essential role. The integration by parts can shift a derivative from the trial variable to the test variable, thus reducing the differential order of the operator and simplifying the construction of finite elements. For (1.1), the integration by parts is not available. Luckily, there is another natural method to reduce the differential order of a PDE operator by introducing another auxiliary variable. We can reduce the second-order equation into a system of a first-order equation by using the new auxiliary variable. Normally, for the first-order system, we have two approaches. One is the mixed method which also involves integration by parts and has difficulties to ensure stability. The other method is the least-squares finite element method (LSFEM). The first-order system least-squares principle first rewrites the PDE into a first-order system, then defines an artificial, externally defined energy-type principle. The energy functional can be defined as the summation of weighted residuals of the system. With the first-order least-squares functional, a corresponding LSFEM can be defined. No integration by parts is needed to define the least-squares principle, and thus the first-order system LSFEM (otherwise knowns as FOSLS-FEM) is ideal for the second-order elliptic equation of nondivergence form.

Beside the obvious advantage of nonrequirement of integration by parts, the LSFEM has other advantages. First, the least-squares weak formulation and its associated LSFEM using conforming finite element spaces are automatically coercive in the least-squares functional norm as long as the first-order system is well-posed. This is a significant advantage over other numerical methods since the well-posedness

Downloaded 01/02/21 to 59.64.129.2. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

theories of the equation in nondivergence form are in general only sufficient theories. On the other hand, even without a rigorous mathematical proof, the elliptic equation in non-divergence form is often a result of some physical process for which a unique solution exists. Thus, in LSFEMs developed in this paper, we can reduce the condition of the PDE into a simple well-posedness without specifying the condition explicitly.

The other advantages of LSFEMs include that conforming discretizations lead to stable and, ultimately, optimally accurate methods in least-squares functional norm, and the resulting algebraic problems are symmetric, positive definite, and can possibly be solved by standard and robust iterative methods including multigrid methods.

The last important advantage of the LSFEM is that it has a built-in a posteriori error estimator. The solution for a PDE can have singularity due to the geometry of the domain or the coefficient matrix. Also, for problems like reaction diffusion equations, interior or boundary layers appear. The a posteriori error estimator and adaptive mesh refinement algorithm are very effective in increasing the accuracy per computational cost for these complicated problems.

In this paper, by introducing the gradient as an auxiliary variable, we first write the equation in nondivergence form into a system of first-order equations, then develop two least-squares minimization principles and two corresponding LSFEMs: one is based on an L^2-norm square sum of the residuals and the other is based on a mesh-size weighted L^2-norm square sum of the residuals. The two methods are called L^2-LSFEM and W-LSFEM, respectively. For the L^2-LSFEM, simplest linear C^0-finite elements are used to approximate both the solution and the gradient. For the W-LSFEM, the C^0-finite element of degree k, $k \geq 2$ is used to approximate the solution, while the degree $k-1$ C^0-finite element is used to approximate the gradient. Under a very weak assumption that the coefficient and domain are good enough to guarantee the existence and uniqueness of a solution, we show that both continuous least-squares weak forms and their corresponding discrete problems are well-posed. A priori and a posteriori error estimates with respect to the least-squares norms are then discussed.

Numerical methods for the nondivergence equation often use the operator regularity assumption

$$\|u\|_2 \leq C\|A : D^2 u\|_0 = C\|f\|_0$$

to construct the method and derive stability and error estimates. Unlike these papers, for the W-LSFEM, the estimates of error in the H^1-norm and the discrete broken H^2-norm are investigated with a weaker assumption; see our discussion in section 4.2. Under stronger regularity assumptions, we show that the L^2-norm of the error of the solution is one order higher than the least-squares norm of the error, provided the approximation degree for the solution is at least three. For the L^2-LSFEM, we show the L^2- and H^1-error estimates with a regularity assumption. Extensive numerical experiments for smooth, nonsmooth, and even degenerate coefficients on smooth and singular solutions are performed to test the accuracy and efficiency of the proposed methods. With uniform refinements, the convergence orders matching with the theory are shown. With adaptive mesh refinements, for singular solutions, optimal convergence, in the sense that the same accuracy can be obtained with the same degrees of freedom as if the solution is smooth, is observed.

The LSFEM is well developed for the elliptic equations in divergence form; see, for example, [14, 15, 9, 7, 13, 21]. A posteriori error estimates and adaptivity algorithms based on LSFEMs can be founded in [3, 1, 42, 20]. Compared to LSFEMs for the elliptic equation in divergence form, the nondivergence equation has many differences

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/02/21 to 59.64.129.2. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

in the stability analysis and choices of the finite element subspaces due to the nondivergence structure. We remark on these differences in various places in the paper as comparisons.

There are two least-squares finite element methods available for the nondivergence equation. Neither of them uses a first-order reformulation. The paper [28] uses a second-order least-squares formulation with C^1-finite element approximations. The simple method developed in [37] uses C^0 -finite element spaces with orders higher than two and penalizes the continuity of the solution and the gradient.

In summary, the LSFEMs developed in this paper have several advantages compared to existing numerical methods: they are automatically stable under very mild assumption; they are easy to program because only simple Lagrange finite elements without jump terms are used; adaptive algorithms with the built-in a posteriori error estimators can handle problems with singular solutions or layers; under a condition on the operator regularity which is weaker than traditionally assumed, error estimates in standard norms are proved.

The remaining parts of this article are as follows. Section 2 defines the first-order system least-squares weak problems and discusses their stability; section 3 presents the corresponding LSFEMs and their a priori and a posteriori error estimates in least-squares norms. Error estimates in other norms are discussed in sections 4 and 5 for the weighted and L^2 versions of the methods, separately. Numerical experiments are presented in section 6.

Standard notation on function spaces apply throughout this article. Norms of functions in Lebesgue and Sobolev space $H^k(\omega)$ $(L^2(\omega) = H^0(\omega))$ are denoted by $\|\cdot\|_{k,\omega}$. The subscript ω is omitted when $\omega = \Omega$. The inner product of real-valued $d \times d$ matrices $A : B$ is denoted by $A : B = \sum_{i,j}^d a_{ij} b_{ij}$. We use $D^2 v$ to denote the Hessian of v.

2. First-order system least-squares problems.

2.1. Existence and uniqueness assumption.

Define

$$V := \{v \in H_0^1(\Omega) \text{ and } A : D^2 v \in L^2(\Omega)\}. \tag{2.1}$$

Notice that the space V is weaker than $H_0^1(\Omega) \cap H^2(\Omega)$, since we can expect that even if an individual $\partial_{ij}^2 v$, $i, j = 1, \ldots, d$, is not in $L^2(\Omega)$, due to the cancellation or good properties of A, $A : D^2 v$ belongs to $L^2(\Omega)$.

For example, letting $A = I$, we can construct a function $w \in H_0^1(\Omega)$ being the solution of the Poisson equation $-\Delta w = f$ with a smooth f on an L-shaped domain $\Omega = (-1, 1)^2 \backslash [0, 1) \times (-1, 0]$, such that $w \in V$ but $w \notin H^2(\Omega)$. In fact $w \in H^{5/3-\epsilon}(\Omega)$ for every $\epsilon > 0$ only; see discussions in [10, section 2.3] and [24].

We first state the assumption of the existence and uniqueness of the solution.

Assumption 2.1 (existence and uniqueness of the solution of the elliptic equation in nondivergence form). Assume that the coefficient matrix A and the domain Ω are nice enough, such that (1.1) has a unique solution $u \in V$ for any $f \in L^2(\Omega)$.

For the special case $f = 0$, it is assumed the coefficient matrix A and the domain Ω are nice enough, such that $u = 0$ is the unique solution.

There are various theories to ensure the existence and uniqueness of the equation, such as the following:

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/02/21 to 59.64.129.2. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Downloaded 01/02/21 to 59.64.129.2. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

1. (Classical solution [32]) A unique classical solution $u \in C^{2,\alpha}(\overline{\Omega})$ for any $f \in C^{0,\alpha}(\overline{\Omega})$ exists if A belongs to $C^{2,\alpha}(\overline{\Omega})$ and Ω is of class $C^{2,\alpha}$ for some $0 < \alpha < 1$.
2. (Strong solution [23]) If $A \in VMO(\Omega) \cap L^\infty(\Omega)$, a vanishing mean oscillation matrix with a uniform VMO-modulus of continuity, and if Ω is of class $C^{1,1}$, then there exists a unique solution $u \in W^{2,p}$ for $1 < p < \infty$ of the problem.
3. (H^2-solution [36, 41]) Let Ω be a bounded convex domain, and let $A \in L^\infty(\Omega)^{d\times d}$ satisfy the Cordes condition (2.3); then for any given $f \in L^2(\Omega)$, there exists a unique solution $u \in H^2(\Omega)$ with

$$\|u\|_{2,\Omega} \le C\|f\|_{0,\Omega}. \tag{2.2}$$

Remark 2.2. The third one, the H^2-solution case, is also the version of the assumption widely used in the constructions and analysis of numerical methods for nondivergence PDEs; see, for example, [41, 26, 27, 43].

The matrix $A \in L^\infty(\Omega)^{d\times d}$ is said to satisfy the Cordes condition if there exists some $\epsilon \in (0, 1]$ such that

$$|A|^2/\mathrm{tr}(A)^2 \le 1/(d - 1 + \epsilon), \tag{2.3}$$

where $|A| = \sqrt{A : A}$ is the Frobenius norm and tr denotes the trace. For the case $d = 2$, uniform ellipticity of A implies the Cordes condition, while for $d = 3$, uniformly ellipticity is not enough; see [41]. In [41], it is shown that for the bounded convex domain Ω, (2.2) is true, with the constant C depending on the matrix A and the constants in the Cordes condition.

Remark 2.3. The following nonuniqueness example was originally due to Gilbarg and Serrin [30] and also can be found in [40]. Let $\Omega \subset \mathbb{R}^3$ be a unit ball centered at 0 and $f = 0$. A discontinuous coefficient A is given by

$$A(x) = I + \frac{1+\alpha}{1-\alpha}\frac{xx^T}{|x|^2}.$$

The equation has two solutions, $u = 0$ and $u = |x|^\alpha - 1$; both of them are of H^2 provided $1/2 < \alpha < 1$.

More detailed discussions on existence and uniqueness theories can be found in the introduction of [40]. We do find these theories are only sufficient theories. Besides the example of the Poisson equation on a nonconvex domain, the Test 3 problem from [26] (see also our numerical test in section 6.2) is an example that the matrix A is not uniformly elliptic but a unique solution still exists.

2.2. Least-squares problems. As discussed at the beginning of section 2.1, the solution space of (1.1) is V, where a second-order derivative is needed. Since no integration by parts can be used to reduce the order of the operator, it is then quite complicated to construct a finite element subspace of V. To overcome this, a new variable is introduced and least-squares formulations are used to reduce the order of differential operators.

Introducing a gradient variable $\boldsymbol{\sigma} = \nabla u$, we have the following first-order system:

$$\left\{ \begin{array}{rcll} \boldsymbol{\sigma} - \nabla u &=& 0 & \text{in } \Omega, \\ -A : \nabla\boldsymbol{\sigma} &=& f & \text{in } \Omega, \end{array} \right. \tag{2.4}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/02/21 to 59.64.129.2. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

with a boundary condition $u = 0$ on $\partial\Omega$. It is clear we must choose $u \in H_0^1(\Omega)$. For the gradient $\boldsymbol{\sigma}$, it is true that $\boldsymbol{\sigma} = \nabla u \in L^2(\Omega)$ and $A : \nabla\boldsymbol{\sigma} = -f \in L^2(\Omega)$, thus the appropriate solution space is

$$Q := \{\boldsymbol{\tau} \in L^2(\Omega)^d : A : \nabla\boldsymbol{\tau} \in L^2(\Omega)\}.$$

By introducing the new variable $\boldsymbol{\sigma}$, the requirement of the solution space V of (1.1) is reduced to $H_0^1(\Omega) \times Q$ of (2.4). The following lemma and theorem discuss the equivalence of (1.1) and (2.4).

