Electronic Transactions on Numerical Analysis.
Volume 41, pp. 262-288, 2014.

ISSN 1068-9613.

ETNA
Kent State University
http://etna.math.kent.edu

A UNIFIED ANALYSIS OF THREE FINITE ELEMENT METHODS FOR THE MONGE-AMPÈRE EQUATION*

MICHAEL NEILAN†

Abstract. It was recently shown in S. C. Brenner et al. [Math. Comp., 80 (2011), pp. 1979–1995] that Lagrange finite elements can be used to approximate classical solutions of the Monge-Ampère equation, a fully nonlinear second order PDE. We expand on these results and give a unified analysis for many finite element methods satisfying some mild structure conditions in two and three dimensions. After proving some abstract results, we lay out a blueprint to construct various finite element methods that inherit these conditions and show how C^1 finite element methods, C^0 finite element methods, and discontinuous Galerkin methods fit into the framework.

Key words. fully nonlinear PDEs, Monge-Ampère equation, finite element methods, discontinuous Galerkin methods

AMS subject classifications. 65N30, 65N12, 35J60.

1. Introduction. In this paper, we consider the finite element approximations of the fully nonlinear Monge-Ampère equation with exact solution u and Dirichlet boundary conditions [14, 15, 28, 32]:

$$0 = \mathcal{F}(u) := \begin{cases} f - \det(D^2 u), & \text{in } \Omega, \\ u - g, & \text{on } \partial\Omega. \end{cases} \tag{1.1}$$

Here, $\det(D^2u)$ denotes the determinant of the Hessian matrix D^2u, $\Omega \subset \mathbf{R}^n$ $(n = 2, 3)$ is a two or three dimensional, smooth, strictly convex domain, and f is a strictly positive function. The goal of this paper is to build and analyze various numerical methods to approximate classical convex solutions of the Monge-Ampère equation. In particular, we develop an abstract framework to construct discretizations, denoted by $\mathrm{F_h}(\cdot)$, of the nonlinear operator $\mathcal{F}(\cdot)$ in order to build finite element approximations of u.

This work is motivated by the recent results of Brenner et al. [10, 11] where the authors developed and analyzed Galerkin methods on smooth domains using the well-known and simple Lagrange finite elements. In order to build convergent methods, the authors constructed *consistent* numerical schemes such that the resulting discrete linearization is *stable*. As emphasized in [10], this simple idea leads to an intricate derivation of a not-so-obvious discretization.

In this paper, we expand on these results and give a unified analysis of many finite element methods in two and three dimensions which satisfy some general structure conditions. Furthermore, we lay out a simple blueprint to construct such schemes with these properties. The key idea, as in [10], is to build consistent and stable (in terms of the linearization) discretizations. Assuming a few more mild conditions on the numerical scheme, we use a relatively simple fixed-point argument to prove the existence of a solution provided the discretization parameter is sufficiently small. Furthermore, the error estimates reduce to error estimates of an auxiliary linear problem. The proof of these results are the main objectives of Sections 2 and 4.

Seemingly powerful, the abstract theory does not explicitly tell us how to build such stable and consistent schemes. As previously mentioned, the construction of the Lagrange finite

*Received June 18, 2012. Accepted June 20, 2014. Published online on September 30, 2014. Recommended by A. Klawonn. The work of the author was supported by the National Science Foundation under grant number DMS-1115421.
†Department of Mathematics, University of Pittsburgh, Pittsburgh, PA 15260 (neilan@pitt.edu).

ETNA
Kent State University
http://etna.math.kent.edu

element method in [10] involves a convoluted derivation, and it is not immediately clear how to extend this methodology to more general schemes. However, intuitively we expect that the discrete linearization should be consistent with the linearization of $\mathcal{F}(\cdot)$ (a justification of this assertion is given in Remark 2.5 below). With this in mind, to construct discretization schemes $\mathrm{F}_h(\cdot)$ that satisfy the conditions presented in Sections 2 and 4, we (i) consider the linearization of $\mathcal{F}(\cdot)$, (ii) discretize the linearization, (iii) 'de-linearize' to obtain discretizations for $\mathcal{F}(\cdot)$. Using this methodology, in Section 3 we are able to systematically develop various numerical methods including C^1 finite element methods, C^0 finite element methods, and discontinuous Galerkin (DG) methods. In the first case, we recover some special cases of the results of Böhmer [7, Theorem 8.7] and derive H^2 error estimates; see also [8]. In addition, we derive L^2 and H^1 error estimates which are new in the literature. In the case of C^0 finite element methods, we recover the results in [10], but again, in a more compact and systematic approach. The development and analysis of DG methods is completely new, and as far as the author is aware, this is the first time DG methods have been either developed or analyzed for fully nonlinear second order equations.

Due to their important role in many application areas such as differential geometry and optimal transport [15, 44, 45], there has been a growing interest in recent years towards developing numerical schemes for fully nonlinear second order equations. Here, we give a brief review in this direction. The first numerical method for the Monge-Ampère equation is due to Oliker and Prussner [39] who constructed a numerical scheme for computing an Aleksandrov measure induced by D^2u and obtained the solution of problem (1.1) as a by-product. More recently, Oberman [26, 38] constructed the first practical wide stencil difference schemes for nonlinear elliptic PDEs which can be written as functions of eigenvalues of the Hessian matrix, such as the Monge-Ampère equation. It was proved that the finite difference scheme satisfies the convergence criterion (consistency, stability, and monotonicity) established by Barles and Souganidis [4], although no rates of convergence were given. The clear advantage of this method is the ability to compute the convex solution even if u is not smooth. However, the method also suffers from low rates of convergence, even if the solution is smooth. Oberman et al. has also constructed a simple finite difference method in two dimensions in [5], although no convergence analysis was presented. Dean and Glowinski [19] presented an augmented Lagrange multiplier method and a least squares method for the Monge-Ampère equation by treating the nonlinear equations as a constraint and using a variational principle to select a particular solution. The convergence analysis of the scheme remains open. Böhmer [7] (see also [8]) introduced a projection method using C^1 finite element functions for classical solutions of general fully nonlinear second order elliptic PDEs (including the Monge-Ampère equation) and analyzes the methods using consistency, stability and linearization arguments. As far as the author is aware, this is the first convergence proof of fully nonlinear second order problems in a finite element setting. Feng and the author considered fourth order singular perturbations of (1.1) by adding a small multiple of the biharmonic operator to the PDE [24, 25]. Other relevant papers include [2, 20, 34, 35, 43, 46].

There are several advantages of the C^0 and DG methods compared to the methods mentioned above. As explained in [10], the advantages of these schemes include the relatively simplicity of the method and the ability to easily implement the method with several finite element software packages. This then allows one to use features such as fast solvers (multigrid and domain decomposition) and adaptivity off-the-shelf. Furthermore, in comparison to finite difference methods, both types of finite element methods can easily handle curved boundaries with little modification of the finite element code. Moreover, the advantages of the DG method for linear and mildly nonlinear problems carry over for the Monge-Ampère equation as well. These include the ease of implementation (especially in the context of

ETNA
Kent State University
http://etna.math.kent.edu

hp-adaptivity), the ability to easily handle inhomogeneous boundary conditions and curved boundaries, the ability to use highly nonuniform unstructured meshes with hanging nodes, and the ability to accurately capture shocks and discontinuities of the function and gradient. Admittedly there are not many advantages of using C^1 finite elements due to their computational complexity; we include this specific case to show that it fits within our framework and to simplify some of the analysis in [7].

There are several notions of solutions for the Monge-Ampère equation including classical, viscosity, and Alekandrov solutions [18, 32, 33, 42, 45]. Solutions in general are not smooth. However, if Ω, g, and f are sufficiently regular, Ω is strictly convex, and f is uniformly positive in Ω then all of these notions of solutions are equivalent, and the regularity of u follows from the results of Caffarelli, Nirenberg, and Spruck [16]. The methods developed in this paper are useful for several applications in differential geometry such as the prescribed Gauss curvature equation and the affine maximal surface equation [2, 28, 29, 30, 31, 44]. In regard to optimal transport, it is important to be able to compute weak solutions. Since the analysis below relies on the smoothness of u, it is not immediately obvious how to extend the results to the case of nonclassical solutions. One way to overcome this issue is to use the methods described below in conjunction with the vanishing moment methodology [24, 25]. This simple procedure involves computing a fourth order perturbation of (1.1) by adding a small multiple of the biharmonic operator to the fully nonlinear PDE. Numerical experiments in [10, 11, 24, 25] indicate that this is a powerful tool not only to compute weak solutions, but it also provides an avenue to obtain good initial guesses to start Newton's method. Unlike the discretizations in [24, 25], the methods in this paper are designed to be stable without regularization. As such, we expect our methods in conjunction with the vanishing moment method will be more robust with respect to the perturbation parameter. Finally, we mention that there are several discretization methods for the biharmonic problem using C^1, C^0, and DG spaces; see, e.g., [3, 13, 17, 22, 27, 36]. Therefore, the formulation of our methods with the vanishing moment method is easily obtained by adding a small multiple of these biharmonic discretizations to the discrete nonlinear operator $\mathrm{F}_h(\cdot)$. However, the convergence analysis of the regularized discretization is beyond the scope of this article.

The results below show that there exists a solution to the numerical schemes provided that the discretization parameter is sufficiently small, namely, $h \le h_0$ for some $h_0 \in (0,1)$. The exact value of h_0 is not addressed in this paper and in general is not known *a priori* as it depends on the the exact solution u. This issue as well as the effect of numerical errors of the schemes will be studied in future work.

The rest of the paper is organized as follows. Sections 2 and 3 are devoted to the development and analysis of finite element methods in two dimensions. In Section 2 we state and prove some abstract results, showing that there exists a solution to the discrete problem if some conditions on the scheme hold. We end this section by deriving L^2 estimates using a standard duality argument. After setting some notation, in Section 3 we apply the abstract framework to three examples, namely, C^1 finite element methods, C^0 finite element methods, and DG methods. In Section 4, we expand on the results of Section 2 by deriving some abstract convergence results in the three dimensional case. Finally, we end the article with some concluding remarks and discuss possible extensions.

2. Abstract results in two dimensions. Throughout the paper, we use $H^r(\Omega)$ $(r \ge 0)$ to denote the set of all $L^2(\Omega)$ functions whose distributional derivatives up to order r are in $L^2(\Omega)$, and $H^r_0(\Omega)$ to denote the set of functions whose traces vanish up to order $r-1$ at $\partial\Omega$. For a normed linear space Y, we denote by Y' its dual and $\langle \cdot, \cdot \rangle$ the pairing between Y' and Y.

Let $(X_h, \|\cdot\|_{X_h})$ be a finite dimensional space such that the inclusion $X_h \subset L^2(\Omega)$

ETNA
Kent State University
http://etna.math.kent.edu

holds. We consider the following discrete version of (1.1): find $u_h \in X_h$ such that

$$\langle \mathrm{F_h}(u_h), v_h \rangle = 0, \qquad \forall v_h \in X_h, \tag{2.1}$$

where $\mathrm{F_h} : X_h \to X_h'$ is a smooth operator.

We make the following assumptions:

(a.1) (a) There exists an auxiliary normed linear space $(Y, \|\cdot\|_Y)$ with $X_h \subset Y$ and $u \in Y$ such that $\|\cdot\|_{X_h}$ is well defined on Y.

(b) The operator $\mathrm{F_h}(\cdot)$ can be extended to a smooth operator $\mathrm{F} : Y \to X_h'$ with $\langle \mathrm{F}(w_h), v_h \rangle = \langle \mathrm{F_h}(w_h), v_h \rangle, \forall w_h, v_h \in X_h$.

(c) There exists a constant $\alpha_h > 0$, which may depend on h, such that the following inverse estimate holds:

$$\|v_h\|_Y \le \alpha_h \|v_h\|_{X_h}, \qquad \forall v_h \in X_h. \tag{2.2}$$

(a.2) (a) The nonlinear operator $\mathrm{F}(\cdot)$ is consistent with $\mathcal{F}(\cdot)$ in the sense that $\mathrm{F}(u) = 0$.