LEMMA 2.4. *Assume that $f \in L^2(\Omega)$; then a solution $u \in H_0^1(\Omega)$ of* (1.1) *is also in V.*

Proof. The result follows from the fact $A : D^2 u = -f \in L^2(\Omega)$. ◻

THEOREM 2.5. *Assume that $f \in L^2(\Omega)$; then $u \in V$ is a solution of* (1.1) *if and only if the pair $(u, \boldsymbol{\sigma}) \in H_0^1(\Omega) \times Q$ is a solution of the first-order system* (2.4).

Proof. If $u \in V$ is a solution of (1.1), let $\boldsymbol{\sigma} = \nabla u$, then $\boldsymbol{\sigma} \in L^2(\Omega)^d$ and $-A : \nabla\boldsymbol{\sigma} = -A : \nabla(\nabla u) = -A : D^2 u = f \in L^2(\Omega)$. Thus $(u, \boldsymbol{\sigma}) \in H_0^1(\Omega) \times Q$ satisfies the first-order system (2.4).

Assume that $(u, \boldsymbol{\sigma}) \in H_0^1(\Omega) \times Q$ is a solution of the first-order system (2.4). Since $\boldsymbol{\sigma} \in Q$ and $\nabla u = \boldsymbol{\sigma}$, $A : \nabla\boldsymbol{\sigma} = A : D^2 u$ is well defined and belongs to L^2. So $A : D^2 u = -f \in L^2(\Omega)$. Thus u belongs to V and satisfies (1.1). ◻

Remark 2.6. The above theorem implies that $u = 0$ is the unique solution of (1.1) if and only if $(u, \boldsymbol{\sigma}) = (0, \mathbf{0})$ is the unique solution of (2.4).

Let $\mathcal{T} = \{K\}$ be a triangulation of Ω using simplicial elements. The mesh $\mathcal{T}$ is assumed to be shape-regular, but it needs not to be quasi-uniform. Let h_K be the diameter of the element $K \in \mathcal{T}$.

Two versions of least-squares functionals are introduced:

$$\mathcal{J}_h(v, \boldsymbol{\tau}; f) := \sum_{K \in \mathcal{T}} h_K^2 \|f + A : \nabla\boldsymbol{\tau}\|_{0,K}^2 + \|\boldsymbol{\tau} - \nabla v\|_0^2 \quad \forall (v, \boldsymbol{\tau}) \in H_0^1(\Omega) \times Q, \tag{2.5}$$

$$\mathcal{J}_0(v, \boldsymbol{\tau}; f) := \|f + A : \nabla\boldsymbol{\tau}\|_0^2 + \|\boldsymbol{\tau} - \nabla v\|_0^2 \quad \forall (v, \boldsymbol{\tau}) \in H_0^1(\Omega) \times Q. \tag{2.6}$$

The functionals $\mathcal{J}_h$ and $\mathcal{J}_0$ are called the W-version and the L^2-version, respectively. The notation $\mathcal{J}$ is used to denote both $\mathcal{J}_h$ and $\mathcal{J}_0$ when two formulations can be presented in a unified framework and no confusion is caused.

The least-squares minimization problem is as follows: seek $(u, \boldsymbol{\sigma}) \in H_0^1(\Omega) \times Q$ such that

$$\mathcal{J}(u, \boldsymbol{\sigma}; f) = \inf_{(v, \boldsymbol{\tau}) \in H_0^1(\Omega) \times Q} \mathcal{J}(v, \boldsymbol{\tau}; f). \tag{2.7}$$

The Euler–Lagrange formulations are as follows: seek $(u, \boldsymbol{\sigma}) \in H_0^1(\Omega) \times Q$ such that

$$a_h((u, \boldsymbol{\sigma}), (v, \boldsymbol{\tau})) = -\sum_{K \in \mathcal{T}} h_K^2 (f, A : \nabla\boldsymbol{\tau})_K \quad \forall (v, \boldsymbol{\tau}) \in H_0^1(\Omega) \times Q, \tag{2.8}$$

and seek $(u, \boldsymbol{\sigma}) \in H_0^1(\Omega) \times Q$ such that

$$a_0((u, \boldsymbol{\sigma}), (v, \boldsymbol{\tau})) = -(f, A : \nabla\boldsymbol{\tau}) \quad \forall (v, \boldsymbol{\tau}) \in H_0^1(\Omega) \times Q, \tag{2.9}$$

where for all $(w, \boldsymbol{\rho})$ and $(v, \boldsymbol{\tau}) \in H_0^1(\Omega) \times Q$, the bilinear forms are defined as

$$a_h((w, \boldsymbol{\rho}), (v, \boldsymbol{\tau})) := (\boldsymbol{\rho} - \nabla w, \boldsymbol{\tau} - \nabla v) + \sum_{K \in \mathcal{T}} h_K^2 (A : \nabla\boldsymbol{\rho}, A : \nabla\boldsymbol{\tau})_K, \tag{2.10}$$

and

$$a_0((w, \boldsymbol{\rho}), (v, \boldsymbol{\tau})) := (\boldsymbol{\rho} - \nabla w, \boldsymbol{\tau} - \nabla v) + (A : \nabla\boldsymbol{\rho}, A : \nabla\boldsymbol{\tau}). \tag{2.11}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

LEMMA 2.7. *Assume that $A\colon D^2u = 0$ in Ω and $u = 0$ on $\partial\Omega$ has a unique solution $u = 0$; then the following are norms for $(v, \boldsymbol{\tau}) \in H_0^1(\Omega) \times Q$:*

$$\|\!|(v,\boldsymbol{\tau})|\!\|_h^2 := \sum_{K\in\mathcal{T}} h_K^2 \|A\colon\nabla\boldsymbol{\tau}\|_{0,K}^2 + \|\boldsymbol{\tau} - \nabla v\|_0^2$$

$$\textit{and} \quad \|\!|(v,\boldsymbol{\tau})|\!\|_0^2 := \|A\colon\nabla\boldsymbol{\tau}\|_0^2 + \|\boldsymbol{\tau} - \nabla v\|_0^2.$$

We use $\|\!|(v,\boldsymbol{\tau})|\!\|$ to denote both versions when no confusion is caused. With Remark 2.6, the lemma can be easily proved by checking conditions of a norm definition.

It is easy to check that the following bounds are true:

$$C\|\!|(v,\boldsymbol{\tau})|\!\|_h \le \sum_{K\in\mathcal{T}} h_K \|A\colon\nabla\boldsymbol{\tau}\|_{0,K} + \|\nabla v\|_0 + \|\boldsymbol{\tau}\|_0 \tag{2.12}$$

$$\text{and} \quad \|\!|(v,\boldsymbol{\tau})|\!\|_0 \le \|A\colon\nabla\boldsymbol{\tau}\|_0 + \|\nabla v\|_0 + \|\boldsymbol{\tau}\|_0. \tag{2.13}$$

THEOREM 2.8. *Assume that $f \in L^2(\Omega)$, the coefficient matrix A, and the domain Ω are nice enough such that Assumption* 2.1 *is true; then the least-squares problem* (2.7) *has a unique solution $(u, \boldsymbol{\sigma}) \in H_0^1(\Omega) \times Q$.*

Proof. The theorem is a consequence of the Lax–Milgram theorem and the facts that $\|\!|(v,\boldsymbol{\tau})|\!\|$ is a norm in $H_0^1(\Omega) \times Q$ (Lemma 2.7) and the term $(f, A : \nabla\boldsymbol{\tau})$ is meaningful since $f \in L^2(\Omega)$. $\square$

The above argument to show the existence and uniqueness of the least-squares formulation is useful when the existence and uniqueness of the PDE are obtained from various nonvariational techniques. A similar argument is used to prove the stability of least-squares formulations for the linear transport equation in [34, 35].

3. Least-squares finite element methods.

In this section, LSFEMs based on the least-squares minimization problems are developed. The a priori and a posteriori error estimates with respect to the least-squares norms $\|\!|(\cdot,\cdot)|\!\|$ are derived.

3.1. LSFEMs.

For an element $K \in \mathcal{T}$ and an integer $k \ge 0$, let $P_k(K)$ be the space of polynomials with degrees less than or equal to k. Define the finite element spaces S_k and $S_{k,0}$, $k \ge 1$, as follows:

$$S_k := \{v \in H^1(\Omega)\colon v|_K \in P_k(K)\ \forall\, K \in \mathcal{T}\} \quad \text{and} \quad S_{k,0} := S_k \cap H_0^1(\Omega).$$

We then define LSFEMs.
(W-LSFEM problem) Seek $(u_h, \boldsymbol{\sigma}_h) \in S_{k,0} \times S_{k-1}^d$, $k \ge 2$, such that

$$\mathcal{J}_h(u_h, \boldsymbol{\sigma}_h; f) = \inf_{(v,\boldsymbol{\tau})\in S_{k,0}\times S_{k-1}^d} \mathcal{J}_h(v, \boldsymbol{\tau}; f). \tag{3.1}$$

Or equivalently, find $(u_h, \boldsymbol{\sigma}_h) \in S_{k,0} \times S_{k-1}^d$, $k \ge 2$, such that

$$a_h((u_h, \boldsymbol{\sigma}_h), (v_h, \boldsymbol{\tau}_h)) = -\sum_{K\in\mathcal{T}} h_K^2 (f, A\colon\nabla\boldsymbol{\tau}_h) \quad \forall (v_h, \boldsymbol{\tau}_h) \in S_{k,0} \times S_{k-1}^d. \tag{3.2}$$

(L^2-LSFEM problem) Seek $(u_h, \boldsymbol{\sigma}_h) \in S_{1,0} \times S_1^d$ such that

$$\mathcal{J}_0(u_h, \boldsymbol{\sigma}_h; f) = \inf_{(v,\boldsymbol{\tau})\in S_{1,0}\times S_1^d} \mathcal{J}_0(v, \boldsymbol{\tau}; f). \tag{3.3}$$

Or equivalently, find $(u_h, \boldsymbol{\sigma}_h) \in S_{1,0} \times S_1^d$ such that

$$a_0((u_h, \boldsymbol{\sigma}_h), (v_h, \boldsymbol{\tau}_h)) = -(f, A\colon\nabla\boldsymbol{\tau}_h) \quad \forall (v, \boldsymbol{\tau}_h) \in S_{1,0} \times S_1^d. \tag{3.4}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/02/21 to 59.64.129.2. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Downloaded 01/02/21 to 59.64.129.2. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

The existence and uniqueness of the LSFEM problems are obvious since

$$S_k^d \subset H^1(\Omega)^d = \left\{ \boldsymbol{\tau} \in L^2(\Omega)^d : \nabla \boldsymbol{\tau} \in L^2(\Omega)^d \right\} \subset Q \text{ and } S_{k,0} \subset H_0^1(\Omega).$$

Remark 3.1. For the approximation space of $u \in H^1(\Omega)$, the H^1-conforming finite element space is a natural choice.

Now, we discuss the choice of finite element subspace of Q. As a comparison, consider the equation in divergence form with $u_d \in H_0^1(\Omega)$ such that

$$-\nabla \cdot (A \nabla u_d) = f \quad \text{in } \Omega. \tag{3.5}$$

It is well known that the flux $A\nabla u_d \in H(\text{div}; \Omega)$ for $f \in L^2(\Omega)$ for (3.5). The space $H(\text{div}; \Omega)$ is well studied [31, 8]. Its finite element subspaces such as the Raviart–Thomas space [8] with good approximation properties are also well known.

For the equation in divergence form with a piecewise constant (discontinuous) A, the tangential component of the flux $A\nabla u_d$ and the normal component of ∇u_d may be discontinuous; see discussions in [17, 18, 12]. Thus

$$A\nabla u_d \in H(\text{div}; \Omega) \quad \text{but} \quad A\nabla u_d \notin H(\mathbf{curl}; \Omega),$$

$$\nabla u_d \in H(\mathbf{curl}; \Omega) \quad \text{but} \quad \nabla u_d \notin H(\text{div}; \Omega),$$

for a discontinuous A and u_d being the solution of (3.5).

For the elliptic equation in nondivergence form (1.1), the property of the space Q is less known. Similar to the space V, for a vector function $\boldsymbol{\tau} \in Q$, we can expect that even though an individual $\partial \boldsymbol{\tau}/\partial x_i$, $i = 1, \dots, d$, is not in $L^2(\Omega)$, due to the cancellation or good properties of A, $A : \nabla \boldsymbol{\tau}$ may belong to $L^2(\Omega)$. But to develop a numerical method for a general nondivergence elliptic equation, any of this information cannot be assumed or used, and an A-intrinsic finite element subspace of Q as $H(\text{div})$-conforming finite elements cannot be designed.

On the other hand, for the nondivergence equation (1.1), the solution u is in $H^2(\Omega)$ with a convex domain and a discontinuous A satisfying the Cordes condition. In such cases, for a discontinuous A, $\nabla u \in H^1(\Omega)^d$ is still continuous in a weak sense, while $A\nabla u$ cannot be continuous in its normal or tangential component. So

$$\nabla u \in H^1(\Omega)^d \subset H(\text{div}; \Omega) \cap H(\mathbf{curl}; \Omega), A\nabla u \notin H(\mathbf{curl}; \Omega), A\nabla u \notin H(\text{div}; \Omega), \tag{3.6}$$

for a discontinuous A satisfying Cordes condition, a convex Ω, and u being the solution of (1.1). For a nonconvex Ω, the solution u of (1.1) may only belong to $H^{3/2+\epsilon}$ for some $\epsilon > 0$, so $\nabla u \in H^{0.5+\epsilon}(\Omega)^d$. This suggests that for (1.1) with a general coefficient A and a general domain Ω, ∇u is very close to continuous in both normal and tangential components.

From the above discussions, $\boldsymbol{\sigma} = \nabla u$ is the right choice of unknown in our method, and $S_k^d \subset H^1(\Omega)^d \subset Q$ is the right choice of the approximation space. This is different from the equation in divergence form (1.2), where the flux $A\nabla u_d \in H(\text{div})$ is the natural choice of the extra unknown and the $H(\text{div})$-conforming Raviart–Thomas space is shown to be a better choice of approximation space compared to the continuous finite element space; see, for example, discussions in [6, 17, 18, 12].

Remark 3.2. The gradient $\boldsymbol{\sigma} = \nabla u$ also satisfies $\nabla \times \boldsymbol{\sigma} = 0$ in Ω and $\mathbf{n} \times \boldsymbol{\sigma} = 0$ on $\partial\Omega$, where $\mathbf{n}$ is the unit outer normal to the boundary. It is also possible to add the equation $\nabla \times \boldsymbol{\sigma} = 0$ (with appropriate boundary conditions on the space) to the first-order system (2.4) and construct corresponding least-squares problems. Such a formulation may be more suitable for a multigrid solver. A discussion of such an approach of LSFEM for the elliptic equation in divergence form can be found in [15, 5, 4].

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

3.2. A priori error estimates.

THEOREM 3.3. *Let $(u, \boldsymbol{\sigma})$ be the solution of least-squares variational problem* (2.7)*. Let $(u_h, \boldsymbol{\sigma}_h) \in S_{k,0} \times S_{k-1}^d$, $k \geq 2$, be the solution of the W-LSFEM problem* (3.1)*. The following best approximation result holds:*

$$\|\!|(u - u_h, \boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\|\!|_h \leq \inf_{(v_h, \boldsymbol{\tau}_h) \in S_{k,0} \times S_{k-1}^d} \|\!|(u - v_h, \boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\|\!|_h. \tag{3.7}$$

Let $(u_h, \boldsymbol{\sigma}_h) \in S_{1,0} \times S_1^d$ be the solution of the L^2-LSFEM problem (3.3)*; the following best approximation result holds:*

$$\|\!|(u - u_h, \boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\|\!|_0 \leq \inf_{(v_h, \boldsymbol{\tau}_h) \in S_{1,0} \times S_1^d} \|\!|(u - v_h, \boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\|\!|_0. \tag{3.8}$$

Proof. The proof of the best approximation result is standard. ∎

THEOREM 3.4 (a priori error estimate for the W-LSFEM). *Assume the solution $u \in H^{r+1}(\Omega)$, for some $r \geq 1$, and $(u_h, \boldsymbol{\sigma}_h) \in S_{k,0} \times S_{k-1}^d$, $k \geq 2$, is the solution of the weighted LSFEM problem* (3.1)*; then there exists a constant $C > 0$ independent of the mesh size h such that*

$$\|\!|(u - u_h, \boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\|\!|_h \leq C h^{\min(k,r)} \|u\|_{\min(k+1,r+1)}. \tag{3.9}$$

Proof. The a priori result is a direct consequence of Theorem 3.3, (2.12), and the approximation properties of functions in S_k. ∎

THEOREM 3.5 (a priori error estimate for the L^2-LSFEM). *Assume the solution $u \in H^3(\Omega)$ and $(u_h, \boldsymbol{\sigma}_h) \in S_{1,0} \times S_1^d$ is the solution of the L^2-LSFEM problem* (3.3)*; then there exists a constant $C > 0$ independent of the mesh size h such that*

$$\|\!|(u - u_h, \boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\|\!|_0 \leq C h \|u\|_3. \tag{3.10}$$

Proof. Let v_h be the interpolation of u in $S_{1,0}$ and $\boldsymbol{\tau}_h$ be the interpolation of ∇u in S_1^d. By the inequality (2.13) and approximation properties of S_1, we get

$$\begin{aligned} C\|\!|(u - u_h, \boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\|\!|_0 &\leq \|A : \nabla(\boldsymbol{\sigma} - \boldsymbol{\tau}_h)\|_0 + \|\boldsymbol{\sigma} - \boldsymbol{\tau}_h\|_0 + \|\nabla(u - v_h)\|_0 \\ &\leq Ch\|\nabla u\|_2 + Ch\|u\|_2 + Ch^2\|\nabla u\|_2 \leq Ch\|u\|_3. \end{aligned}$$ ∎

Remark 3.6. From the above discussion of a priori estimates, we can clearly see that with respect to the least-squares norm, the weighted version is optimal when the regularity of the solution is high and a suitable high-order finite element pair is used. For the L^2-LSFEM, the optimal interpolation order for the L^2-norm of $\boldsymbol{\sigma}$ is 2, which is one order higher than the other two components, and thus suboptimal. So high-order approximations of the L^2-LSFEM are not suggested. But the L^2-LSFEM can use the simplest linear conforming finite element space for u and has reasonable approximation orders. For example, the L^2-estimates of $u - u_h$ are of the same order as the $S_{2,0} \times S_1^d$ W-LSFEM if assuming enough smoothness of the coefficient A and the solution u; see Theorems 4.8 and 5.1 and our numerical tests.

Remark 3.7. In our a priori error estimates, in order to get a convergence order, we assumed that the regularity of u is at least of H^2 or H^3. To discuss the convergence without a high regularity, we first discuss a case that the coefficient matrix A is nice enough such that for any $\epsilon > 0$, there exists a $\boldsymbol{\sigma}^\epsilon \in C^\infty(\Omega)^d$ such that

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}^\epsilon\|_0 \leq \epsilon \quad \text{and} \quad \|f - A : \nabla \boldsymbol{\sigma}^\epsilon\|_0 \leq \epsilon. \tag{3.11}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/02/21 to 59.64.129.2. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Downloaded 01/02/21 to 59.64.129.2. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Note that this condition is weaker than $\|\boldsymbol{\sigma} - \boldsymbol{\sigma}^\epsilon\|_1 \le \epsilon$, since $\nabla\boldsymbol{\sigma}$ may not be in L^2. If (3.11) is true, then a convergence result is easy to prove for $u \in H_0^1(\Omega)$ only.

Second, consider the special case that $A|_K$, the restriction of the coefficient matrix on each $K \in \mathcal{T}$, is a constant matrix. For a low regularity problem, we use the linear finite element to approximate $\boldsymbol{\sigma}$, then $A:\nabla\boldsymbol{\tau}_h$ contains a piecewise constant on each element K, and we have

$$\|f - A:\nabla\boldsymbol{\tau}_h\|_{0,K} \le C h_K^r \|f\|_{r,K} \quad \text{for } 0 < r \le 1.$$

This can be used to establish the convergence for a low regularity problem.

For the standard LSFEM for the PDEs in divergence form, such problems do not exist since the Raviart–Thomas element is used to approximate $\boldsymbol{\sigma}_d = -A\nabla u$, and the regularity requirement is on $f = \nabla\cdot\boldsymbol{\sigma}_d$, which is weaker than the requirement on $\nabla\boldsymbol{\sigma} = D^2 u$. Again, this is due to the special structure of the divergence equation, which is not available for the general nondivergence equation.

3.3. A posteriori error estimates. The least-squares functional can be used to define the following fully computable a posteriori local indicator and global error estimator.

Let $(u,\boldsymbol{\sigma})$ be the solution of least-squares variational problem (2.7), and let $(u_h,\boldsymbol{\sigma}_h) \in S_{k,0} \times S_{k-1}^d$, $k \ge 2$, be the solution of the W-LSFEM problem (3.1); define

$$\eta_{h,K}^2 := h_K^2\|f + A:\nabla\boldsymbol{\sigma}_h\|_{0,K}^2 + \|\boldsymbol{\sigma}_h - \nabla u_h\|_{0,K}^2 \quad \forall K \in \mathcal{T}$$

$$\text{and} \quad \eta_h^2 := \sum_{K\in\mathcal{T}} \eta_{h,K}^2 = \sum_{K\in\mathcal{T}} h_K^2\|f + A:\nabla\boldsymbol{\sigma}_h\|_{0,K}^2 + \|\boldsymbol{\sigma}_h - \nabla u_h\|_0^2.$$

By the standard approach in a posteriori error estimates in LSFEM, we have

$$\eta_h = |||(u-u_h,\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)|||_h \quad \text{and} \quad \eta_{h,K} = |||(u-u_h,\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)|||_{h,K}. \tag{3.12}$$

By a local version of (2.12), the following local efficiency bound is also true:

$$C\eta_{h,K} \le h_K\|A:\nabla(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_{0,K} + \|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{0,K} + \|\nabla(u-u_h)\|_{0,K} \quad \forall K \in \mathcal{T}. \tag{3.13}$$

For the L^2-LSFEM, the a posteriori error estimator η_0 can be defined accordingly, and the corresponding results can be obtained in a similar fashion.