(b) $\mathrm{F}(\cdot)$ can be decomposed as

$$\mathrm{F}(\cdot) = \mathrm{F}^{(2)}(\cdot) + \mathrm{F}^{(1)}(\cdot) + \mathrm{F}^{(0)}, \tag{2.3}$$

where $\mathrm{F}^{(2)}(\cdot)$ is quadratic (i.e., $\mathrm{F}^{(2)}(tw) = t^2 \mathrm{F}^{(2)}(w)$ for all $w \in Y$ and $t \in \mathbb{R}$), $\mathrm{F}^{(1)}(\cdot)$ is linear, and $\mathrm{F}^{(0)}$ is constant in their arguments.

(a.3) Define the linear operator

$$\mathrm{L}(w) := \mathrm{DF}[u](w) := \lim_{t\to 0} \frac{\mathrm{F}(u+tw) - \mathrm{F}(u)}{t} = \mathrm{F}^{(1)}(w) + \mathrm{DF}^{(2)}[u](w), \tag{2.4}$$

where

$$\mathrm{DF}^{(2)}[u](w) := \lim_{t\to 0} \frac{\mathrm{F}^{(2)}(u+tw) - \mathrm{F}^{(2)}(u)}{t}. \tag{2.5}$$

Note that, by definition of L and F, we have $\mathrm{L} : Y \mapsto X_h'$. Then there are constants $\beta_h, C_{cont} > 0$ such that

$$\beta_h \|v_h\|_{X_h}^2 \le \langle \mathrm{L_h}(v_h), v_h \rangle, \qquad \forall v_h \in X_h, \tag{2.6}$$

$$\langle \mathrm{L}(w), v_h \rangle \le C_{cont} \|w\|_{X_h} \|v_h\|_{X_h}, \qquad \forall w \in Y,\ v_h \in X_h, \tag{2.7}$$

where $\mathrm{L_h} : X_h \to X_h'$ denotes the restriction of L to X_h, that is, $\langle \mathrm{L_h}(w_h), v_h \rangle = \langle \mathrm{L}(w_h), v_h \rangle$ for all $v_h, w_h \in X_h$.

(a.4) There exists a constant $\gamma_h > 0$ such that for all $v, w \in Y$

$$\left\| \mathrm{DF}^{(2)}[v](w) \right\|_{X_h'} \le \gamma_h \|v\|_Y \|w\|_Y,$$

where

$$\left\| \mathrm{DF}^{(2)}[v](w) \right\|_{X_h'} := \sup_{y_h \in X_h} \frac{\langle \mathrm{DF}^{(2)}[v](w), y_h \rangle}{\|y_h\|_{X_h}}.$$

THEOREM 2.1. *Suppose that assumptions (a.1)–(a.4) are satisfied. Let* $u_{ch} = \mathrm{L_h^{-1}L}(u) \in X_h$*, i.e.,*

$$\langle \mathrm{L_h}(u_{ch}), v_h \rangle = \langle \mathrm{L}(u), v_h \rangle, \qquad \forall v_h \in X_h. \tag{2.8}$$

Assume further that

$$\|u - u_{ch}\|_Y \le \tau_0 \frac{\beta_h}{2\alpha_h \gamma_h} \qquad \text{for some } \tau_0 \in (0,1). \tag{2.9}$$

Then there exists a locally unique solution $u_h \in X_h$ *to* (2.1)*. Moreover, there holds*

$$\begin{aligned} \|u - u_h\|_{X_h} &\le \|u - u_{ch}\|_{X_h} + \frac{1}{\alpha_h}\|u - u_{ch}\|_Y, \\ \|u - u_h\|_Y &\le 2\|u - u_{ch}\|_Y. \end{aligned} \tag{2.10}$$

We prove Theorem 2.1 using the Banach fixed-point theorem as our main tool. The essential ingredients of the proof of Theorem 2.1 is to construct a mapping such that (i) the mapping is a contraction in a subset (in our case, a ball $\mathbb{B}_\rho$ with radius ρ) of X_h; (ii) the mapping maps this ball into itself. Both of these results are derived by the following lemma.

LEMMA 2.2. *Suppose that conditions (a.1)–(a.4) hold. Define the mapping* $\mathcal{M} : Y \to X_h$ *by*

$$\mathcal{M}w = \mathrm{L_h^{-1}}\Big(\mathrm{L}(w) - \mathrm{F}(w)\Big). \tag{2.11}$$

Then for any $v, w \in Y$*, we have*

$$\big\|\mathcal{M}w - \mathcal{M}v\big\|_{X_h} \le \frac{\gamma_h}{2\beta_h}\big(\|u - w\|_Y + \|u - v\|_Y\big)\|w - v\|_Y. \tag{2.12}$$

Proof. By the decomposition (2.3), we have $\mathrm{F}(w) - \mathrm{F}(v) = \mathrm{F}^{(2)}(w) - \mathrm{F}^{(2)}(v) + \mathrm{F}^{(1)}(w - v)$, where we have used the property that $\mathrm{F}^{(1)}$ is a linear operator. Moreover, by (2.4) there holds $\mathrm{L}(w - v) = \mathrm{F}^{(1)}(w - v) + \mathrm{DF}^{(2)}[u](w - v)$. Consequently, we have

$$\begin{aligned} \mathrm{L}(w - v) - \big(\mathrm{F}(w) - \mathrm{F}(v)\big) &= \mathrm{DF}^{(2)}[u](w - v) - \big(\mathrm{F}^{(2)}(w) - \mathrm{F}^{(2)}(v)\big) \\ &= \mathrm{DF}^{(2)}[u](w - v) - \int_0^1 \mathrm{DF}^{(2)}[tw + (1 - t)v](w - v)\, dt. \end{aligned}$$

Since $\mathrm{F}^{(2)}(\cdot)$ is smooth and quadratic, the mapping $(w, v) \to \mathrm{DF}^{(2)}[w](v)$ is bilinear. It then follows that

$$\mathrm{L}(w - v) - \big(\mathrm{F}(w) - \mathrm{F}(v)\big) = \mathrm{DF}^{(2)}\left[u - \frac{1}{2}(w + v)\right](w - v). \tag{2.13}$$

Therefore, by the definition of $\mathcal{M}$ (2.11) along with the identity (2.13), we arrive at

$$\begin{aligned} \mathcal{M}w - \mathcal{M}v &= \mathrm{L_h^{-1}}\Big(\mathrm{L}(w - v) - \big(\mathrm{F}(w) - \mathrm{F}(v)\big)\Big) \\ &= \mathrm{L_h^{-1}}\left(\mathrm{DF}^{(2)}\left[u - \frac{1}{2}(w + v)\right](w - v)\right). \end{aligned}$$

ETNA
Kent State University
http://etna.math.kent.edu

By the stability estimate in (a.3), we easily obtain $\|\mathcal{M}w - \mathcal{M}v\|_{X_h} \le \beta_h^{-1}\|\mathrm{DF}^{(2)}[u - \frac{1}{2}(w+v)](w-v)\|_{X_h'}$, and therefore by (a.4), we have

$$\|\mathcal{M}w - \mathcal{M}v\|_{X_h} \le \frac{\gamma_h}{2\beta_h}\big(\|u-w\|_Y + \|u-v\|_Y\big)\|w-v\|_Y. \qquad \square$$

Proof of Theorem 2.1. Define the closed discrete ball with center u_{ch} as

$$\mathbb{B}_\rho(u_{ch}) = \{v_h \in X_h;\ \|u_{ch} - v_h\|_{X_h} \le \rho\}, \tag{2.14}$$

and let $\mathcal{M}_h : X_h \to X_h$ be the restriction of $\mathcal{M}$ to X_h. The proof proceeds by showing that $\mathcal{M}_h$ has a fixed point in a $\mathbb{B}_{\rho_0}(u_{ch})$ with $\rho_0 := \frac{1}{\alpha_h}\|u - u_{ch}\|_Y$. By the definition of $\mathcal{M}_h$, (2.11), we clearly see that this fixed point is a solution to (2.1).

First, by (2.12), (a.1c), (2.14), and the definition of ρ_0, we have for all $v_h, w_h \in \mathbb{B}_{\rho_0}(u_{ch})$,

$$\begin{aligned}\|\mathcal{M}_h w_h - \mathcal{M}_h v_h\|_{X_h} &\le \frac{\gamma_h}{2\beta_h}\big(\|u-w_h\|_Y + \|u-v_h\|_Y\big)\|w_h - v_h\|_Y \\ &\le \frac{\alpha_h\gamma_h}{\beta_h}\big(\|u-u_{ch}\|_Y + \alpha_h\rho_0\big)\|w_h - v_h\|_{X_h} \\ &= \frac{2\alpha_h\gamma_h}{\beta_h}\|u-u_{ch}\|_Y\|w_h - v_h\|_{X_h}.\end{aligned}$$

Hence, by (2.9) we obtain

$$\|\mathcal{M}_h w_h - \mathcal{M}_h v_h\|_{X_h} \le \tau_0\|w_h - v_h\|_{X_h} \qquad \text{for some } \tau_0 \in (0,1). \tag{2.15}$$

Next, it is clear from (2.8), (2.11), and the consistency of $\mathrm{F}(\cdot)$ that $u_{ch} = \mathcal{M}u$. Therefore by (2.12), (a.1c), the definition of ρ_0, and (2.9), we have for any $w_h \in \mathbb{B}_{\rho_0}(u_{ch})$,

$$\begin{aligned}\|u_{ch} - \mathcal{M}_h w_h\|_{X_h} &= \|\mathcal{M}u - \mathcal{M}w_h\|_{X_h} \le \frac{\gamma_h}{2\beta_h}\|u-w_h\|_Y^2 \\ &\le \frac{\gamma_h}{\beta_h}\big(\|u-u_{ch}\|_Y^2 + \|u_{ch}-w_h\|_Y^2\big) \\ &\le \frac{\gamma_h}{\beta_h}\big(\|u-u_{ch}\|_Y^2 + \alpha_h^2\rho_0^2\big) = \frac{2\alpha_h\gamma_h}{\beta_h}\|u-u_{ch}\|_Y\rho_0 \le \rho_0.\end{aligned} \tag{2.16}$$

From (2.15) and (2.16) it then follows that $\mathcal{M}_h$ has a unique fixed point u_h in the ball $\mathbb{B}_{\rho_0}(u_{ch})$ which is a solution to (2.1). Also, by the triangle inequality we have

$$\|u-u_h\|_{X_h} \le \|u-u_{ch}\|_{X_h} + \rho_0 = \|u-u_{ch}\|_{X_h} + \frac{1}{\alpha_h}\|u-u_{ch}\|_Y.$$

Finally, to prove (2.10), we use the triangle inequality once again, the inverse estimate (2.2), and the definition of ρ_0 to get $\|u-u_h\|_Y \le \|u-u_{ch}\|_Y + \alpha_h\rho_0 = 2\|u-u_{ch}\|_Y$. $\square$

COROLLARY 2.3. *Suppose that assumptions (a.1)–(a.4) hold and define* $C_h := 1 + C_{cont}/\beta_h$. *Then, there holds*

$$\begin{aligned}\|u-u_h\|_{X_h} &\le \inf_{v_h\in X_h}\Big(2C_h\|u-v_h\|_{X_h} + \frac{1}{\alpha_h}\|u-v_h\|_Y\Big), \\ \|u-u_h\|_Y &\le 2\inf_{v_h\in X_h}\big(\|u-v_h\|_Y + \alpha_h C_h\|u-v_h\|_{X_h}\big),\end{aligned}$$

provided $\inf_{v_h\in X_h}(\|u-v_h\|_Y + \alpha_h C_h\|u-v_h\|_{X_h}) \le \tau_0\beta_h/(2\alpha_h\gamma_h)$ *for some* $\tau_0 \in (0,1)$.