4. More a priori error estimates for the W-LSFEM.

4.1. Coefficient A and h-weighted broken H^2-norm estimate. For the W-LSFEM, $k \ge 2$ polynomial spaces are used to approximate u, and due to the nondivergence structure of the equation, the W-LSFEM can also be viewed as a method via an approximation of the D^2 operator. In this subsection, a coefficient A and h-weighted broken H^2-norm estimate of the error is derived.

The following notation is used to denote a mesh-dependent norm: $\|hv\|_0 := (\sum_{K\in\mathcal{T}} h_K^2\|v\|_{0,K}^2)^{1/2}$. For example, $\|hA:D_h^2 v_h\|_0 = (\sum_{K\in\mathcal{T}} h_K^2\|A:D^2 v_h\|_{0,K}^2)^{1/2}$.

In two dimensions, define $\tilde{V}_{k+2,2d}$ to be the C^1-conforming finite element space of degree $k+2$, $k \ge 2$ on $\mathcal{T}$, which is the high-order version of the classical Hsieh–Clough–Tocher macroelement [25]. In three dimensions, let $\tilde{V}_{h,3d}$ be the classical C^1-conforming piecewise cubic Hsieh–Clough–Tocher macroelement space associated with the mesh $\mathcal{T}$; see [22, 39].

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

The proof of the following lemma can be found in [11, 29] for the two-dimensional case and in [39] for the three-dimensional case. Although in these papers, the results are all presented in a global setting, a careful look into their proofs can find that the result is also true locally due to the shape regularity assumption.

LEMMA 4.1. *For any function $v_h \in S_{k,0}$, where $2 \le k \le 3$ if the space dimension $d = 3$, and $k \ge 2$ if $d = 2$, there exists an averaging linear map $E_h : S_k \to \tilde{V}_{k+2,2d} \cap H_0^1(\Omega)$ in two dimensions and $E_h : S_k \to \tilde{V}_{h,3d} \cap H_0^1(\Omega)$ in three dimensions, such that the following estimate is true:*

$$\|v_h - E_h v_h\|_{0,K} \le C h_K^{3/2} \sum_{F \in \mathcal{E}_K} \|[\![\nabla v_h \cdot \mathbf{n}]\!]\|_{0,F} \quad \forall K \in \mathcal{T}, \tag{4.1}$$

where $\mathcal{E}_K$ is the collection of interior edges on elements that share a common node with the element K in two dimensions and is the collection of interior faces on elements that share a common vertex or a common edge middle point with the element K in three dimensions.

LEMMA 4.2. *There exists a constant $C > 0$ independent of the mesh size h such that, for any $(v_h, \boldsymbol{\tau}_h) \in S_{k,0} \times S_{k-1}^d$, with $2 \le k \le 3$ if the space dimension $d = 3$, and $k \ge 2$ if $d = 2$, the following estimates are true:*

$$\|\nabla(v_h - E_h v_h)\|_0 \le C\|\boldsymbol{\tau}_h - \nabla v_h\|_0, \quad \|D_h^2(v_h - E_h v_h)\|_0 \le C h^{-1}\|\boldsymbol{\tau}_h - \nabla v_h\|_0, \tag{4.2}$$

$$\text{and} \quad \|\boldsymbol{\tau}_h - \nabla(E_h v_h)\|_0 \le C\|\boldsymbol{\tau}_h - \nabla v_h\|_0. \tag{4.3}$$

Proof. Since $\boldsymbol{\tau}_h \in (H^1(\Omega))^d$, then $[\![\boldsymbol{\tau}_h \cdot \mathbf{n}]\!]_F = 0$ on an interior edge (two-dimensional)/face (three-dimensional) F. By the discrete trace inequality,

$$\|[\![\nabla v_h \cdot \mathbf{n}]\!]\|_{0,F} = \|[\![\nabla v_h - \boldsymbol{\tau}_h]\!] \cdot \mathbf{n}\|_{0,F} \le C h_F^{-1/2} \|\nabla v_h - \boldsymbol{\tau}_h\|_{0,\omega_F},$$

where ω_F is the collection of two elements that share the common F. By (4.1),

$$\|v_h - E_h v_h\|_{0,K} \le C h_K \sum_{F \in \mathcal{E}_K} \|\nabla v_h - \boldsymbol{\tau}_h\|_{0,\omega_F}. \tag{4.4}$$

Combined with the inverse inequalities $\|\nabla(v_h - E_h v_h)\|_{0,K} \le C h_K^{-1}\|v_h - E_h v_h\|_{0,K}$, $\|D_h^2(v_h - E_h v_h)\|_{0,K} \le C h_K^{-2}\|v_h - E_h v_h\|_{0,K}$, and the shape regularity of the mesh $\mathcal{T}$, inequalities in (4.2) are proved. The inequality of (4.3) is a simple consequence of the first inequality of (4.2) and the triangle inequality. ∎

LEMMA 4.3. *The following inequality holds for all $(v_h, \boldsymbol{\tau}_h) \in S_{k,0} \times S_{k-1}^d$, with $2 \le k \le 3$ if the space dimension $d = 3$, and $k \ge 2$ if $d = 2$:*

$$\|hA : D_h^2 v_h\|_0 \le C(\|hA : \nabla\boldsymbol{\tau}_h\|_0 + \|\boldsymbol{\tau}_h - \nabla v_h\|_0) \le C|\!|\!|(v_h, \boldsymbol{\tau}_h)|\!|\!|_h. \tag{4.5}$$

Proof. Let $\tilde{v}_h = E_h v_h$; by the triangle inequality, we get

$$\|hA : D_h^2 v_h\|_0 \le \|hA : D_h^2(v_h - \tilde{v}_h)\|_0 + \|hA : D^2 \tilde{v}_h\|_0. \tag{4.6}$$

For the term $\|hA : D_h^2(v_h - \tilde{v}_h)\|_0$, by the inverse estimate, the fact that $A \in L^\infty(\Omega)^{d\times d}$, and the second inequality in (4.2), we have

$$\|hA : D_h^2(v_h - \tilde{v}_h)\|_0 \le C\|\nabla(v_h - \tilde{v}_h)\|_0 \le C\|\boldsymbol{\tau}_h - \nabla v_h\|_0. \tag{4.7}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/02/21 to 59.64.129.2. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Downloaded 01/02/21 to 59.64.129.2. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

On the other hand, by the triangle inequality,

$$\|hA : D^2\tilde{v}_h\|_0 \le \|hA : \nabla\boldsymbol{\tau}_h\|_0 + \|hA : \nabla(\boldsymbol{\tau}_h - \nabla\tilde{v}_h)\|_0.$$

By the inverse estimate, the fact that $A \in L^\infty(\Omega)^{d\times d}$, and the inequality (4.3),

$$\|hA : \nabla(\boldsymbol{\tau}_h - \nabla\tilde{v}_h)\|_0 \le C\|\boldsymbol{\tau}_h - \nabla\tilde{v}_h\|_0 \le C\|\boldsymbol{\tau}_h - \nabla v_h\|_0.$$

Combining the above results, the first inequality of the lemma is proved. The second inequality is a consequence of a simple calculation. ◻

THEOREM 4.4. *Assume the solution $u \in H^{r+1}(\Omega)$, for some $r > 1$, and $(u_h, \boldsymbol{\sigma}_h) \in S_{k,0} \times S^d_{k-1}$, with $2 \le k \le 3$ if the space dimension $d = 3$, and $k \ge 2$ if $d = 2$, be the solution of the W-LSFEM* (3.1)*; then there exists a constant $C > 0$ independent of the mesh size h such that the following error estimate is true:*

$$\|hA : D_h^2(u - u_h)\|_0 \le Ch^{\min(k,r)}\|u\|_{\min(k+1,r+1)}. \tag{4.8}$$

Proof. By the triangle inequality, for an arbitrary $v_h \in S_{k,0}$,

$$\|hA : D_h^2(u - u_h)\|_0 \le \|hA : D_h^2(u - v_h)\|_0 + \|hA : D_h^2(u_h - v_h)\|_0.$$

By (4.5), for an arbitrary $\boldsymbol{\tau}_h \in S^d_{k-1}$,

$$C\|hA : D_h^2(u_h - v_h)\|_0 \le |\!|\!|(u_h - v_h, \boldsymbol{\sigma}_h - \boldsymbol{\tau}_h)|\!|\!|_h.$$

On the other hand, by the orthogonality result $a_h((u-u_h, \boldsymbol{\sigma}-\boldsymbol{\sigma}_h), (u_h-v_h, \boldsymbol{\sigma}_h-\boldsymbol{\tau}_h)) = 0$, for all $(v_h, \boldsymbol{\tau}_h) \in S_{k,0} \times S^d_{k-1}$,

$$\begin{aligned}|\!|\!|(u_h - v_h, \boldsymbol{\sigma}_h - \boldsymbol{\tau}_h)|\!|\!|_h^2 &= a_h((u_h - v_h, \boldsymbol{\sigma}_h - \boldsymbol{\tau}_h), (u_h - v_h, \boldsymbol{\sigma}_h - \boldsymbol{\tau}_h))\\ &= a_h((u - v_h, \boldsymbol{\sigma} - \boldsymbol{\sigma}_h), (u_h - v_h, \boldsymbol{\sigma}_h - \boldsymbol{\tau}_h))\\ &\le |\!|\!|(u - v_h, \boldsymbol{\sigma} - \boldsymbol{\sigma}_h)|\!|\!|_h |\!|\!|(u_h - v_h, \boldsymbol{\sigma}_h - \boldsymbol{\tau}_h)|\!|\!|_h.\end{aligned}$$

Thus, $|\!|\!|(u_h - v_h, \boldsymbol{\sigma}_h - \boldsymbol{\tau}_h)|\!|\!|_h \le |\!|\!|(u - v_h, \boldsymbol{\sigma} - \boldsymbol{\sigma}_h)|\!|\!|_h$ for all $(v_h, \boldsymbol{\sigma}_h) \in S_{k,0} \times S^d_{k-1}$. Combining the above results, we get

$$\|hA : D_h^2(u - u_h)\|_0 \le \inf_{(v_h,\boldsymbol{\sigma}_h)\in S_{k,0}\times S^d_{k-1}} (\|hA : D_h^2(u - v_h)\|_0 + C|\!|\!|(u - v_h, \boldsymbol{\sigma} - \boldsymbol{\sigma}_h)|\!|\!|_h).$$

By approximation properties of functions in S_k, the theorem is proved. ◻

4.2. Error estimates based on an assumption of the nondivergence operator. In this subsection, we prove error estimates based on the following assumption on the nondivergence operator.

Assumption 4.5 (assumption on the nondivergence operator). It is assumed that for the given A and the domain Ω, the following result holds for the operator $A : D^2 v$ in nondivergence form:

$$\|v\|_{1+\delta,\Omega} \le C\|A : D^2 v\|_{0,\Omega} \quad \forall v \in V, \text{ for some } \delta \in [0, 1], \tag{4.9}$$

where the space V is defined in (2.1). For a special case $\delta = 1$, (4.9) is (2.2).