ETNA
Kent State University
http://etna.math.kent.edu

Proof. In light of (2.6), (2.8), and (2.7), we find

$$\begin{aligned}(2.17)\qquad \beta_h\|u_{ch}-v_h\|_{X_h}^2 &\le \big\langle \mathrm{L_h}(u_{ch}-v_h), u_{ch}-v_h\big\rangle = \big\langle \mathrm{L}(u-v_h), u_{ch}-v_h\big\rangle\\ &\le C_{cont}\|u_{ch}-v_h\|_{X_h}\|u_{ch}-v_h\|_{X_h},\end{aligned}$$

for any $v_h \in X_h$, and therefore $\|u-u_{ch}\|_{X_h} \le C_h \inf_{v_h\in X_h}\|u-v_h\|_{X_h}$. Furthermore, by (2.17) and (2.2) we have

$$(2.18)\qquad \|u-u_{ch}\|_Y \le \inf_{v_h\in X_h}\big(\|u-v_h\|_Y + \alpha_h C_h\|u-v_h\|_{X_h}\big).$$

Hence, by Theorem 2.1 we obtain

$$\begin{aligned}\|u-u_h\|_{X_h} &\le \inf_{v_h\in X_h}\big(2C_h\|u-v_h\|_{X_h} + \frac{1}{\alpha_h}\|u-v_h\|_Y\big),\\ \|u-u_h\|_Y &\le 2\inf_{v_h\in X_h}\big(\|u-v_h\|_Y + \alpha_h C_h\|u-v_h\|_{X_h}\big),\end{aligned}$$

provided that the right-hand side of (2.18) is smaller than $\tau_0\beta_h/(2\alpha_h\gamma_h)$. ∎

REMARK 2.4. In all the examples considered below, $\|\cdot\|_{X_h}$ is a discrete H^1-type norm, $\|\cdot\|_Y$ is a discrete H^2-type norm, $\alpha_h = O(h^{-1})$, $\beta_h = O(1)$, and $C_h = O(1)$.

REMARK 2.5 (Some practical considerations). Theorem 2.1 states that if the discrete linearization is stable and if some other mild conditions hold, then there exists a solution to (2.1) close to u. However, it does not indicate how to construct discretizations with stable linearizations. One natural way to do this is to construct a scheme such that the operator $\mathrm{L_h}(\cdot)$ is consistent with $\mathcal{L}(\cdot)$, where

$$(2.19)\quad \mathcal{L}(w) := \lim_{t\to 0}\frac{\mathcal{F}(u+tw)-\mathcal{F}(u)}{t} = -\operatorname{cof}(D^2u) : D^2w = -\nabla\cdot(\operatorname{cof}(D^2u)\nabla w).$$

Here, $\operatorname{cof}(D^2u)$ denotes the cofactor matrix of D^2u,

$$\operatorname{cof}(D^2u) : D^2w = \sum_{i,j=1}^n (\operatorname{cof}(D^2u))_{i,j}(D^2w)_{i,j},$$

and we have used the divergence-free row property of cofactor matrices (cf. Lemma 3.1 below) to obtain the last equality. In other words, it is desirable that the diagram in Fig-

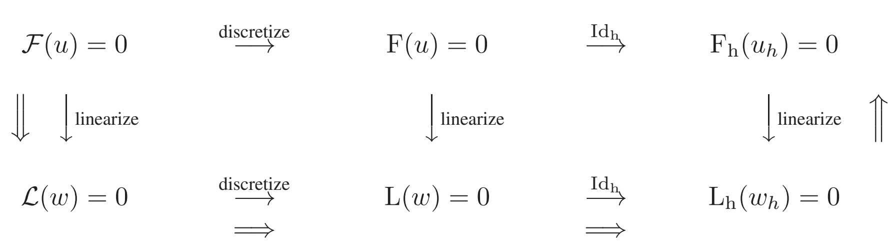

FIGURE 2.1. *An abstract commuting diagram. Here,* $\mathrm{Id_h}$ *denotes the restriction of an operator to the finite element space* X_h*. The path we take to derive convergent finite element methods is indicated by the double-lined arrows.*

ure 2.1 commutes. Since u is a classical convex solution to the Monge-Ampère equation, the matrix $\operatorname{cof}(D^2u)$ is positive definite, and so the operator $\mathcal{L}(\cdot)$ is a uniformly elliptic. As

ETNA
Kent State University
http://etna.math.kent.edu

discretization schemes for elliptic second order linear PDEs are well understood and developed (namely, many *stable* discretizations of $\mathcal{L}(\cdot)$ exist), it is better to build the nonlinear discretization $\mathrm{F_h}(\cdot)$ based on the discrete linear problem $\mathrm{L_h}(\cdot)$. In fact we use this property when deriving L^2 estimates below; cf. Assumption (a.5c). This is the approach we take in Section 3 when constructing finite element schemes.

2.1. L^2 estimates.

We end this section by deriving some L^2 error estimates. This is achieved by using duality arguments in conjunction with the following additional set of assumptions.

(a.5) (a) The operator $\mathrm{L}(\cdot)$ is symmetric and can be naturally extended such that $\mathrm{L} : H^2(\Omega) \to Y'$.

(b) The norm $\|\cdot\|_{X_h}$ is well-defined on $H^2(\Omega)$.

(c) The operator $\mathrm{L}(\cdot)$ is consistent with $\mathcal{L}(\cdot)$ (defined by (2.19)) in the sense that

$$\big\langle \mathrm{L}(v), w\big\rangle = \big\langle \mathcal{L}(v), w\big\rangle, \qquad \forall v \in H^2(\Omega) \cap H^1_0(\Omega),\ w \in Y.$$

(d) $\mathrm{L}(\cdot)$ is bounded in the sense that there exists an $M > 0$ such that

$$\big\langle \mathrm{L}(v), w\big\rangle \le M\|v\|_{X_h}\|w\|_{X_h}, \qquad \forall v, w \in X_h + H^2(\Omega).$$

(e) u is strictly convex and $u \in W^{3,\infty}(\Omega)$.

THEOREM 2.6. *In addition to the assumptions of Theorem 2.1, suppose that condition (a.5) is satisfied. Then there holds*

$$\|u-u_h\|_{L^2(\Omega)} \le \sup_{\varphi\in H^2(\Omega)} \inf_{\varphi_h\in X_h} \frac{C_E}{\|\varphi\|_{H^2(\Omega)}}\Big(M\|u-u_h\|_{X_h}\|\varphi-\varphi_h\|_{X_h} + \frac{\gamma_h}{2}\|u-u_h\|_Y^2\|\varphi_h\|_{X_h}\Big), \tag{2.20}$$

where C_E is defined by (2.22) below.

Proof. Let $\psi \in H^1_0(\Omega)$ be the solution to

$$\begin{aligned} \mathcal{L}(\psi) &= u - u_h, && \text{in } \Omega,\\ \psi &= 0, && \text{on } \partial\Omega. \end{aligned} \tag{2.21}$$

Since u is strictly convex in Ω and $u \in W^{3,\infty}(\Omega)$, by elliptic regularity there holds $\psi \in H^2(\Omega)$ and

$$\|\psi\|_{H^2(\Omega)} \le C_E\|u-u_h\|_{L^2(\Omega)} \tag{2.22}$$

for some $C_E > 0$. It follows from (a.5a) and (2.21) that for any $\psi_h \in X_h$

$$\begin{aligned} \|u-u_h\|^2_{L^2(\Omega)} &= \big\langle \mathcal{L}(\psi), u-u_h\big\rangle = \big\langle \mathrm{L}(\psi), u-u_h\big\rangle\\ &= \big\langle \mathrm{L}(u-u_h), \psi-\psi_h\big\rangle + \big\langle \mathrm{L}(u-u_h), \psi_h\big\rangle. \end{aligned} \tag{2.23}$$

Bounding the first term in (2.23), we use (a.5d) to obtain

$$\big\langle \mathrm{L}(u-u_h), \psi-\psi_h\big\rangle \le M\|u-u_h\|_{X_h}\|\psi-\psi_h\|_{X_h}.$$

ETNA
Kent State University
http://etna.math.kent.edu

To bound the second term in (2.23), we first note by (2.3) and (2.4) that

$$\begin{aligned}\big\langle \mathrm{L}(u_h-u),\psi_h\big\rangle &= \Big\langle \mathrm{L}(u_h-u)-\big(\mathrm{F}(u_h)-\mathrm{F}(u)\big),\psi_h\Big\rangle\\ &= \Big\langle \mathrm{DF}^{(2)}[u](u_h-u)-\big(\mathrm{F}^{(2)}(u_h)-\mathrm{F}^{(2)}(u)\big),\psi_h\Big\rangle\\ &= \Big\langle \mathrm{F}^{(2)}[u](u_h-u)-\int_0^1 \mathrm{F}^{(2)}[tu_h+(1-t)u](u_h-u)\,dt,\psi_h\Big\rangle\\ &= \frac{1}{2}\Big\langle \mathrm{DF}^{(2)}[u-u_h](u_h-u),\psi_h\Big\rangle,\end{aligned}$$

where we have again used the fact that the mapping $(w,v)\to \mathrm{F}^{(2)}[w](v)$ is bilinear. Combining this last identity with assumption (a.4), we obtain the estimate

$$\big\langle \mathrm{L}(u_h-u),\psi_h\big\rangle \le \frac{\gamma_h}{2}\|u-u_h\|_Y^2\|\psi_h\|_{X_h}. \tag{2.24}$$

Finally, combining (2.23)–(2.24), we have

$$\|u-u_h\|_{L^2(\Omega)}^2 \le \Big(M\|u-u_h\|_{X_h}\|\psi-\psi_h\|_{X_h}+\frac{\gamma_h}{2}\|u-u_h\|_Y^2\|\psi_h\|_{X_h}\Big).$$

Dividing both terms by $\|u-u_h\|_{L^2(\Omega)}$ and using the elliptic regularity estimate (2.22), we obtain (2.20). □

3. Some specific examples in two dimensions. In this section we apply the abstract framework set in the previous section to some concrete examples, namely, C^1 finite element, C^0 finite element, and discontinuous Galerkin methods. In the first case, we recover some of the error estimates obtained by Böhmer in [7], but also obtain L^2 and H^1 error estimates. In the second case, we recover the same results recently shown in [10], but in a more compact form. The method and error analysis of DG methods for the Monge-Ampère equation is completely new. Before proceeding, we first give some notation and standard lemmas that will be used throughout the rest of the paper.

3.1. Notation and some preliminary lemmas. Let $\mathcal{T}_h$ be a quasi-uniform, simplicial, and conforming triangulation [6, 12, 17] of the domain Ω where each triangle on the boundary has at most one curved side. We denote by $\mathcal{E}_h^i$ the set of interior edges, $\mathcal{E}_h^b$ the set of boundary edges, and $\mathcal{E}_h=\mathcal{E}_h^i\cup\mathcal{E}_h^b$ the set of all edges in $\mathcal{T}_h$. We set $h_T=\mathrm{diam}(T)$ for all $T\in\mathcal{T}_h$, $h_e=\mathrm{diam}(e)$ for all $e\in\mathcal{E}_h$, and note that by the assumption of the quasiuniformity of the mesh, $h_T\approx h_e\approx h:=\max_{T\in\mathcal{T}_h}h_T$.

Define the broken Sobolev space, norm, and semi-norm associated with the mesh as

$$H^r(\mathcal{T}_h):=\prod_{T\in\mathcal{T}_h}H^r(T),\quad \|v\|_{H^r(\mathcal{T}_h)}^2:=\sum_{T\in\mathcal{T}_h}\|v\|_{H^r(T)}^2,\quad |v|_{H^r(\mathcal{T}_h)}^2:=\sum_{T\in\mathcal{T}_h}|v|_{H^r(T)}^2.$$

We define the jump of a vector function $\mathbf{w}$ on an interior edge $e=\partial T^+\cap\partial T^-$ as follows:

$$[\![\mathbf{w}]\!]\big|_e=\mathbf{w}^+\cdot\mathbf{n}_+\big|_e+\mathbf{w}^-\cdot\mathbf{n}_-\big|_e\in\mathbf{R},$$

where $\mathbf{w}^\pm=\mathbf{w}\big|_{T^\pm}$ and $\mathbf{n}_\pm$ is the outward unit normal of $T^\pm$. On a boundary edge $e\in\mathcal{E}_h^b$, we define $[\![\mathbf{w}]\!]\big|_e=\mathbf{w}\cdot\mathbf{n}\big|_e\in\mathbf{R}$. The jump of a scalar function w is a vector and is defined as

$$\begin{aligned}&[\![w]\!]\big|_e=w^+\mathbf{n}_+\big|_e+w^-\mathbf{n}_-\big|_e, && e=\partial T^+\cap T^-\in\mathcal{E}_h^i,\\ &[\![w]\!]\big|_e=w\mathbf{n}\big|_e, && e=\partial T\cap\partial\Omega\in\mathcal{E}_h^b.\end{aligned}$$

ETNA
Kent State University
http://etna.math.kent.edu

For a matrix $\underline{\mathbf{w}} \in \mathbf{R}^{2\times 2}$, we define the average of $\underline{\mathbf{w}}$ on $e = \partial T^+ \cap \partial T^-$ by

$$\{\!\{\underline{\mathbf{w}}\}\!\}\big|_e = \frac{1}{2}\left(\underline{\mathbf{w}}^+\big|_e + \underline{\mathbf{w}}^-\big|_e\right) \in \mathbf{R}^{2\times 2},$$

and on a boundary edge $e \in \mathcal{E}_h^b$ we take $\{\!\{\underline{\mathbf{w}}\}\!\}\big|_e = \underline{\mathbf{w}}\big|_e \in \mathbf{R}^{2\times 2}$. Similarly, for a vector $\mathbf{w} \in \mathbf{R}^2$, we define the average of $\mathbf{w}$ on e by

$$\begin{aligned} \{\!\{\mathbf{w}\}\!\}\big|_e &= \frac{1}{2}\left(\mathbf{w}^+\big|_e + \mathbf{w}^-\big|_e\right) \in \mathbf{R}^2, &\qquad e = \partial T^+ \cap \partial T^- \in \mathcal{E}_h^i, \\ \{\!\{\mathbf{w}\}\!\}\big|_e &= \mathbf{w}\big|_e \in \mathbf{R}^2, &\qquad e \in \mathcal{E}_h^b. \end{aligned}$$

We end this subsection with some lemmas that will be used many times throughout the paper. The first states the divergence-free row property of cofactor matrices [23, p. 440].

LEMMA 3.1. *For any smooth function v,*

$$\nabla \cdot \left(\operatorname{cof}(D^2 v)_i\right) = \sum_{j=1}^{n} \frac{\partial}{\partial x_j}\left(\operatorname{cof}(D^2 v)_{ij}\right) = 0, \qquad \textit{for } i = 1, 2, ..., n,$$

where $\operatorname{cof}(D^2 v)_i$ and $\operatorname{cof}(D^2 v)_{ij}$ denote respectively the ith row and the (i, j)-entry of the cofactor matrix $\operatorname{cof}(D^2 v)$.

Next, we state some standard inverse inequalities [12, 17], as well as a discrete Sobolev inequality [9].

LEMMA 3.2. *There holds for all $T \in \mathcal{T}_h$*

$$\|v_h\|_{H^m(T)} \lesssim h_T^{q-m} \|v_h\|_{H^q(T)}, \qquad \forall v_h \in \mathbb{P}_k(T),\ 0 \le q \le m,$$

where $\mathbb{P}_k(T)$ denotes the set of all polynomials up to degree k restricted to T. Furthermore, for any piecewise polynomial with respect to the partition $\mathcal{T}_h$, there holds

$$\|v_h\|^2_{L^\infty(\Omega)} \lesssim (1 + |\ln h|)\left(\|v_h\|^2_{H^1(\mathcal{T}_h)} + \sum_{e \in \mathcal{E}_h} \frac{1}{h_e} \|\,[\![v_h]\!]\,\|^2_{L^2(e)}\right). \tag{3.1}$$

REMARK 3.3. In order to avoid the proliferation of constants, we shall use the notation $A \lesssim B$ to represent the relation $A \le \text{constant} \times B$, where the constant is independent of the mesh parameter h and any penalty parameters.

3.2. C^1 finite element methods. As a primer for more complicated looking methods to come, we consider a simple example to use the abstract framework set in Section 2, namely C^1 finite element methods. To simplify matters, we assume in this subsection that Ω is a polygonal domain and that $g = 0$ in (1.1) so that the Dirichlet boundary conditions can be imposed exactly in the finite element space. The assumptions do not guarantee the smoothness of u, but this example has the advantage of being simple. The issues of curved boundaries and inhomogeneous boundary data will be handled using penalization techniques in the next two subsections, and it is straightforward to apply this methodology to C^1 finite element methods.

We take our finite element space and auxiliary space to be

$$X_h = \{v_h \in H^2(\Omega) \cap H_0^1(\Omega);\ v_h\big|_T \in \mathbb{P}_k(T)\ \forall T \in \mathcal{T}_h\}, \quad Y = H^2(\Omega), \tag{3.2}$$

ETNA
Kent State University
http://etna.math.kent.edu

with norms

$$\|v\|_{X_h} = \|v\|_{H^1(\Omega)}, \qquad \|v\|_Y = \|v\|_{H^2(\Omega)}. \tag{3.3}$$

REMARK 3.4. To ensure that the inclusion $X_h \subset C^1(\Omega)$ holds, we require $k > 4$ in the definition (3.2); see [17].

As discussed in Remark 2.5, we first consider finite element discretizations of the linear operator (2.19). To this end, we define

$$\langle \mathrm{L}(w), y_h \rangle = \int_\Omega \big(\operatorname{cof}(D^2 u)\nabla w\big)\cdot\nabla y_h\,dx, \qquad \forall w\in Y, y_h\in X_h. \tag{3.4}$$

The goal now is to build $\mathrm{F}(\cdot)$ such that (2.4) holds.

If we integrate by parts in (3.4) and use Lemma 3.1 and the C^1 continuity of the finite element space, we obtain the following identity:

$$\langle \mathrm{L}(w), y_h \rangle = -\int_\Omega \big(\operatorname{cof}(D^2 u) : D^2 w\big) y_h\,dx. \tag{3.5}$$

Based on (3.5), we then define the nonlinear operator $\mathrm{F}(\cdot)$ as

$$\langle \mathrm{F}(w), v_h \rangle = \int_\Omega \Big(f - \det(D^2 w)\Big) v_h\,dx.$$

REMARK 3.5. This is the same discretization one gets without considering the linear problem, but this will not be the case for other discretization schemes derived below.

THEOREM 3.6. *Suppose that $u \in H^s(\Omega)$ with $s > 3$. Then there exists an $h_0 > 0$ depending on u such that for $h \le h_0$ there exists a solution $u_h \in X_h$ to*

$$\langle \mathrm{F_h}(u_h), v_h \rangle = 0, \qquad \forall v_h \in X_h, \tag{3.6}$$

where $\mathrm{F_h}(\cdot)$ is the restriction of $\mathrm{F}(\cdot)$ to X_h. Moreover, there holds

$$h\|u-u_h\|_Y + \|u-u_h\|_{X_h} \lesssim h^{\ell-1}\|u\|_{H^\ell(\Omega)},$$

where the norms $\|\cdot\|_{X_h}$ and $\|\cdot\|_Y$ are defined by (3.3) and $\ell = \min\{k+1, s\}$. In addition, if $u \in W^{3,\infty}(\Omega)$ then

$$\|u-u_h\|_{L^2(\Omega)} \lesssim h^\ell \|u\|_{H^\ell(\Omega)} + h^{2\ell-4}(1+|\ln h|)^{\frac12}\|u\|^2_{H^\ell(\Omega)}.$$

Proof. The proof is achieved by verifying that conditions (a.1)–(a.5) hold and employing Corollary 2.3 and Theorem 2.6.

First, by the definitions of Y, X_h, and $\mathrm{F}(\cdot)$, and the assumptions on u, conditions (a.1a)–(a.1b) are satisfied. Furthermore, by the inverse inequality we have

$$\|v_h\|_{X_h} = \|v_h\|_{H^1(\Omega)} \lesssim h^{-1}\|v_h\|_{H^2(\Omega)} = \|v_h\|_Y,$$

and so (a.1c) holds with $\alpha_h = O(h^{-1})$.

Next we observe that $\mathrm{F}(\cdot)$ has the decomposition (2.3) with

$$\begin{aligned} \langle \mathrm{F}^{(0)}, v_h \rangle &= \int_\Omega f v_h\,dx, \qquad \langle \mathrm{F}^{(1)}(u_h), v_h\rangle = 0,\\ \langle \mathrm{F}^{(2)}(u_h), v_h \rangle &= -\int_\Omega \det(D^2 u_h) v_h\,dx, \end{aligned} \tag{3.7}$$

ETNA
Kent State University
http://etna.math.kent.edu

and since $\mathrm{F}(u) = 0$, assumption (a.2) holds.

Continuing, we use (3.7) and (2.5) to conclude that for any $v, w \in Y$ and $y_h \in X_h$,

$$\big\langle \mathrm{DF}^{(2)}[v](w), y_h \big\rangle = -\int_\Omega \big(\mathrm{cof}(D^2 v) : D^2 w\big) y_h \, dx. \tag{3.8}$$

Therefore by (3.5), we have

$$\begin{aligned}\big\langle \mathrm{DF}[u](w), y_h \big\rangle &= \big\langle \mathrm{F}^{(1)}(w) + \mathrm{DF}^{(2)}[u](w), y_h \big\rangle \\ &= -\int_\Omega \big(\mathrm{cof}(D^2 u) : D^2 w\big) y_h \, dx = \big\langle \mathrm{L}(w), y_h \big\rangle.\end{aligned} \tag{3.9}$$

Since u is convex, the matrix $\mathrm{cof}(D^2 u)$ is positive definite. Therefore by (3.4) and an application of the Poincaré inequality, we have

$$\|v_h\|_{X_h}^2 \lesssim \|\nabla v_h\|_{L^2(\Omega)}^2 \lesssim \int_\Omega \big(\mathrm{cof}(D^2 u)\nabla v_h\big) \cdot \nabla v_h \, dx = \big\langle \mathrm{L_h}(v_h), v_h \big\rangle.$$

Moreover since $u \in H^s(\Omega)$, with $s > 3$, implies $u \in W^{2,\infty}(\Omega)$, we have

$$\big\langle \mathrm{L}(v), w \big\rangle \le |u|_{W^{2,\infty}(\Omega)} \|v\|_{H^1(\Omega)} \|w\|_{H^1(\Omega)}, \qquad \forall v, w \in H^2(\Omega) + X_h.$$

Hence both assumptions (a.3) and (a.5d) hold. Furthermore, it is easy to see that assumptions (a.5a)–(a.5c) are true by the definitions of X_h, Y, and $\mathrm{L}(\cdot)$.

The last assumption to verify is (a.4); that is, to bound the operator $\mathrm{DF}^{(2)}$. To this end, we use (3.8), the discrete Sobolev inequality (3.1), and the definition of the norms $\|\cdot\|_{X_h}$ and $\|\cdot\|_Y$ to bound $\mathrm{DF}^{(2)}$ as follows:

$$\big\langle \mathrm{DF}^{(2)}[v](w), y_h \big\rangle \le |v|_{H^2(\Omega)} |w|_{H^2(\Omega)} \|y_h\|_{L^\infty(\Omega)} \lesssim (1 + |\ln h|)^{\frac{1}{2}} \|v\|_Y \|w\|_Y \|y_h\|_{X_h}.$$

Therefore, condition (a.4) holds with $\gamma_h = O(1 + |\ln h|)^{\frac{1}{2}}$.

It remains to show that

$$\inf_{v_h \in X_h} \big(\|u - v_h\|_Y + \alpha_h C_h \|u - u_h\|_{X_h}\big) \le \tau_0 \beta_h / (2\alpha\gamma_h) \quad \text{for some } \tau_0 \in (0,1),$$

in order to apply Corollary 2.1 and Theorem 2.6. Here, $C_h = 1 + C_{cont}/\beta_h = O(1)$. Thus, by (3.3) and since $\alpha_h = O(h^{-1})$ and $\gamma_h = O(1 + |\ln h|)^{1/2}$, this last expression reduces to

$$\inf_{v_h \in X_h} \left(\|u - v_h\|_{H^2(\Omega)} + \frac{1}{h} \|u - u_h\|_{H^1(\Omega)} \right) = O(h(1 + |\ln h|)^{-1/2}). \tag{3.10}$$

By standard approximation properties of X_h [12, 17], we have

$$\inf_{v_h \in X_h} \left(\|u - v_h\|_{H^2(\Omega)} + \frac{1}{h} \|u - u_h\|_{H^1(\Omega)} \right) \lesssim h^{\ell-2} \|u\|_{H^\ell(\Omega)}.$$

Therefore since $s > 3$ and $k > 4$ (cf. Remark 3.4), condition (3.10) holds provided h is sufficiently small.