In this paper, we assume the domain can be nonconvex; thus only a weaker assumption on the operator (4.9) holds.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/02/21 to 59.64.129.2. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

THEOREM 4.6. *Assume that the assumption of the operator* (4.9) *is true, the mesh is quasi-uniform,* $(u_h, \boldsymbol{\sigma}_h) \in S_{k,0} \times S^d_{k-1}$*, with* $k = 2$ *or* 3 *if the space dimension* $d = 3$*, and* $k \ge 2$ *for* $d = 2$*, is the W-LSFEM solution; then for* $u \in H^{1+r}(\Omega)$*,* $r \ge 1$*, we have*

$$\|u - E_h u_h\|_{1+\delta} \le C h^{\min(k-1,r-1)} \|u\|_{\min(k+1,r+1)} \quad \text{for some } \delta \in [0,1]. \tag{4.10}$$

Specifically, for the weakest case $\delta = 0$*,*

$$\|u - E_h u_h\|_1 \le C h^{\min(k-1,r-1)} \|u\|_{\min(k+1,r+1)}, \tag{4.11}$$

and for the strongest case, $\delta = 1$*,*

$$\|u - E_h u_h\|_2 \le C h^{\min(k-1,r-1)} \|u\|_{\min(k+1,r+1)}. \tag{4.12}$$

Here, E_h *is the Hsieh–Clough–Toucher-element averaging operator defined in Lemma* 4.1.

Proof. Let $\tilde{u}_h = E_h u_h$; then by the assumption of the operator (4.9) and the triangle inequality,

$$C\|u - \tilde{u}_h\|_{1+\delta,\Omega} \le \|A : D^2(u - \tilde{u}_h)\|_0 \le \|A : D_h^2(u - u_h)\|_0 + \|A : D_h^2(u_h - \tilde{u}_h)\|_0.$$

For the first term on the right-hand side, by Theorem 4.4, $k \ge 2$, and the fact the mesh is quasi-uniform,

$$\|A : D_h^2(u - u_h)\|_{0,\Omega} \le C h^{\min(k-1,r-1)} \|u\|_{\min(k+1,r+1)}.$$

For the second term, by the same argument as in (4.7),

$$\|A : D^2(u_h - \tilde{u}_h)\|_{0,\Omega} \le C h^{-1} \|\boldsymbol{\sigma}_h - \nabla u_h\|_0.$$

Then by the fact that $\boldsymbol{\sigma} = \nabla u$ and Theorem 3.4,

$$\begin{aligned} \|\boldsymbol{\sigma}_h - \nabla u_h\|_0 &\le \|\boldsymbol{\sigma}_h - \boldsymbol{\sigma} + \nabla(u_h - u)\|_0 \le |\!|\!|(u - v_h, \boldsymbol{\sigma} - \boldsymbol{\sigma}_h)|\!|\!|_h \\ &\le C h^{\min(k,r)} \|u\|_{\min(k+1,r+1)}. \end{aligned}$$

Combining the above estimates, the theorem is proved. ∎

THEOREM 4.7. *Assume that the assumption of the operator* (4.9) *is true, the mesh is quasi-uniform, the solution* $u \in H^{r+1}(\Omega)$*, for some* $r > 1$*, and the numerical solution* $(u_h, \boldsymbol{\sigma}_h) \in S_{k,0} \times S^d_{k-1}$*, with* $2 \le k \le 3$ *if the space dimension* $d = 3$*, and* $k \ge 2$ *for* $d = 2$*, be the solution of the W-LSFEM* (3.1)*; then we also have the following* H^1*-estimate:*

$$\|\nabla(u - u_h)\|_0 \le C h^{\min(k-1,r-1)} \|u\|_{\min(k+1,r+1)}. \tag{4.13}$$

If we further assume that (2.2) *is true, then the following broken* H^2*-norm estimate is also true:*

$$\|D_h^2(u - u_h)\|_0 \le C h^{\min(k-1,r-1)} \|u\|_{\min(k+1,r+1)}. \tag{4.14}$$

Proof. Let $\tilde{u}_h = E_h u_h$, then by the triangle inequality,

$$\|\nabla(u - u_h)\|_{0,\Omega} \le \|\nabla(u - \tilde{u}_h)\|_{0,\Omega} + \|\nabla(\tilde{u}_h - u_h)\|_{0,\Omega}.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/02/21 to 59.64.129.2. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

The first term is good by Theorem 4.6. For the second term, by (4.2) and the a priori error estimate result of Theorem 3.4,

$$\begin{aligned}\|\nabla(\tilde{u}_h - u_h)\|_{0,\Omega} &\le C\|\boldsymbol{\sigma}_h - \nabla u_h\|_0 \le C|\!|\!|(u - u_h, \boldsymbol{\sigma} - \boldsymbol{\sigma}_h)|\!|\!|_h \\ &\le Ch^{\min(k,r)}\|u\|_{\min(k+1,r+1)}.\end{aligned}$$

Then we prove (4.13).

For (4.14), similarly, we have

$$\|D_h^2(u - u_h)\|_{0,\Omega} \le \|D^2(u - \tilde{u}_h)\|_{0,\Omega} + \|D_h^2(\tilde{u}_h - u_h)\|_{0,\Omega}$$

and

$$\begin{aligned}\|D_h^2(\tilde{u}_h - u_h)\|_{0,\Omega} &\le Ch^{-1}\|\boldsymbol{\sigma}_h - \nabla u_h\|_0 \le Ch^{-1}|\!|\!|(u - u_h, \boldsymbol{\sigma} - \boldsymbol{\sigma}_h)|\!|\!|_h \\ &\le Ch^{\min(k-1,r-1)}\|u\|_{\min(k+1,r+1)}.\end{aligned}$$

The result (4.14) then can be proved by combining the estimates of $\|D^2(u - \tilde{u}_h)\|_{0,\Omega}$ and $\|D_h^2(\tilde{u}_h - u_h)\|_{0,\Omega}$. ∎

4.3. L^2-error estimate. In this subsection, we discuss the L^2-error estimate of the W-LSFEM with extra regularity conditions of the equation. The proof is based on a modification of the argument of Cai and Ku [16] for the LSFEM of the elliptic equations in divergence form. The existence of the weight h in the W-LSFEM adds extra difficulty to the L^2-error analysis, and it requires the polynomial degree to approximate u to be at least three to have an extra order.

Denote by

$$E = \boldsymbol{\sigma} - \boldsymbol{\sigma}_h \quad \text{and} \quad e = u - u_h$$

the respective errors of the gradient and the solution.

Assume that A is smooth enough that the operation $\nabla \cdot (\nabla \cdot (Az))$ is meaningful for a smooth z, where Az is a matrix with items $a_{i,j}z$, and the divergence of a matrix B is a column vector with each item being the divergence of the row of B.

Let $z \in H_0^1(\Omega)$ be the solution of the following equation:

$$\sum_{i,j=1}^{d} \partial_{ij}^2(a_{i,j}u) = \nabla \cdot (\nabla \cdot (Az)) = e \text{ in } \Omega, \quad z = 0 \text{ on } \partial\Omega. \tag{4.15}$$

We assume that both the original nondivergence PDE (1.1) and the dual equation (4.15) satisfy the full H^2-regularity:

$$\|u\|_2 \le C\|f\|_0 \quad \text{and} \quad \|z\|_2 \le C\|e\|_0. \tag{4.16}$$

In addition, it is also assumed that the solution of (1.1) satisfies the following stronger regularity assumption:

$$\|u\|_4 \le C\|f\|_2. \tag{4.17}$$

Making $\nabla \cdot (\nabla \cdot (Az))$ well defined does not require that $A \in C^1(\Omega)^{d\times d}$, but only $z \in \{v \in L^2(\Omega) : Av \in H(\text{div};\Omega)^d, \nabla \cdot (Av) \in H(\text{div};\Omega)\}$.

THEOREM 4.8. *Assuming that the mesh is quasi-uniform with a mesh size h, the regularity assumptions* (4.16) *and* (4.17) *are true, and the W-LSFEM solutions* $(u_h, \boldsymbol{\sigma}_h)$ *belong to* $S_{k,0} \times S_{k-1}^d$, *for* $k \ge 3$, *we have the following L^2-error estimate:*

$$\|u - u_h\|_0 \le Ch|\!|\!|(u - u_h, \boldsymbol{\sigma} - \boldsymbol{\sigma}_h)|\!|\!|_h. \tag{4.18}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/02/21 to 59.64.129.2. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Proof. Using the integration by parts, we have

$$\begin{aligned}\|e\|_0^2 &= (e,e) = (e, \nabla\cdot(\nabla\cdot(Az))) = -(\nabla e, \nabla\cdot(Az)) = (E-\nabla e, \nabla\cdot(Az)) - (E, \nabla\cdot(Az))\\ &= (E-\nabla e, \nabla\cdot(Az)) + (\nabla E, Az) = (E-\nabla e, \nabla\cdot(Az)) + (A:\nabla E, z).\end{aligned}$$

To match with the bilinear form of the W-LSFEM, two subauxiliary problems for $w_1 \in H_0^1(\Omega)$ and $w_2 \in H_0^1(\Omega)$ are introduced:

$$\left\{\begin{array}{lcl}\boldsymbol{\phi}_1 - \nabla w_1 &=& 0,\\ A:\nabla\boldsymbol{\phi}_1 &=& h^{-2}z,\end{array}\right. \quad \text{and} \quad \left\{\begin{array}{lcl}\boldsymbol{\phi}_2 - \nabla w_2 &=& \nabla\cdot(Az),\\ A:\nabla\boldsymbol{\phi}_2 &=& 0,\end{array}\right. \tag{4.19}$$

which are, in the PDE forms,

$$A:D^2 w_1 = h^{-2}z \quad \text{and} \quad A:D^2 w_2 = -A:\nabla(\nabla\cdot(Az)). \tag{4.20}$$

Let $w = w_1 + w_2$ and $\boldsymbol{\phi} = \boldsymbol{\phi}_1 + \boldsymbol{\phi}_2$, then $w = 0$ on $\partial\Omega$, and

$$\left\{\begin{array}{lcll}\boldsymbol{\phi} - \nabla w &=& \nabla\cdot(Az) & \text{in } \Omega,\\ A:\nabla\boldsymbol{\phi} &=& h^{-2}z & \text{in } \Omega.\end{array}\right. \tag{4.21}$$

Substitute (4.19) into the representation of $\|e\|_0^2$:

$$\begin{aligned}\|e\|_0^2 &= (E-\nabla e, \nabla\cdot(Az)) + (A:\nabla E, z)\\ &= a_h((e,E),(w_1,\boldsymbol{\phi}_1)) + a_h((e,E),(w_2,\boldsymbol{\phi}_2)).\end{aligned}$$

Letting $(w_{i,h}, \boldsymbol{\phi}_{i,h}) \in S_{k,0}\times S_{k-1}^d$ for $k \ge 3$ and $i = 1$ and 2 and using the orthogonality of the error equation, we have

$$\begin{aligned}\|e\|_0^2 &= a_h((e,E),(w_1 - w_{1,h}, \boldsymbol{\phi}_1 - \boldsymbol{\phi}_{1,h})) + a_h((e,E),(w_2 - w_{2,h}, \boldsymbol{\phi}_2 - \boldsymbol{\phi}_{2,h}))\\ &\le |\!|\!|(e,E)|\!|\!|_h (|\!|\!|(w_1 - w_{1,h}, \boldsymbol{\phi}_1 - \boldsymbol{\phi}_{1,h})|\!|\!|_h + |\!|\!|(w_2 - w_{2,h}, \boldsymbol{\phi}_2 - \boldsymbol{\phi}_{2,h})|\!|\!|_h).\end{aligned} \tag{4.22}$$

By the approximation properties of $S_{k,0}\times S_{k-1}^d$ with $k \ge 3$,

$$\inf_{(w_{1,h},\boldsymbol{\phi}_{1,h})\in S_{k,0}\times S_{k-1}^d} |\!|\!|(w_1 - w_{1,h}, \boldsymbol{\phi}_1 - \boldsymbol{\phi}_{1,h})|\!|\!|_h \le Ch^3(\|w_1\|_4 + \|\boldsymbol{\phi}_1\|_3 + \|A:\nabla\boldsymbol{\phi}_1\|_2).$$

Since $A:\nabla\boldsymbol{\phi}_1 = h^{-2}z$, then $\|A:\nabla\boldsymbol{\phi}_1\|_2 = h^{-2}\|z\|_2$. By the regularity assumptions (4.17), we have $\|\boldsymbol{\phi}_1\|_3 \le C\|w_1\|_4 \le Ch^{-2}\|z\|_2$. Combined with the regularity assumption $\|z\|_2 \le C\|e\|_0$,

$$\inf_{(w_{1,h},\boldsymbol{\phi}_{1,h})\in S_{k,0}\times S_{k-1}^d} |\!|\!|(w_1 - w_{1,h}, \boldsymbol{\phi}_1 - \boldsymbol{\phi}_{1,h})|\!|\!|_h \le Ch\|e\|_0. \tag{4.23}$$

For the w_2 and $\boldsymbol{\phi}_2$ terms, using approximation properties and the fact that $A:\nabla\boldsymbol{\phi}_2 = 0$,

$$\inf_{(w_{2,h},\boldsymbol{\phi}_{2,h})\in S_{k,0}\times S_{k-1}^d} |\!|\!|(w_2 - w_{2,h}, \boldsymbol{\phi}_2 - \boldsymbol{\phi}_{2,h})|\!|\!|_h \le Ch(\|w_2\|_2 + \|\boldsymbol{\phi}_2\|_1).$$

By the PDE form (4.20) and using the regularity assumption for the nondivergence PDE (4.16) for two times, we have

$$\|w_2\|_2 \le C\|A:\nabla(\nabla\cdot(Az))\|_0 \le C\|z\|_2 \le C\|e\|_0,$$

and by the fact $\boldsymbol{\phi}_2 = \nabla w_2 + \nabla\cdot(Az)$, the following is true:

$$\|\boldsymbol{\phi}_2\|_1 \le C(\|w_2\|_2 + \|\nabla\cdot(Az)\|_1) \le C(\|w_2\|_2 + \|z\|_2) \le C\|e\|_0.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/02/21 to 59.64.129.2. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Combining the results, we get

$$\inf_{(w_{2,h},\boldsymbol{\phi}_{2,h})\in S_{k,0}\times S^d_{k-1}} |||(w_2 - w_{2,h}, \boldsymbol{\phi}_2 - \boldsymbol{\phi}_{2,h})||| \le Ch\|e\|_0. \tag{4.24}$$

The result

$$\|e\|_0^2 \le Ch|||(e, E)|||_h\|e\|_0 \tag{4.25}$$

can be obtained from (4.22), (4.23), and (4.24), and the theorem is proved. ∎

Remark 4.9. The result of this theorem requires some high regularity and at least a degree three polynomial approximation for u. From the numerical experiments, we do find that this degree three requirement is necessary. The proof can be generalized to other h-weighted LSFEMs.

5. More a priori error estimates for the L^2-LSFEM.

In this section, we discuss the L^2-error estimation of the L^2-LSFEM with a standard H^2-regularity assumption. The proof is also based on a modification of that of Cai and Ku [16] but it is simpler than that of the W-LSFEM. With the L^2-error estimate available, we discuss the H^1-norm estimate with the same assumption.

THEOREM 5.1. *Assuming that the mesh is quasi-uniform with a mesh size h, the regularity assumptions* (4.16) *are true, and* $(u_h, \boldsymbol{\sigma}_h)$ *is the* $S_{1,0}\times S_1^d$ L^2*-LSFEM solution, we have the following* L^2*- and* H^1*-error estimates:*

$$\|u - u_h\|_0 \le Ch|||(u - u_h, \boldsymbol{\sigma} - \boldsymbol{\sigma}_h)|||_0, \tag{5.1}$$

$$\text{and} \quad \|\nabla(u - u_h)\|_0 \le C|||(u - u_h, \boldsymbol{\sigma} - \boldsymbol{\sigma}_h)|||_0 + Ch\|u\|_2. \tag{5.2}$$

Proof. We have the same error representation as in the W-LSFEM case. To match with the bilinear form of L^2-LSFEM, we also introduce two subauxiliary problems for $w_1 \in H^1_0(\Omega)$ and $w_2 \in H^1_0(\Omega)$:

$$\left\{\begin{array}{rcl} \boldsymbol{\phi}_1 - \nabla w_1 &=& 0, \\ A:\nabla\boldsymbol{\phi}_1 &=& z, \end{array}\right. \quad \text{and} \quad \left\{\begin{array}{rcl} \boldsymbol{\phi}_2 - \nabla w_2 &=& \nabla\cdot(Az), \\ A:\nabla\boldsymbol{\phi}_2 &=& 0. \end{array}\right. \tag{5.3}$$

Let $w = w_1 + w_2$ and $\boldsymbol{\phi} = \boldsymbol{\phi}_1 + \boldsymbol{\phi}_2$, and substitute (5.3) into the representation of $\|e\|_0^2$:

$$\|e\|_0^2 = a_0((e, E), (w_1, \boldsymbol{\phi}_1)) + a_0((e, E), (w_2, \boldsymbol{\phi}_2)).$$

Letting $(w_{i,h}, \boldsymbol{\phi}_{i,h}) \in S_{1,0}\times S_1^d$ for $i = 1$ and 2 and using the orthogonality of the error equation, we have

$$\begin{aligned} \|e\|_0^2 &= a_0((e, E), (w_1 - w_{1,h}, \boldsymbol{\phi}_1 - \boldsymbol{\phi}_{1,h})) + a_0((e, E), (w_2 - w_{2,h}, \boldsymbol{\phi}_2 - \boldsymbol{\phi}_{2,h})) \\ &\le |||(e, E)|||_0(|||(w_1 - w_{1,h}, \boldsymbol{\phi}_1 - \boldsymbol{\phi}_{1,h})|||_0 + |||(w_2 - w_{2,h}, \boldsymbol{\phi}_2 - \boldsymbol{\phi}_{2,h})|||_0). \end{aligned} \tag{5.4}$$

By the approximation properties of $S_{1,0}\times S_1^d$,

$$\inf_{(w_{1,h},\boldsymbol{\phi}_{1,h})\in S_{1,0}\times S_1^d} |||(w_1 - w_{1,h}, \boldsymbol{\phi}_1 - \boldsymbol{\phi}_{1,h})|||_0 \le Ch(\|w_1\|_2 + \|\boldsymbol{\phi}_1\|_1 + \|A:\nabla\boldsymbol{\phi}_1\|_1).$$

Since $A:\nabla\boldsymbol{\phi}_1 = z$, then $\|A:\nabla\boldsymbol{\phi}_1\|_1 \le \|A:\nabla\boldsymbol{\phi}_1\|_2 = \|z\|_2 \le C\|e\|_0$. By the regularity assumptions (4.16), then $\|\boldsymbol{\phi}_1\|_1 \le C\|w_1\|_2 \le C\|z\|_0 \le C\|z\|_2 \le C\|e\|_0$. Thus

$$\inf_{(w_{1,h},\boldsymbol{\phi}_{1,h})\in S_{1,0}\times S_1^d} |||(w_1 - w_{1,h}, \boldsymbol{\phi}_1 - \boldsymbol{\phi}_{1,h})|||_0 \le Ch\|e\|_0. \tag{5.5}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/02/21 to 59.64.129.2. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

For the w_2 and $\boldsymbol{\phi}_2$ terms, the proof of

$$\inf_{(w_{2,h},\boldsymbol{\phi}_{2,h})\in S_{1,0}\times S_1^d} |||(w_2 - w_{2,h}, \boldsymbol{\phi}_2 - \boldsymbol{\phi}_{2,h})|||_0 \le Ch\|e\|_0 \tag{5.6}$$

is identical to the estimate of the same term in the proof in the L^2-estimate of the W-LSFEM.

From (5.4), (5.5), and (5.6), we get

$$\|e\|_0^2 \le Ch|||(e, E)|||_0\|e\|_0, \tag{5.7}$$

thus the L^2-error estimate is proved.

To prove the H^1-norm error estimate, let v_h be the nodal interpolation of u in $S_{0,1}$, by the triangle inequality,

$$\|\nabla(u - u_h)\|_0 \le \|\nabla(u - v_h)\|_0 + \|\nabla(u_h - v_h)\|_0.$$

By the inverse estimates and the triangle inequality,

$$\|\nabla(u_h - v_h)\|_0 \le Ch^{-1}\|u_h - v_h\|_0 \le Ch^{-1}(\|u - u_h\|_0 + \|u - v_h\|_0).$$

Then, by the L^2-estimate of $u - u_h$ and the approximation property of v_h,

$$\|\nabla(u-u_h)\|_0 \le \|\nabla(u-v_h)\|_0 + Ch^{-1}(\|u-u_h\|_0 + \|u-v_h\|_0) \le C(|||(e, E)|||_0 + h\|u\|_2).$$

The theorem is proved. $\square$

6. Numerical experiments.

In this section, we present various numerical experiments. In the figures showing convergences, reference lines like N^{-1} (order h^2) are given, where N is the number of degrees of freedom. For example, for a uniform mesh with size h in two dimensions, $N^{-1} = O(h^2)$, while for an adaptive mesh, we use the same notation to denote the order that the errors converge as if the solution is smooth. In the examples, $r = \sqrt{x^2 + y^2}$.

6.1. Examples with a smooth solution.

We consider several different cases:

$$A_u = \begin{pmatrix} -\frac{5}{\ln(r)} + 15 & 1 \\ 1 & -\frac{1}{\ln(r)} + 3 \end{pmatrix}, \quad A_{deg} = \begin{pmatrix} |x|^{2/3} & -|x|^{1/3}|y|^{1/3} \\ -|x|^{1/3}|y|^{1/3} & |y|^{2/3} \end{pmatrix},$$

$$A_{dc} = \begin{pmatrix} 2 & s \\ s & 2 \end{pmatrix}, \text{ where } s = \begin{cases} 1, & \text{Quadrants I and III,} \\ -1, & \text{Quadrants II and IV.} \end{cases}$$

The coefficients of the matrices A_u, A_{dc}, and A_{deg} are uniformly continuous, discontinuous, and degenerate, respectively. The matrix A_{deg} is degenerate since $\det(A_{deg}) = 0$, thus the equation is not uniformly elliptic and does not satisfy the Cordes condition. The domain is chosen to be $\Omega = (-1/2, 1/2)^2$. The right-hand side f is chosen for the different coefficient A such that the solution is

$$u(x, y) = \sin(2\pi x)\sin(2\pi y)e^{x\cos(y)}. \tag{6.1}$$

The initial mesh contains four triangles by connecting two diagonals. Eight uniform refinements are performed to generate a series of numerical solutions.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

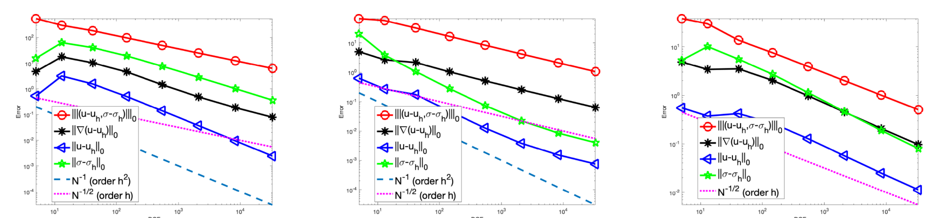
Downloaded 01/02/21 to 59.64.129.2. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

6.1.1. L^2-LSFEM. From Figure 1, we can see that for a problem with a smooth solution, the convergence order of the error in LSFEM norm $|||(u-u_h, \boldsymbol{\sigma}-\boldsymbol{\sigma}_h)|||_0$ is always one for all cases if the L^2-LSFEM with $S_{1,0} \times S_1^2$ approximation is used.