Finally, applying Corollary 2.3 we obtain

$$\begin{aligned}\|u - u_h\|_{X_h} &\lesssim \inf_{v_h \in X_h} (\|u - v_h\|_{X_h} + h\|u - v_h\|_Y) \lesssim h^{\ell-1} \|u\|_{H^\ell(\Omega)}, \\ \|u - u_h\|_Y &\lesssim \inf_{v_h \in X_h} (\|u - v_h\|_Y + h^{-1}\|u - v_h\|_{X_h}) \lesssim h^{\ell-2} \|u\|_{H^\ell(\Omega)},\end{aligned}$$

ETNA
Kent State University
http://etna.math.kent.edu

and by applying Theorem 2.6 we obtain

$$\begin{aligned}\|u-u_h\|_{L^2(\Omega)} &\lesssim \sup_{\varphi\in H^2(\Omega)} \inf_{\varphi_h\in X_h} \|\varphi\|_{H^2(\Omega)}^{-1}\Big(\|u-u_h\|_{X_h}\|\varphi-\varphi_h\|_{X_h} \\ &\qquad + (1+|\ln h|)^{\frac12}\|u-u_h\|_Y^2\|\varphi_h\|_{X_h}\Big) \\ &\lesssim h^\ell\|u\|_{H^\ell(\Omega)} + (1+|\ln h|)^{\frac12} h^{2\ell-4}\|u\|_{H^\ell(\Omega)}^2. \qquad \square\end{aligned}$$

3.3. C^0 finite element methods. The use of Lagrange finite elements and Nitsche's method [37] to compute the solution of the Monge-Ampère equation was recently introduced and analyzed in [10]. In this section, we show how this method can fit into the abstract framework set in Section 2. To this end, we define the finite element space $X_h \subset H^1(\Omega)$ as follows:

- if $T\in\mathcal{T}_h$ does not have a curved edge, then $v\big|_T$ is a polynomial of (total) degree $\le k$ in the rectilinear coordinates for T;
- if $T\in\mathcal{T}_h$ has one curved edge, then $v\big|_T$ is a polynomial of degree $\le k$ in the curvilinear coordinates of T that correspond to the rectilinear coordinates on the reference triangle. see [6, Example 2, p. 1216].

We set $Y = H^3(\mathcal{T}_h)$, and define the norms

$$\|v\|_{X_h}^2 = \|v\|_{H^1(\Omega)}^2 + \sum_{e\in\mathcal{E}_h^b}\Big(\frac{1}{h_e}\|v\|_{L^2(e)}^2 + h_e\|\nabla v\|_{L^2(e)}^2\Big), \tag{3.11}$$

$$\begin{aligned}\|v\|_Y^2 = |v|_{H^2(\mathcal{T}_h)}^2 &+ \sum_{e\in\mathcal{E}_h^b}\frac{1}{h_e^3}\|v\|_{L^2(e)}^2 \\ &+ \sum_{e\in\mathcal{E}_h}\Big(\frac{1}{h_e}\|[\![\nabla v]\!]\|_{L^2(e)}^2 + h_e\|\{\!\{D^2 v\}\!\}\|_{L^2(e)}^2\Big).\end{aligned} \tag{3.12}$$

Applying Nitsche's method to the linear operator $\mathcal{L}(\cdot)$, we define $\mathrm{L}(\cdot)$ as

$$\begin{aligned}\langle \mathrm{L}(w), y_h\rangle &= \int_\Omega \big(\mathrm{cof}(D^2u)\nabla w\big)\cdot\nabla y_h\,dx \\ &+ \sum_{e\in\mathcal{E}_h^b}\int_e\Big(\frac{\eta}{h_e}wy_h - [\![\mathrm{cof}(D^2u)\nabla w]\!]\,y_h - [\![\mathrm{cof}(D^2u)\nabla y_h]\!]\,w\Big)ds,\end{aligned} \tag{3.13}$$

where η is a positive penalization parameter. Here, the third term in the right-hand side of (3.13) ensures consistency of the operator, while the fourth term imposes symmetry.

We now derive the discretization $\mathrm{F}(\cdot)$ based on (3.13). Integrating by parts and using the divergence-free row property of cofactor matrices, we obtain

$$\begin{aligned}\langle \mathrm{L}(w), y_h\rangle = &-\sum_{T\in\mathcal{T}_h}\int_T\big(\mathrm{cof}(D^2u):D^2w\big)y_h\,dx \\ &+ \sum_{e\in\mathcal{E}_h^i}\int_e [\![\{\!\{\mathrm{cof}(D^2u)\}\!\}\nabla w]\!]\,y_h\,ds \\ &+ \sum_{e\in\mathcal{E}_h^b}\int_e\Big(\frac{\eta}{h_e}wy_h - [\![\mathrm{cof}(D^2u)\nabla y_h]\!]\,w\Big)ds.\end{aligned} \tag{3.14}$$

ETNA
Kent State University
http://etna.math.kent.edu

Based on the identity (3.14), we define the discrete nonlinear operator as

$$\begin{aligned}\langle \mathrm{F}(w), v_h\rangle = \sum_{T\in\mathcal{T}_h} \big(f-\det(D^2w)\big)\, v_h\,dx + \sum_{e\in\mathcal{E}_h^i}\int_e \big[\!\big[\big\{\!\!\big\{ \operatorname{cof}(D^2w)\big\}\!\!\big\}\nabla w\big]\!\big]\, v_h\,ds \\ + \sum_{e\in\mathcal{E}_h^b}\Big(\frac{\eta}{h_e}(w-g)v_h - \big[\!\big[\operatorname{cof}(D^2w)\nabla v_h\big]\!\big]\,(w-g)\Big)ds.\end{aligned} \tag{3.15}$$

REMARK 3.7. The operator $\mathrm{F}(\cdot)$ can be decomposed as in (2.3) with

$$\langle \mathrm{F}^{(0)}, v_h\rangle = \sum_{T\in\mathcal{T}_h}\int_T f v_h\,dx - \sum_{e\in\mathcal{E}_h^b}\int_e \frac{\eta}{h_e} g v_h\,ds,$$

$$\langle \mathrm{F}^{(1)}(w), v_h\rangle = \sum_{e\in\mathcal{E}_h^b}\int_e\Big(\frac{\eta}{h_e} w v_h + \big[\!\big[\operatorname{cof}(D^2w)\nabla v_h\big]\!\big]\, g\Big)\,ds, \tag{3.16}$$

$$\begin{aligned}\langle \mathrm{F}^{(2)}(w), v_h\rangle = -\sum_{T\in\mathcal{T}_h}\int_T \det(D^2w) v_h\,dx + \sum_{e\in\mathcal{E}_h^i}\int_e \big[\!\big[\big\{\!\!\big\{ \operatorname{cof}(D^2w)\big\}\!\!\big\}\nabla w\big]\!\big]\, v_h\,ds \\ - \sum_{e\in\mathcal{E}_h^b}\int_e \big[\!\big[\operatorname{cof}(D^2w)\nabla v_h\big]\!\big]\, w\,ds.\end{aligned} \tag{3.17}$$

THEOREM 3.8. *Suppose that $u\in H^s(\Omega)$ for some $s>3$ and that $k\geq 3$ in the definition of X_h. Then there exists an $\eta_0>0$ and $h_0=h_0(\eta)$ such that for $\eta\geq\eta_0$ and $h\leq h_0(\eta)$ there exists a solution to*

$$\langle \mathrm{F_h}(u_h), v_h\rangle = 0, \qquad \forall v_h\in X_h, \tag{3.18}$$

where $\mathrm{F_h}(\cdot)$ is the restriction to X_h of $\mathrm{F}(\cdot)$ defined by (3.15). Moreover, there holds the following error estimates

$$h\|u-u_h\|_Y + \|u-u_h\|_{X_h} \lesssim (1+\eta)h^{\ell-1}\|u\|_{H^\ell(\Omega)},$$

where $\ell=\min\{k+1,s\}$. If $u\in W^{3,\infty}(\Omega)$, then

$$\|u-u_h\|_{L^2(\Omega)} \lesssim (1+\eta)^2\Big(h^\ell\|u\|_{H^\ell(\Omega)} + (1+|\ln h|)^{\frac12} h^{2\ell-4}\|u\|^2_{H^\ell(\Omega)}\Big).$$

Proof. The proof is very similar to that of the proof of Theorem 3.6; that is, we verify that conditions (a.1)–(a.5) hold and apply the abstract results set in Section 2.

First, we observe that (a.1a)–(a.1b) and (a.2) hold by the definitions of $\mathrm{F}(\cdot)$, X_h, Y, Remark 3.7, and the assumptions of u. Moreover, by scaling arguments, $\|v_h\|_Y \lesssim h^{-1}\|v_h\|_{X_h}$, and so assumption (a1.c) holds with $\alpha_h = O(h^{-1})$.

Next, we use the definition of $\mathrm{F}^{(2)}$ (3.17) and $\mathrm{DF}^{(2)}$ (2.5), to conclude that for any $v,w\in Y$ and $y_h\in X_h$, there holds

$$\begin{aligned}\langle \mathrm{DF}^{(2)}[v](w), y_h\rangle = -\sum_{T\in\mathcal{T}_h}\int_T\big(\operatorname{cof}(D^2v) : D^2w\big) y_h\,dx \\ + \sum_{e\in\mathcal{E}_h^i}\int_e\Big(\big[\!\big[\big\{\!\!\big\{ \operatorname{cof}(D^2v)\big\}\!\!\big\}\nabla w\big]\!\big] + \big[\!\big[\big\{\!\!\big\{ \operatorname{cof}(D^2w)\big\}\!\!\big\}\nabla v\big]\!\big]\Big) y_h\,ds \\ - \sum_{e\in\mathcal{E}_h^b}\int_e\Big(\big[\!\big[\operatorname{cof}(D^2w)\nabla y_h\big]\!\big]\, v + \big[\!\big[\operatorname{cof}(D^2v)\nabla y_h\big]\!\big]\, w\Big)\,ds.\end{aligned} \tag{3.19}$$

ETNA
Kent State University
http://etna.math.kent.edu

In particular, by setting $v = u$, noting the boundary condition (1.1) and

$$\sum_{e\in\mathcal{E}_h^i}\int_e [\![\{\!\{\operatorname{cof}(D^2w)\}\!\}\nabla u]\!]\, y_h\, ds = 0,$$

we have

$$\begin{aligned}\big\langle \mathrm{DF}^{(2)}[u](w), y_h\big\rangle = &-\sum_{T\in\mathcal{T}_h}\int_T \big(\operatorname{cof}(D^2u) : D^2w\big) y_h\, dx\\ &+\sum_{e\in\mathcal{E}_h^i}\int_e [\![\{\!\{\operatorname{cof}(D^2u)\}\!\}\nabla w]\!]\, y_h\, ds\\ &-\sum_{e\in\mathcal{E}_h^b}\int_e \Big([\![\operatorname{cof}(D^2w)\nabla y_h]\!]\, g + [\![\operatorname{cof}(D^2u)\nabla y_h]\!]\, w\Big)\, ds.\end{aligned}\tag{3.20}$$

Therefore by (2.4), (3.16), (3.20), and (3.14), we obtain

$$\begin{aligned}\big\langle \mathrm{DF}[u](w), y_h\big\rangle =&\big\langle \mathrm{DF}^{(2)}[u](w) + \mathrm{F}^{(1)}(w), y_h\big\rangle\\ =&-\sum_{T\in\mathcal{T}_h}\int_T \big(\operatorname{cof}(D^2u) : D^2w\big) y_h\, dx\\ &+\sum_{e\in\mathcal{E}_h^i}\int_e [\![\{\!\{\operatorname{cof}(D^2u)\}\!\}\nabla w]\!]\, y_h\, ds\\ &+\sum_{e\in\mathcal{E}_h^b}\Big(\frac{\eta}{h_e} w y_h - [\![\operatorname{cof}(D^2u)\nabla y_h]\!]\, w\Big)\, ds = \big\langle \mathrm{L}(w), y_h\big\rangle.\end{aligned}$$

By using standard finite element techniques (cf. [10, Lemma 3.1] and [37]), the restriction $\mathrm{L_h}(\cdot)$ of $\mathrm{L}(\cdot)$ to X_h is coercive on X_h provided η_0 is sufficiently large and is bounded in the sense of assumption (a.5d) with $M \approx C_{cont} \approx (1+\eta)$. Thus, assumptions (a.3) and (a.5) hold.