For the error of u in H^1-seminorm $\|\nabla(u-u_h)\|_0$, the order is the optimal one for all cases. For the L^2-norm error $\|u-u_h\|_0$, the uniformly continuous problem has an order 2 (optimal interpolation order), while the discontinuous case has an order slightly less than two and the degenerate case has an order slightly larger than one. These results are not covered by our theoretical analysis, but the numerical experiments suggest that the errors in L^2-norm are more sensitive to the smoothness of the coefficients, while the errors in H^1-norm is less sensitive.

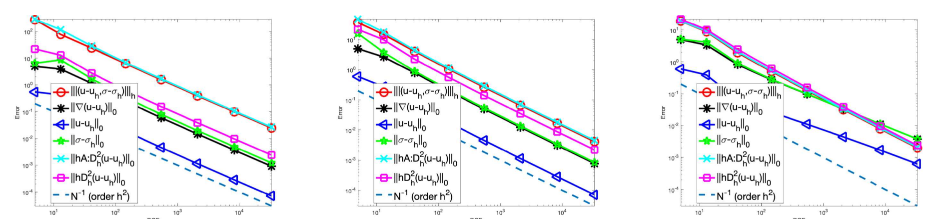
For the error of $\boldsymbol{\sigma}$ in L^2-norm $\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_0$, we do not have theoretical analysis, and numerical results do show that the convergence order depends on the problem. The observed order is between one, which is the approximation order of $\|\nabla(u-v_h)\|_0$, and two, which is the optimal interpolation order of $\|\boldsymbol{\sigma}-\boldsymbol{\tau}_h\|_0$.

6.1.2. W-LSFEM. In Figure 2, we show the numerical results for the smooth solution problems using $S_{k,0} \times S_{k-1}^2$ W-LSFEM with $k = 2$ and 3.

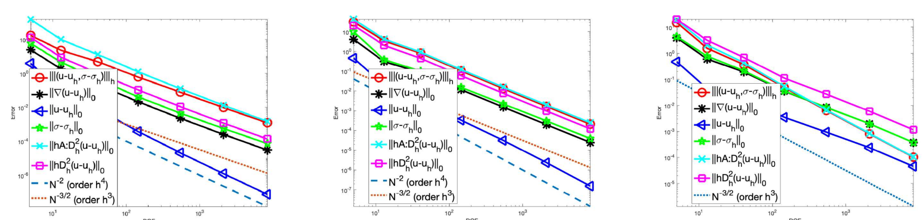


(a) Uniformly continuous A_u (b) Discontinuous A_{dc} (c) Degenerate A_{deg}

FIG. 1. *Convergence for L^2-LSFEM with $S_{1,0} \times S_1^2$ smooth solution* (6.1).



(a) Uniformly continuous A_u (b) Discontinuous A_{dc} (c) Degenerate A_{deg}



(d) Uniformly continuous A_u (e) Discontinuous A_{dc} (f) Degenerate A_{deg}

FIG. 2. *Convergence for $S_{k,0} \times S_{k-1}^2$ W-LSFEM with a smooth solution* (6.1)*:* (a)–(c) (k = 2)*;* (d)–(f) (k = 3).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/02/21 to 59.64.129.2. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

As discussed earlier, the error in LSFEM norms $\|\!|(u-u_h, \boldsymbol{\sigma}-\boldsymbol{\sigma}_h)|\!\|_h$ and $\|A:D_h^2(u-u_h)\|_0$ is of order k. For all nondegenerate cases, $\|\nabla(u-u_h)\|_0$, $\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_0$, $\|u-u_h\|_0$, and $\|hD_h^2(u-u_h)\|_0$ are of optimal interpolation orders. For the degenerate case with $k=2$, $\|\nabla(u-u_h)\|_0 = O(h^{1.5})$ and $\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_0 = O(h^{1.5})$ are less than optimal interpolation orders; $\|u-u_h\|_0 = O(h^{1.4})$ is even worse than the order of the H^1-seminorm error, but $\|hD_h^2(u-u_h)\|_0$ is of an optimal interpolation order 2. For $k=3$, $\|\nabla(u-u_h)\|_0$, $\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_0$, and $\|u-u_h\|_0$ are of order 2.4, which is less than the optimal interpolation order; $\|hD_h^2(u-u_h)\|_0$ is of order 2.3, which is also less than the optimal interpolation order.

In conclusion, for the degenerate case, $\|\nabla(u-u_h)\|_0$, $\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_0$, $\|u-u_h\|_0$, and $\|hD_h^2(u-u_h)\|_0$ are often worse than the optimal interpolation order, while the other cases are fine with a smooth solution.

6.2. A singular solution example with a degenerate matrix from [26]. Let the coefficient matrix $A = A_{deg}$ and $\Omega = (0,1)^2$. The right-hand side f is 0 and the exact solution is chosen to be

$$u = x^{4/3} - y^{4/3}.$$

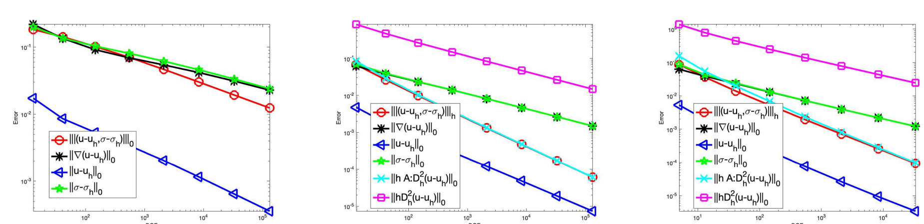
The gradient is singular along the x and y axes. This example is motivated by Aronson's example for the infinity-Laplace equation; see Test 3 of [26].

In Figure 3, numerical results with uniform refinements using $S_{1,0}\times S_1^2$ L^2-LSFEM and $S_{2,0}\times S_1^2$ and $S_{3,0}\times S_2^2$ W-LSFEM are shown.

For the $S_{1,0}\times S_1^2$ L^2-LSFEM, $\|\!|(u-u_h, \boldsymbol{\sigma}-\boldsymbol{\sigma}_h)|\!\|_0$ has an order 0.63, $\|\nabla(u-u_h)\|_0$ and $\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_0$ have an order 0.45, and $\|u-u_h\|_0$ has an order 0.85.

The W-LSFEMs have a convergence order 1.5 for $\|\!|(u-u_h, \boldsymbol{\sigma}-\boldsymbol{\sigma}_h)|\!\|_h$ and $\|hA:D_h^2(u-u_h)\|_0$, a convergence order 0.84 for $\|\nabla(u-u_h)\|_0$ and $\|\nabla(u-u_h)\|_0$, and a convergence order 1.4 for $\|u-u_h\|_0$. This again suggests that for the low regularity problem with uniformly mesh refinements, the higher-order method is unnecessary for the W-LSFEM. Not surprisingly, due to the degenerate nature of A, $\|hD_h^2(u-u_h)\|_0$ is only of order 0.83, which is much worse than order 1.5 of $\|hA:D_h^2(u-u_h)\|_0$.

In (a) and (b) of Figure 4, the numerical results with adaptive refinements using the adaptive $S_{1,0}\times S_1^2$ L^2-LSFEM and $S_{2,0}\times S_1^2$ W-LSFEM are shown. For the adaptive $S_{1,0}\times S_1^2$ L^2-LSFEM, all norms except $\|u-u_h\|_0$ converge at a rate less than one. This is partly due to the fact that the term $\|A:\nabla\boldsymbol{\sigma}_h - f\|_0$ requires a high regularity. The error $\|u-u_h\|_0$ converges at order 1, which is also less than the optimal interpolation order 2, which is also partly due to the degenerate nature of the matrix A. For the adaptive W-LSFEM, $\|\!|(u-u_h, \boldsymbol{\sigma}-\boldsymbol{\sigma}_h)|\!\|_h$ and $\|hA:D_h^2(u-u_h)\|_0$ converge at $O(N^{-1})$, which corresponds to $O(h^2)$ for a smooth solution in a uniform mesh.

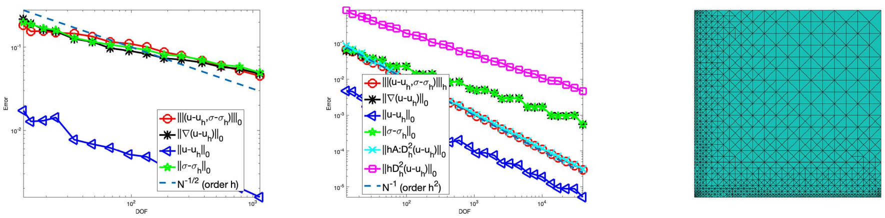


(a) L^2-LSFEM with $S_{1,0}\times S_1^2$ (b) W-LSFEM with $S_{2,0}\times S_1^2$ (c) W-LSFEM with $S_{3,0}\times S_2^2$

FIG. 3. *Convergence histories for a degenerate coefficient problem with* $u = x^{4/3} - y^{4/3}$.

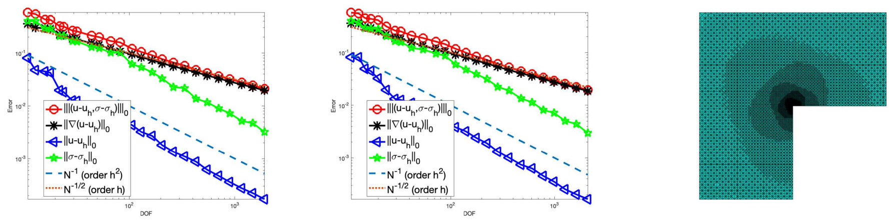
Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/02/21 to 59.64.129.2. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms



(a) L^2-LSFEM with $S_{1,0} \times S_1^2$ (b) W-LSFEM with $S_{2,0} \times S_1^2$ (c) Refined mesh generated by adaptive L^2-LSFEM

FIG. 4. *Convergence histories for a degenerate coefficient problem with $u = x^{4/3} - y^{4/3}$ with adaptive mesh refinements.*



(a) Uniformly continuous coefficient A_{u2} (b) Discontinuous coefficient A_{dc2} (c) A refined mesh when $||\nabla(u-u_h)||_0 \le 0.01$, $N = 7618$

FIG. 5. *Convergence histories for adaptive L^2-LSFEM with $S_{1,0} \times S_1^2$ for the L-shaped problem.*

All other norms converge with a lower than optimal rate due to the degenerate A. The error $\|u - u_h\|_0$ for $k = 3$ only has a rate about 2, which is also worse than the optimal 3. We also discover that the higher-order method behaves much better than the lower-order methods with adaptive methods.