Next by (3.19), (3.11), the inverse inequality, and the discrete Sobolev inequality (3.1), we have for $v, w \in Y$ and $y_h \in X_h$,

$$\begin{aligned}\big\langle \mathrm{DF}^{(2)}[v](w), y_h\big\rangle \le &\sum_{e\in\mathcal{E}_h^i}\Big(\big\|\{\!\{D^2v\}\!\}\big\|_{L^2(e)}\big\|[\![\nabla w]\!]\big\|_{L^2(e)}\\ &+\big\|\{\!\{D^2w\}\!\}\big\|_{L^2(e)}\big\|[\![\nabla v]\!]\big\|_{L^2(e)}\Big)\|y_h\|_{L^\infty(\Omega)}\\ &+\sum_{e\in\mathcal{E}_h^b}\Big(\|D^2w\|_{L^2(e)}\|v\|_{L^2(e)} + \|D^2v\|_{L^2(e)}\|w\|_{L^2(e)}\Big)\|\nabla y_h\|_{L^\infty(\Omega)}\\ &+|v|_{H^2(\mathcal{T}_h)}|w|_{H^2(\mathcal{T}_h)}\|y_h\|_{L^\infty(\Omega)}\\ \le &(1+|\ln h|)^{\frac12}\Bigg(\sum_{e\in\mathcal{E}_h^i}\Big(\big\|\{\!\{D^2v\}\!\}\big\|_{L^2(e)}\big\|[\![\nabla w]\!]\big\|_{L^2(e)}\\ &+\big\|\{\!\{D^2w\}\!\}\big\|_{L^2(e)}\big\|[\![\nabla v]\!]\big\|_{L^2(e)}\Big) + \sum_{e\in\mathcal{E}_h^b}\frac{1}{h_e}\Big(\|D^2w\|_{L^2(e)}\|v\|_{L^2(e)}\\ &+\|D^2v\|_{L^2(e)}\|w\|_{L^2(e)}\Big) + |v|_{H^2(\mathcal{T}_h)}|w|_{H^2(\mathcal{T}_h)}\Bigg)\|y_h\|_{X_h}.\end{aligned}$$

ETNA
Kent State University
http://etna.math.kent.edu

Thus by (3.12) and many applications of the Cauchy-Schwarz inequality, we conclude

$$\left\langle \mathrm{DF}^{(2)}[v_h](w_h), y_h \right\rangle \lesssim (1 + |\ln h|)^{\frac{1}{2}} \|v_h\|_Y \|w_h\|_Y \|y_h\|_{X_h}.$$

From this calculation, condition (a.4) holds with $\gamma_h \lesssim (1 + |\ln h|)^{\frac{1}{2}}$.

As all of conditions (a.1)–(a.5) have been verified, it remains to show that

$$\begin{aligned}\inf_{v_h \in X_h} \left(\|u - v_h\|_Y + h^{-1}(1+\eta)\|u - v_h\|_{X_h} \right) &= O\left(\frac{\beta_h}{\alpha_h \gamma_h} \right) \\ &= O\left(\frac{h}{(1 + |\ln h|)^{\frac{1}{2}}} \right)\end{aligned} \tag{3.21}$$

to apply Corollary 2.3 and Theorem 2.6. By approximation properties of the finite element space X_h [6] and scaling, we have

$$\inf_{v_h \in X_h} \left(\|u - v_h\|_Y + h^{-1}(1+\eta)\|u - v_h\|_{X_h} \right) \lesssim (1+\eta) h^{\ell - 2} \|u\|_{H^\ell(\Omega)}.$$

Thus, by the definition of ℓ, we see that (3.21) holds provided $s > 3$, $k \geq 3$, and h is sufficiently small.

Finally, applying Corollary 2.3 and Theorem 2.6, we obtain

$$\begin{aligned}
\|u - u_h\|_{X_h} &\lesssim \inf_{v_h \in X_h} \left((1+\eta)\|u - v_h\|_{X_h} + h\|u - v_h\|_Y \right) \lesssim (1+\eta) h^{\ell-1} \|u\|_{H^\ell(\Omega)}, \\
\|u - u_h\|_Y &\lesssim \inf_{v_h \in X_h} \left(\|u - v_h\|_Y + h^{-1}(1+\eta)\|u - v_h\|_{X_h} \right) \lesssim (1+\eta) h^{\ell-2} \|u\|_{H^\ell(\Omega)}, \\
\|u - u_h\|_{L^2(\Omega)} &\lesssim \sup_{\varphi \in H^2(\Omega)} \inf_{\varphi_h \in X_h} \|\varphi\|^{-1}_{H^2(\Omega)} \Big((1+\eta)\|u - u_h\|_{X_h} \|\varphi - \varphi_h\|_{X_h} \\
&\qquad + (1 + |\ln h|)^{\frac{1}{2}} \|u - u_h\|_Y^2 \|\varphi_h\|_{X_h} \Big) \\
&\lesssim (1+\eta)^2 \Big(h^\ell \|u\|_{H^\ell(\Omega)} + (1 + |\ln h)^{\frac{1}{2}} h^{2\ell - 4} \|u\|^2_{H^\ell(\Omega)} \Big). \qquad \square
\end{aligned}$$

3.4. Discontinuous Galerkin methods. As our last example, we construct and analyze discontinuous Galerkin methods for the Monge-Ampère equation. We take our finite element space to consist of totally discontinuous piecewise polynomial functions. In particular, we define $X_h \subset L^2(\Omega)$ to consist of functions v_h such that

- if $T \in \mathcal{T}_h$ does not have a curved edge, then $v_h\big|_T$ is a polynomial of (total) degree $\leq k$ in the rectilinear coordinates for T;
- if $T \in \mathcal{T}_h$ has one curved edge, then $v_h\big|_T$ is a polynomial of degree $\leq k$ in the curvilinear coordinates of T that correspond to the rectilinear coordinates on the reference triangle.

We set $Y = H^3(\mathcal{T}_h)$ and define the norms

$$\begin{aligned}
\|v\|^2_{X_h} &= \|v\|^2_{H^1(\mathcal{T}_h)} + \sum_{e \in \mathcal{E}_h} \left(\frac{1}{h_e} \| [\![v]\!] \|^2_{L^2(e)} + h_e \| \{\!\{\nabla v\}\!\} \|^2_{L^2(e)} \right), \\
\|v\|^2_Y &= |v|^2_{H^2(\mathcal{T}_h)} + \sum_{e \in \mathcal{E}_h} \Big(\frac{1}{h_e^3} \| [\![v]\!] \|^2_{L^2(e)} + \frac{1}{h_e} \| [\![\nabla v]\!] \|^2_{L^2(e)} \\
&\qquad + h_e \| \{\!\{D^2 v\}\!\} \|^2_{L^2(e)} \Big).
\end{aligned} \tag{3.22}$$

ETNA
Kent State University
http://etna.math.kent.edu

REMARK 3.9. By the discrete Sobolev inequality (3.1) and the definition of $\|\cdot\|_{X_h}$, we have

$$\|v_h\|_{L^\infty(\Omega)} \lesssim (1+|\ln h|)^{\frac{1}{2}}\|v_h\|_{X_h}, \qquad \forall v_h \in X_h.$$

Similarly to the previous two subsections, we base the nonlinear method $\mathrm{F_h}(\cdot)$ on the corresponding discrete linear problem. In this case, we define the discrete linear problem corresponding to the linear operator (2.19) as

$$\begin{aligned}\langle \mathrm{L}(w), y_h\rangle = \sum_{T\in\mathcal{T}_h}\int_T \big(\mathrm{cof}(D^2u)\nabla w\big)\cdot\nabla y_h\,dx \\ - \sum_{e\in\mathcal{E}_h}\int_e\Big(\{\!\{\mathrm{cof}(D^2u)\nabla w\}\!\}\cdot[\![y_h]\!] \\ + \gamma\{\!\{\mathrm{cof}(D^2u)\nabla y_h\}\!\}\cdot[\![w]\!] - \frac{\eta}{h_e}[\![w]\!]\cdot[\![y_h]\!]\Big)ds,\end{aligned} \tag{3.23}$$

where γ is a parameter that can take the values $\{1,-1,0\}$, which correspond to the SIPG method ($\gamma=1$), NIPG method ($\gamma=-1$), and IIPG method ($\gamma=0$) [1, 8, 21, 40, 41]. The constant $\eta>0$ is again a penalty parameter.

Integrating by parts of the first term in (3.23) gives us

$$\begin{aligned}\langle \mathrm{L}(w), y_h\rangle = -\sum_{T\in\mathcal{T}_h}\int \big(\mathrm{cof}(D^2u):D^2w\big)y_h\,dx \\ + \sum_{e\in\mathcal{E}_h^i}\int_e [\![\{\!\{\mathrm{cof}(D^2u)\}\!\}\nabla w]\!]\,\{\!\{y_h\}\!\}ds \\ - \sum_{e\in\mathcal{E}_h}\int_e\Big(\gamma\{\!\{\mathrm{cof}(D^2u)\nabla y_h\}\!\}\cdot[\![w]\!] \\ - \frac{\eta}{h_e}[\![w]\!]\cdot[\![y_h]\!]\Big)ds.\end{aligned} \tag{3.24}$$

Based on the identity (3.24), we define $\mathrm{F}(\cdot)$ such that

$$\begin{aligned}\langle \mathrm{F}(w), v_h\rangle = \sum_{T\in\mathcal{T}_h}\int_T\Big(f-\det(D^2w)\Big)v_h\,dx + \sum_{e\in\mathcal{E}_h^i}\int_e\Big(\frac{\eta}{h_e}[\![w]\!]\cdot[\![v_h]\!] \\ + [\![\{\!\{\mathrm{cof}(D^2w)\}\!\}\nabla w]\!]\,\{\!\{v_h\}\!\} - \gamma\{\!\{\mathrm{cof}(D^2w)\nabla v_h\}\!\}\cdot[\![w]\!]\Big)ds \\ + \sum_{e\in\mathcal{E}_h^b}\int_e\Big(\frac{\eta}{h_e}(w-g)v_h - \gamma\,[\![\mathrm{cof}(D^2w)\nabla v_h]\!]\,(w-g)\Big)ds.\end{aligned}$$

REMARK 3.10. The decomposition (2.3) holds with

$$\langle \mathrm{F}^{(0)}, v_h \rangle = \sum_{T\in\mathcal{T}_h} \int_T f v_h \, dx - \sum_{e\in\mathcal{E}_h^b} \int_e g v_h \, ds,$$

$$\langle \mathrm{F}^{(1)}(w), v_h \rangle = \sum_{e\in\mathcal{E}_h} \int_e \frac{\eta}{h_e} [\![w]\!] \cdot [\![v_h]\!] \, ds + \gamma \sum_{e\in\mathcal{E}_h^b} \int_e \left[\!\left[\operatorname{cof}(D^2 w)\nabla v_h\right]\!\right] g \, ds, \tag{3.25}$$

$$\langle \mathrm{F}^{(2)}(w), v_h \rangle = -\sum_{T\in\mathcal{T}_h} \int_T \det(D^2 w) v_h \, dx + \sum_{e\in\mathcal{E}_h^i} \int_e \Big(\left[\!\left[\{\!\{ \operatorname{cof}(D^2 w) \}\!\} \nabla w \right]\!\right] \{\!\{ v_h \}\!\} - \gamma \sum_{e\in\mathcal{E}_h} \int_e \{\!\{ \operatorname{cof}(D^2 w) \nabla v_h \}\!\} \cdot [\![w]\!] \Big) ds. \tag{3.26}$$

THEOREM 3.11. *Suppose that $u \in H^s(\Omega)$ for some $s > 3$ and that $k \geq 3$ in the definition of X_h. Then there exists an $\eta_1 = \eta_1(\gamma) > 0$ and $h_1 = h_1(\eta) > 0$ such that for $\eta \geq \eta_1$ and $h \leq h_1(\eta)$, there exists a solution $u_h \in X_h$ satisfying*

$$\langle \mathrm{F_h}(u_h), v_h \rangle = 0, \qquad \forall v_h \in X_h. \tag{3.27}$$

Moreover, there holds

$$h\|u - u_h\|_Y + \|u - u_h\|_{X_h} \lesssim (1+\eta) h^{\ell-1} \|u\|_{H^\ell(\Omega)},$$

where $\ell = \min\{k+1, s\}$. If $u \in W^{3,\infty}(\Omega)$ and $\gamma = 1$, then

$$\|u - u_h\|_{L^2(\Omega)} \lesssim (1+\eta)^2 \Big(h^\ell \|u\|_{H^\ell(\Omega)} + (1 + |\ln h|)^{\frac{1}{2}} h^{2\ell-4} \|u\|^2_{H^\ell(\Omega)} \Big).$$

REMARK 3.12. For the NIPG case $\gamma = -1$, the penalization parameter can be taken to be any positive number.

Proof. The proof follows the same arguments as the proof of Theorems 3.6 and 3.8. First, by (3.26) and (2.5) for any $v, w \in Y$ and $y_h \in X_h$, we have

$$\begin{aligned} \langle \mathrm{DF}^{(2)}[v](w), y_h \rangle = &-\sum_{T\in\mathcal{T}_h} \int_T \big(\operatorname{cof}(D^2 v) : D^2 w\big) y_h \, dx \\ &+ \sum_{e\in\mathcal{E}_h^i} \int_e \Big(\left[\!\left[\{\!\{ \operatorname{cof}(D^2 v) \}\!\} \nabla w \right]\!\right] + \left[\!\left[\{\!\{ \operatorname{cof}(D^2 w) \}\!\} \nabla v \right]\!\right] \Big) \{\!\{ y_h \}\!\} \\ &- \gamma \sum_{e\in\mathcal{E}_h} \int_e \Big(\left[\!\left[\{\!\{ \operatorname{cof}(D^2 v \}\!\} \nabla y_h \right]\!\right] \cdot [\![w]\!] + \left[\!\left[\{\!\{ \operatorname{cof}(D^2 w) \}\!\} \nabla y_h \right]\!\right] \cdot [\![v]\!] \Big) \, ds. \end{aligned} \tag{3.28}$$

Therefore, setting $v=u$ in (3.28) we have by (2.4), (3.25)–(3.26), and (3.24),

$$\begin{aligned}\big\langle \mathrm{DF}[u](w), y_h\big\rangle &= \big\langle \mathrm{DF}^{(2)}[u](w) + \mathrm{F}^{(1)}(w), y_h\big\rangle\\ &= -\sum_{T\in\mathcal{T}_h}\int_T \big(\operatorname{cof}(D^2u) : D^2 w\big) y_h\,dx\\ &\quad -\sum_{e\in\mathcal{E}_h}\int_e \Big(\gamma\{\!\{\operatorname{cof}(D^2u)\nabla y_h\}\!\}\cdot[\![w]\!]\\ &\quad -\frac{\eta}{h_e}[\![w]\!]\cdot[\![y_h]\!]\Big)ds + \sum_{e\in\mathcal{E}_h^i}\int_e\Big([\![\{\!\{\operatorname{cof}(D^2u)\}\!\}\nabla w]\!]\ \{\!\{y_h\}\!\}ds\\ &= \big\langle \mathrm{L}(w_h), y_h\big\rangle.\end{aligned}$$

Furthermore, using standard DG techniques (e.g. [40]), conditions (a.3) and (a.5d) (coercivity and continuity) holds with $C_{cont}\approx M\approx(1+\eta)$ and β_h independent of h provided that η_1 is sufficiently large (for the case $\gamma=-1$, η_1 can be taken to be any positive number).

Next, it is easy to see that assumptions (a.1a)–(a.1b), (a.2), and (a.5b)–(a.5c) hold by the definition of Y, X_h, $\mathrm{L}(\cdot)$, Remark 3.10, and the assumptions on u. Furthermore by the definitions of the norms (3.22) and the inverse inequality, assumption (a.1c) holds as well with $\alpha_h = O(h^{-1})$. Lastly, by the definition of $\mathrm{L}(\cdot)$, assumption (a.5a) (symmetry) holds provided $\gamma=1$.

It remains to show that assumption (a.4) holds. This is achieved by using (3.28), the inverse inequality, (3.1), and the Cauchy-Schwarz inequality as follows:

$$\begin{aligned}&\big\langle \mathrm{DF}^{(2)}[v](w), y_h\big\rangle \le |v|_{H^2(\mathcal{T}_h)}|w|_{H^2(\mathcal{T}_h)}\|y_h\|_{L^\infty(\Omega)}\\ &+\sum_{e\in\mathcal{E}_h^i}\Big(\big\|\{\!\{D^2v\}\!\}\big\|_{L^2(e)}\big\|[\![\nabla w]\!]\big\|_{L^2(e)} + \big\|\{\!\{D^2w\}\!\}\big\|_{L^2(e)}\big\|[\![\nabla v]\!]\big\|_{L^2(e)}\Big)\|y_h\|_{L^\infty(\Omega)}\\ &\quad+\sum_{e\in\mathcal{E}_h}\Big(\big\|\{\!\{D^2v\}\!\}\big\|_{L^2(e)}\big\|\,[\![w]\!]\big\|_{L^2(e)} + \big\|\{\!\{D^2w\}\!\}\big\|_{L^2(e)}\big\|\,[\![v]\!]\big\|_{L^2(e)}\Big)\|\nabla y_h\|_{L^\infty(\Omega)}\\ &\lesssim (1+|\ln h|)^{\frac12}\bigg[|v|_{H^2(\mathcal{T}_h)}|w|_{H^2(\mathcal{T}_h)}\\ &\quad+\sum_{e\in\mathcal{E}_h^i}\Big(\big\|\{\!\{D^2v\}\!\}\big\|_{L^2(e)}\big\|[\![\nabla w]\!]\big\|_{L^2(e)} + \big\|\{\!\{D^2w\}\!\}\big\|_{L^2(e)}\big\|[\![\nabla v]\!]\big\|_{L^2(e)}\Big)\\ &\quad+\sum_{e\in\mathcal{E}_h}h_e^{-1}\Big(\big\|\{\!\{D^2v\}\!\}\big\|_{L^2(e)}\big\|\,[\![w]\!]\big\|_{L^2(e)} + \big\|\{\!\{D^2w\}\!\}\big\|_{L^2(e)}\big\|\,[\![v]\!]\big\|_{L^2(e)}\Big)\bigg]\|y_h\|_{X_h}\\ &\lesssim (1+|\ln h|)^{\frac12}\|v\|_Y\|w\|_Y\|y_h\|_{X_h}.\end{aligned}$$

Therefore assumption (a.4) holds with $\gamma_h \lesssim (1+|\ln h|)^{\frac12}$.

To apply Corollary 2.3 we must verify that

$$\begin{aligned}\inf_{v_h\in X_h}\big(\|u-v_h\|_Y + h^{-1}(1+\eta)\|u-v_h\|_{X_h}\big) &= O\left(\beta_h/(\alpha_h\gamma_h)\right)\\ &= O\left(h(1+|\ln h|)^{-\frac12}\right).\end{aligned}$$

By standard approximation properties of X_h, this requirement reduces to

$$(1+\eta)h^{\ell-2}\|u\|_{H^\ell(\Omega)} = O\left(h(1+|\ln h|)^{-\frac12}\right),$$

ETNA
Kent State University
http://etna.math.kent.edu

which holds provided $s > 3$, $k \geq 3$, and h is sufficiently small. Therefore by Corollary 2.3, we have

$$\|u-u_h\|_{X_h} \lesssim \inf_{v_h\in X_h}\big((1+\eta)\|u-u_{ch}\|_{X_h} + h\|u-u_{ch}\|_{Y}\big) \lesssim (1+\eta)h^{\ell-1}\|u\|_{H^\ell(\Omega)},$$
$$\|u-u_h\|_{Y} \lesssim \inf_{v_h\in X_h}\big(\|u-v_h\|_{Y} + h^{-1}(1+\eta)\|u-v_h\|_{X_h}\big) \lesssim (1+\eta)h^{\ell-2}\|u\|_{H^\ell(\Omega)}.$$

If $\gamma = 1$, then by Theorem 2.6 there holds

$$\begin{aligned}\|u-u_h\|_{L^2(\Omega)} &\lesssim \sup_{\varphi\in H^2(\Omega)}\inf_{\varphi_h\in X_h}\|\varphi\|_{H^2(\Omega)}^{-1}\Big((1+\eta)\|u-u_h\|_{X_h}\|\varphi-\varphi_h\|_{X_h} \\ &\qquad + (1+\ln h)^{\frac12}\|u-u_h\|_Y\|\varphi_h\|_{X_h}\Big) \\ &\lesssim (1+\eta)^2\Big(h^\ell\|u\|_{H^\ell(\Omega)} + (1+|\ln h|)^{\frac12}h^{2\ell-4}\|u\|_{H^\ell(\Omega)}^2\Big). \qquad \square\end{aligned}$$

4. Abstract results in three dimensions. In this section, we extend the abstract results of Section 2 to the three dimensional case. As before, we let $\big(X_h, \|\cdot\|_{X_h}\big)$ be a finite dimensional space and we consider the problem of finding $u_h \in X_h$ such that (2.1) holds. We make similar assumptions as for the two dimensional counterpart, but there are some subtle differences. First, since the PDE (1.1) is cubic in 3D, the decomposition (2.3) needs to change to reflect this feature. Another difference is that we must introduce *two* auxiliary normed linear spaces to effectively analyze the method. This in turn will allow us to effectively estimate the second Gâteaux derivative of the nonlinear component of $\mathrm{F_h}(\cdot)$, which is a key element in the proof of Theorem 4.1 below.

Specifically, we make the following assumptions:

(A.1) (a) There exists two auxiliary normed linear spaces

$$(Y^{(1)}, \|\cdot\|_{Y^{(1)}}), \qquad (Y^{(2)}, \|\cdot\|_{Y^{(2)}})$$

with $X_h \subset Y^{(2)} \subset Y^{(1)}$ and $u \in Y^{(2)}$ such that $\|\cdot\|_{X_h}$ is well-defined on $Y^{(1)}$.

(b) The operator $\mathrm{F_h}(\cdot)$ can be extended to a smooth operator $\mathrm{F} : Y^{(1)} \to X_h'$ with

$$\big\langle \mathrm{F}(w_h), v_h\big\rangle = \big\langle \mathrm{F_h}(w_h), v_h\big\rangle, \qquad \forall w_h, v_h \in X_h.$$

(c) There exist constants $a_h^{(1)}, a_h^{(2)} > 0$ such that for all $v_h \in X_h$,

$$\|v_h\|_{Y_h^{(1)}} \leq a_h^{(1)}\|v_h\|_{X_h}, \qquad \|v_h\|_{Y_h^{(2)}} \leq a_h^{(2)}\|v_h\|_{X_h}. \tag{4.1}$$

(A.2) (a) The nonlinear operator $\mathrm{F}(\cdot)$ is consistent with $\mathcal{F}(\cdot)$ in the sense that $\mathrm{F}(u) = 0$.

(b) $\mathrm{F}(\cdot)$ can be decomposed as

$$\mathrm{F}(\cdot) = \mathrm{F}^{(3)}(\cdot) + \mathrm{F}^{(1)}(\cdot) + \mathrm{F}^{(0)}, \tag{4.2}$$

where $\mathrm{F}^{(3)}(\cdot)$ is cubic, $\mathrm{F}^{(1)}(\cdot)$ is linear, and $\mathrm{F}^{(0)}$ is constant in their arguments.

(A.3) Define the linear operator $\mathrm{L} : Y^{(1)} \mapsto X_h'$ as

$$\mathrm{L}(w) := \mathrm{DF}[u](w) = \lim_{t\to0}\frac{\mathrm{F}(u+tw)-\mathrm{F}(u)}{t} = \mathrm{F}^{(1)}(w) + \mathrm{DF}^{(3)}[u](w), \tag{4.3}$$

and denote by $\mathrm{L_h}$ the restriction of L to X_h. Then there exists a constant $b_h > 0$ such that the following coercivity condition holds:

$$b_h\|v_h\|_{X_h}^2 \leq \big\langle \mathrm{L_h}(v_h), v_h\big\rangle, \qquad \forall v_h \in X_h. \tag{4.4}$$

ETNA
Kent State University
http://etna.math.kent.edu

(A.4) Define

$$\mathrm{D}^2\mathrm{F}^{(3)}[v](w,z) := \lim_{t\to 0} \frac{\mathrm{DF}^{(3)}[v+tz](w) - \mathrm{DF}^{(3)}[v](w)}{t}.$$

Then there exists a constant c_h such that for all $v, w \in Y^{(1)}$,

$$\big\|\mathrm{D}^2\mathrm{F}^{(3)}[u](w,v)\big\|_{X_h'} \le c_h \|v\|_{Y^{(1)}} \|w\|_{Y^{(1}}, \tag{4.5}$$

and for all $v, w, z \in Y^{(2)}$,

$$\big\|\mathrm{D}^2\mathrm{F}^{(3)}[v](w,z)\big\|_{X_h'} \le c_h \|v\|_{Y^{(2)}} \|w\|_{Y^{(2)}} \|z\|_{Y^{(2)}}.$$

(A.5) (a) The operator $\mathrm{L}(\cdot)$ is symmetric and can be naturally extended such that $\mathrm{L} : H^2(\Omega) \to Y^{(1)\prime}$.
(b) The norm $\|\cdot\|_{X_h}$ is well-defined on $H^2(\Omega)$.
(c) The operator $\mathrm{L}(\cdot)$ is consistent with $\mathcal{L}(\cdot)$ (defined by (2.19)) in the sense that

$$\big\langle \mathrm{L}(v), w\big\rangle = \big\langle \mathcal{L}(v), w\big\rangle, \qquad \forall v \in H^2(\Omega) \cap H_0^1(\Omega),\ w \in Y^{(1)}.$$

(d) $\mathrm{L}(\cdot)$ is bounded in the sense that there exists an $M > 0$ such that

$$\big\langle \mathrm{L}(v), w\big\rangle \le M \|v\|_{X_h} \|w\|_{X_h}, \qquad \forall v, w \in X_h + H^2(\Omega).$$

(e) u is strictly convex in Ω and $u \in W^{3,\infty}(\Omega)$.