In (c) of Figure 4, an adaptively refined mesh generated by the adaptive L^2-LSFEM is shown. It is clear from the graph that many refinements are along the x and y axes and around the origin.

6.3. An L-shaped domain problem. We choose the L-shaped domain $\Omega = (-1,1)^2 \backslash [0,1) \times (-1,0]$. The exact solution u in polar coordinates is

$$u(r,\theta) = r^{2/3} \sin(2\theta/3). \tag{6.2}$$

It is easy to check that $\Delta u = 0$. Consider the following cases:

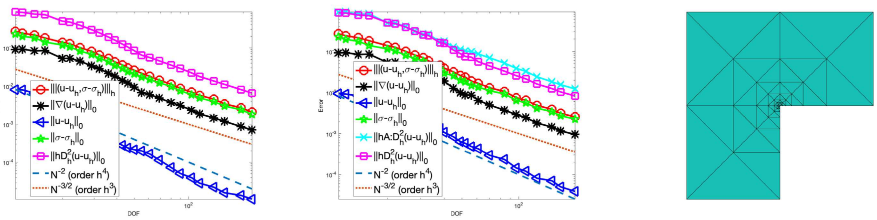
$$A_{u2} = \begin{pmatrix} 5 - 1/\ln(r) & r^2/2 \\ r^2/2 & 5 - 1/\ln(r) \end{pmatrix} \quad \text{and} \quad A_{dc2} = \begin{pmatrix} 2 & r^2 s \\ r^2 s & 2 \end{pmatrix},$$

where $s = \begin{cases} 1, & \text{Quadrants I and III,} \\ -1, & \text{Quadrants II and IV.} \end{cases}$ The coefficients of the matrices A_{u2} and A_{dc2} are uniformly continuous and discontinuous, respectively. We choose $a_{11} = a_{22}$ to use the fact $u_{xx} + u_{yy} = 0$, and $a_{12} = a_{21}$ with an r^2 factor to make sure that the right-hand side $f = -a_{12}u_{xy} - a_{21}u_{yx}$ belongs to L^2.

In Figures 5 and 6, numerical results for the L-shaped problems with adaptive refinements using the adaptive $S_{1,0} \times S_1^2$ L^2-LSFEM and $S_{3,0} \times S_2^2$ W-LSFEMs are shown. All convergences orders behave as if the solution and the matrix are smooth. In (c) of Figures 5 and 6, we show the mesh when $||\nabla(u - u_h)||_0 \le 0.01$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 01/02/21 to 59.64.129.2. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms



(a) Uniformly continuous coefficient A_{u2} (b) Discontinuous coefficient A_{dc2} (c) A refined mesh when $||\nabla(u-u_h)||_0 \leq 0.01$, $N = 50$

Fig. 6. *Convergence histories for adaptive W-LSFEM with $S_{3,0} \times S_2^2$ for the L-shaped problem.*

Acknowledgment. The authors sincerely thank the anonymous referee for the helpful comments and valuable suggestions, which considerably improved the exposition of this work.

REFERENCES

[1] J. H. Adler, T. A. Manteuffel, S. F. McCormick, J. W. Nolting, J. W. Ruge, and L. Tang, *Efficiency based adaptive local refinement for first-order system least-squares formulations*, SIAM J. Sci. Comput., 33 (2011), pp. 1–24.

[2] A. K. Aziz, R. B. Kellogg, and A. B. Stephens, *Least squares methods for elliptic systems*, Math. Comp., 44 (1985), pp. 53–70.

[3] M. Berndt, T. Manteuffel, and S. McCormick, *Local error estimates and adaptive refinement for first-order system least squares*, Electron. Trans. Numer. Anal., 6 (1997), pp. 35–43.

[4] M. Berndt, T. A. Manteuffel, S. F. McCormick, *Analysis of first-order system least-squares (FOSLS) for elliptic problems with discontinuous coefficients: Part* II, SIAM J. Numer. Anal., 43 (2005), pp. 409–436.

[5] M. Berndt, T. A. Manteuffel, S. F. McCormick, and G. Starke, *Analysis of first-order system least-squares (FOSLS) for elliptic problems with discontinuous coefficients: Part* II, SIAM J. Numer. Anal., 43 (2005), pp. 386–408.

[6] P. Bochev and M. Gunzburger, *On least-squares finite element methods for the Poisson equation and their connection to the Dirichlet and Kelvin principles*, SIAM J. Numer. Anal., 43 (2006), pp. 340–362.

[7] P. Bochev and M. Gunzburger, *Least-Squares Finite Element Methods*, Springer, New York, 2009.

[8] D. Boffi, F. Brezzi, and M. Fortin, *Mixed Finite Element Methods and Applications*, Springer Ser. Comput. Math. 44, Springer, New York, 2013.

[9] J. H. Bramble, R. D. Lazarov, and J. E. Pasciak, *A least-squares approach based on a discrete minus one inner product for first-order system*, Math. Comp., 66 (1997), pp. 935–955.

[10] S. C. Brenner and C. Carstensen, *Finite element methods*, in Encyclopedia of Computational Mechanics, 2nd ed., E. Stein, R. d. Borst, and T. J. R. Hughes, eds., John Wiley and Sons, New York, 2017.

[11] S. C. Brenner, T. Gudi, and L. Sung, *An a posteriori error estimator for a quadratic C0-interior penalty method for the biharmonic problem*, IMA J. Numer. Anal. 30 (2010), pp. 777–798.

[12] Z. Cai, C. He, and S. Zhang, *Improved ZZ a posteriori error estimators for diffusion problems: Conforming linear elements*, Comput. Methods Appl. Mech. Engrg., 313 (2017), pp. 433–449.

[13] Z. Cai, R. Falgout, and S. Zhang, *Div first-order system LL*(FOSLL*) for second-order elliptic partial differential equations*, SIAM J. Numer. Anal., 53 (2015), pp. 405–420.

[14] Z. Cai, R. Lazarov, T. A. Manteuffel, and S. F. McCormick, *First-order system least squares for second-order partial differential equations: Part* I, SIAM J. Numer. Anal., 31 (1994), pp. 1785–1799.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[15] Z. Cai, T. A. Manteuffel, and S. F. McCormick, *First-order system least squares for second-order partial differential equations: Part* II, SIAM J. Numer. Anal., 34 (1997), pp. 425–454.
[16] Z. Cai and J. Ku, *The L^2 norm error estimates for the div least-squares method*, SIAM J. Numer. Anal., 44 (2006), pp. 1721–1734.
[17] Z. Cai and S. Zhang, *Recovery-based error estimator for interface problems: Conforming linear elements*, SIAM J. Numer. Anal., 47 (2009), pp. 2132–2156.
[18] Z. Cai and S. Zhang, *Recovery-based error estimators for interface problems: Mixed and nonconforming finite elements*, SIAM J. Numer. Anal., 48 (2010), pp. 30–52.
[19] Z. Cai and S. Zhang, *Flux recovery and a posteriori error estimators: Conforming elements for scalar elliptic equations*, SIAM J. Numer. Anal., 48 (2010), pp. 578–602.
[20] C. Carstensen and E.-J. Park, *Convergence and optimality of adaptive least squares finite element methods*, SIAM J. Numer. Anal., 53 (2015), pp. 43-–62.
[21] H. Chen and W. Qiu, *A first-order system least squares method for the Helmholtz equation*, J. Comput. Appl. Math., 309 (2017), pp. 145–162.
[22] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, North-Holland, Amsterdam, 1978.
[23] F. Chiarenza, M. Frasca, and P. Longo, *$W^{p,2}$-solvability of the Dirichlet problem for non-divergence elliptic equations with VMO coefficients*, Trans. Amer. Math. Soc., 336 (1993), pp. 841–853.
[24] M. Dauge, *Elliptic Boundary Value Problems on Corner Domains*, Lecture Notes in Math. 1341, Springer-Verlag, Berlin, 1988.
[25] J. Douglas, Jr., T. Dupont, P. Percell, and R. Scott, *A family of $C1$ finite element with optimal approximation properties for various Galerkin methods for 2nd and 4th order problems*, RAIRO Anal. Numer., 13 (1979), pp. 227–255.
[26] X. Feng, L. Hennings, and M Neilan, *Finite element methods for second order linear elliptic partial differential equations in non-divergence form*, Math. Comp., 86 (2017), pp. 2025–2051.
[27] X. Feng, M. Neilan, and S. Stephan, *Interior penalty discontinuous Galerkin methods for second order linear non-divergence form elliptic PDEs*, J. Sci. Comput., 74 (2018), pp. 1651–1676.
[28] D. Gallistl, *Variational formulation and numerical analysis of linear elliptic equations in nondivergence form with Cordes coefficients*, SIAM J. Numer. Anal., 55 (2017), pp. 737–757.
[29] E. H. Georgoulis, P. Houston, and J. Virtanen, *An a posteriori error indicator for discontinuous Galerkin approximations of fourth-order elliptic problems*, IMA J. Numer. Anal., 31 (2011), pp. 281–298.
[30] D. Gilbarg and J. Serin, *On isolated singularities of solutions of second order elliptic differential equations*, J. Anal. Math., 4 (1955), pp. 309–340.
[31] V. Girault and P.-A. Raviart, *Finite Element Methods for Navier-Stokes Equations: Theory and Algorithms*, Springer Ser. Comput. Math. 5, Springer, New York, 1986.
[32] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Classics in Math., Springer, New York, 2001.
[33] J. Ku, *A remark on the coercivity for a first-order least-squares method*, Numer. Methods Partial Differential Equations, 23 (2007), pp. 1577–1581.
[34] Q. Liu and S. Zhang, *Adaptive least-squares finite element methods for linear transport equations based on an H(div) flux reformulation*, Comput. Methods Appl. Mech. Engrg., 366 (2020), 113041.
[35] Q. Liu and S. Zhang, *Adaptive flux-only least-squares finite element methods for linear transport equations*, J. Sci. Comput., 84 (2020).
[36] A. Maugeri, D. K. Palagachev, and L. G. Softova, *Elliptic and Parabolic Equations with Discontinuous Coefficients*, Math. Res. 109, Wiley-VCH Verlag, Weinheim, 2000.
[37] L. Mu and X. Ye, *A simple finite element method for non-divergence form elliptic equations*, Int. J. Numer. Anal. Model., 14 (2017), pp. 306–311.
[38] M. Neilan, *Convergence analysis of a finite element method for second order non-variational elliptic problems*, J. Numer. Math., 25 (2017), pp. 169–184.
[39] M. Neilan and M. Wu, *Discrete Miranda–Talenti estimates and applications to linear and nonlinear PDEs*, J. Comput. Appl. Math., 356 (2019), pp. 358–376.
[40] R. H. Nochetto and W. Zhang, *Discrete ABP estimate and convergence rates for linear elliptic equations in non-divergence form*, Found. Comput. Math., 18 (2018), pp. 537–593.
[41] I. Smears and E. Süli, *Discontinuous Galerkin finite element approximation of nondivergence form elliptic equations with Cordes coefficients*, SIAM J. Numer. Anal., 51 (2013), pp. 2088–2106.

Downloaded 01/02/21 to 59.64.129.2. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[42] G. Starke, *An adaptive least-squares mixed finite element method for elasto-plasticity*, SIAM J. Numer. Anal., 45 (2007), pp. 371–388.

[43] C. Wang and J. Wang, *A primal-dual weak Galerkin finite element method for second order elliptic equations in non-divergence form*, Math. Comp., 87 (2018), pp. 515–545.

Downloaded 01/02/21 to 59.64.129.2. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.