THEOREM 4.1. *Suppose that assumptions (A.1)–(A.4) hold, and let $u_{ch} \in X_h$ be the unique solution to*

$$\big\langle \mathrm{L_h}(u_{ch}), v_h\big\rangle = \big\langle \mathrm{L}(u), v_h\big\rangle, \qquad \forall v_h \in X_h.$$

Suppose that

$$\|u - u_{ch}\|_{Y^{(1)}}^2 + \|u - u_{ch}\|_{Y^{(2)}}^3 \le \tau_1 \frac{b_h}{4c_h} \min\left\{ \frac{1}{a_h^{(1)}} \|u - u_{ch}\|_{Y^{(1)}}, \frac{1}{a_h^{(2)}} \|u - u_{ch}\|_{Y^{(2)}} \right\} \tag{4.6}$$

for some $\tau_1 \in (0,1)$. Then there exists a solution $u_h \in X_h$ to (2.1) satisfying

$$\|u - u_h\|_{X_h} \le \|u - u_{ch}\|_{X_h} + \min\left\{ \frac{1}{a_h^{(1)}} \|u - u_{ch}\|_{Y^{(1)}}, \frac{1}{a_h^{(2)}} \|u - u_{ch}\|_{Y^{(2)}} \right\}, \tag{4.7}$$

$$\|u - u_h\|_{Y^{(1)}} \le 2\|u - u_{ch}\|_{Y^{(1)}}, \tag{4.8}$$

$$\|u - u_h\|_{Y^{(2)}} \le 2\|u - u_{ch}\|_{Y^{(2)}}. \tag{4.9}$$

If in addition assumption (A.5) holds, then

$$\|u - u_{ch}\|_{L^2(\Omega)} \le \sup_{\varphi \in H^2(\Omega)} \inf_{\varphi_h \in X_h} \frac{C_E}{\|\varphi\|_{H^2(\Omega)}} \Bigg(M \|u - u_h\|_{X_h} \|\varphi - \varphi_h\|_{X_h} + c_h \Big(\frac{1}{2} \|u - u_h\|_{Y^{(1)}}^2 + \frac{1}{6} \|u - u_h\|_{Y^{(2)}}^3 \Big) \|\varphi_h\|_{X_h} \Bigg), \tag{4.10}$$

ETNA
Kent State University
http://etna.math.kent.edu

with C_E *defined by* (2.22).

REMARK 4.2. As in Corollary 2.3, it is easy to show that the quantity $\|u - u_{ch}\|_{X_h}$ can be bounded by $C \inf_{v_h \in X_h} \|u - v_h\|_{X_h}$ for some constant $C > 0$.

Proof. The proof proceeds by using similar arguments to those of Theorem 2.1. Namely, we consider the mapping $\mathcal{M} : Y^{(1)} \to X_h$ defined in Lemma 2.2, that is,

$$\mathcal{M}w = \mathrm{L_h}^{-1}\Big(\mathrm{L}(w) - \mathrm{F}(w)\Big). \tag{4.11}$$

The goal is to show that $\mathcal{M}$, when restricted to X_h, has a fixed point.

First, we note that, by (4.2), (4.3), and the consistency of $\mathrm{F}(\cdot)$,

$$\begin{aligned}
\mathrm{F}(w) &= \mathrm{F}^{(0)} + \mathrm{F}^{(1)}(w) + \mathrm{F}^{(3)}(w) \\
&= -\mathrm{F}^{(1)}(u) - \mathrm{F}^{(3)}(u) + \mathrm{L}(w) - \mathrm{DF}^{(3)}[u](w) + \mathrm{F}^{(3)}(w) \\
&= \mathrm{L}(w-u) + \int_0^1 \Big(\mathrm{DF}^{(3)}[tw + (1-t)u](w-u) - \mathrm{DF}^{(3)}[u](w-u)\Big)\,dt \\
&= \mathrm{L}(w-u) + \int_0^1 \int_0^1 \mathrm{D}^2\mathrm{F}^{(3)}\big[st(w-u) + u\big]\big(w-u, t(w-u)\big)\,dt\,ds.
\end{aligned}$$

Since $\mathrm{F}^{(3)}(\cdot)$ is cubic, the mapping $(w, v, z) \to \mathrm{D}^2\mathrm{F}^{(3)}[w](v, z)$ is trilinear. It then follows that

$$\begin{aligned}
&\int_0^1 \int_0^1 \mathrm{D}^2\mathrm{F}^{(3)}\big[st(w-u) + u\big]\big(w-u, t(w-u)\big)\,dt\,ds \\
&\qquad = \mathrm{D}^2\mathrm{F}^{(3)}[w-u](w-u, w-u) \int_0^1 \int_0^1 st^2\,dt\,ds \\
&\qquad\qquad + \mathrm{D}^2\mathrm{F}^{(3)}[u](w-u, w-u) \int_0^1 \int_0^1 t\,dt\,ds \\
&\qquad = \frac{1}{2}\mathrm{D}^2\mathrm{F}^{(3)}[u](w-u, w-u) + \frac{1}{6}\mathrm{D}^2\mathrm{F}^{(3)}[w-u](w-u, w-u).
\end{aligned}$$

Substituting this identity into (4.11) we obtain

$$\begin{aligned}
\mathrm{F}(w) &= \mathrm{L}(w-u) - \frac{1}{2}\mathrm{D}^2\mathrm{F}^{(3)}[u](w-u, u) \\
&\qquad + \frac{1}{2}\int_0^1 \mathrm{D}^2\mathrm{F}^{(3)}[tw + (1-t)u](w-u, tw + (1-t)u)\,dt \\
&= \mathrm{L}(w-u) + \frac{1}{2}\mathrm{D}^2\mathrm{F}^{(3)}[u](w-u, w-u) \\
&\quad + \frac{1}{6}\mathrm{D}^2\mathrm{F}^{(3)}[w-u](w-u, w-u),
\end{aligned} \tag{4.12}$$

and using this last identity in (4.11), we arrive at

$$\mathcal{M}w = \mathrm{L_h^{-1}}\Big(\mathrm{L}(u) - \frac{1}{2}\mathrm{D}^2\mathrm{F}^{(3)}[u](w-u, w-u) - \frac{1}{6}\mathrm{D}^2\mathrm{F}^{(3)}[w-u](w-u, w-u)\Big).$$

ETNA
Kent State University
http://etna.math.kent.edu

Hence, for any $v, w \in Y^{(1)}$ we have

$$\begin{aligned}\mathcal{M}w - \mathcal{M}v = \mathrm{L}_{\mathrm{h}}^{-1}\bigg(&\frac{1}{2}\Big(\mathrm{D}^2\mathrm{F}^{(3)}[u](v-u,v-u) - \mathrm{D}^2\mathrm{F}^{(3)}[u](w-u,w-u)\Big)\\ &+\frac{1}{6}\Big(\mathrm{D}^2\mathrm{F}^{(3)}[v-u](v-u,v-u) - \mathrm{D}^2\mathrm{F}^{(3)}[w-u](w-u,w-u)\Big)\bigg)\\ = L_h^{-1}\bigg(&\frac{1}{2}\Big(\mathrm{D}^2\mathrm{F}^{(3)}[u](v-w,v-u) + \mathrm{D}^2\mathrm{F}^{(3)}[u](w-u,v-w)\Big)\\ &+\frac{1}{6}\Big(\mathrm{D}^2\mathrm{F}^{(3)}[v-w](v-u,v-u) + \mathrm{D}^2\mathrm{F}^{(3)}[w-u](v-w,v-u)\\ &\qquad + \mathrm{D}^2\mathrm{F}^{(3)}[w-u](w-u,v-w)\Big)\bigg).\end{aligned}$$

Next we apply (4.4) and (4.5) to obtain

$$\begin{aligned}\big\|\mathcal{M}w-\mathcal{M}v\big\|_{X^h} \le \frac{1}{b_h}\bigg(&\frac{1}{2}\big\|\mathrm{D}^2\mathrm{F}^{(3)}[u](v-w,v-u)\big\|_{X_h'}\\ &+\frac{1}{2}\big\|\mathrm{D}^2\mathrm{F}^{(3)}[u](w-u,v-w)\big\|_{X_h'}\\ &+\frac{1}{6}\big\|\mathrm{D}^2\mathrm{F}^{(3)}[v-w](v-u,v-u)\big\|_{X_h'}\\ &+\frac{1}{6}\big\|\mathrm{D}^2\mathrm{F}^{(3)}[w-u](v-w,v-u)\big\|_{X_h'}\\ &+\frac{1}{6}\big\|\mathrm{D}^2\mathrm{F}^{(3)}[w-u](w-u,v-w)\big\|_{X_h'}\bigg)\\ \le \frac{c_h}{2b_h}&\Big(\|u-v\|_{Y^{(1)}} + \|u-w\|_{Y^{(1)}}\Big)\|w-v\|_{Y^{(1)}}\\ &+\frac{c_h}{6b_h}\Big(\|u-v\|_{Y^{(2)}}^2 + \|u-v\|_{Y^{(2)}}\|u-w\|_{Y^{(2)}}\\ &+\|u-w\|_{Y^{(2)}}^2\Big)\|w-v\|_{Y^{(2)}}\\ \le \frac{c_h}{2b_h}&\bigg[\Big(\|u-v\|_{Y^{(1)}} + \|u-w\|_{Y^{(1)}}\Big)\|w-v\|_{Y^{(1)}}\\ &+\frac{1}{2}\Big(\|u-v\|_{Y^{(2)}}^2 + \|u-w\|_{Y^{(2)}}^2\Big)\|w-v\|_{Y^{(2)}}\bigg].\end{aligned} \tag{4.13}$$

In particular, since $u_{ch} = \mathcal{M}u$, we have

$$\big\|u_{ch} - \mathcal{M}w\big\|_{X_h} \le \frac{c_h}{2b_h}\Big(\|u-w\|_{Y^{(1)}}^2 + \frac{1}{2}\|u-w\|_{Y^{(2)}}^3\Big). \tag{4.14}$$

Let $\mathcal{M}_h$ be the restriction of $\mathcal{M}$ to X_h, let $\mathbb{B}_\rho(u_{ch})$ be defined by (2.14), and set

$$\rho_1 := \min\left\{\frac{1}{a_h^{(1)}}\|u-u_{ch}\|_{Y^{(1)}}, \frac{1}{a_h^{(2)}}\|u-u_{ch}\|_{Y^{(2)}}\right\}. \tag{4.15}$$

ETNA
Kent State University
http://etna.math.kent.edu

Then by (4.13), (4.1), and (4.15) for $v_h, w_h \in \mathbb{B}_{\rho_1}(u_{ch})$,

$$\begin{aligned}\big\|\mathcal{M}_h w_h - \mathcal{M}_h v_h\big\|_{X^h} &\le \frac{c_h}{b_h}\Bigg[a_h^{(1)}\Big(\|u-u_{ch}\|_{Y^{(1)}} + a_h^{(1)}\rho_1\Big) \\ &\qquad + a_h^{(2)}\Big(\|u-u_{ch}\|_{Y^{(2)}}^2 + \big(a_h^{(2)}\rho_1\big)^2\Big)\Bigg]\|w_h - v_h\|_{X_h} \\ &\le \frac{2c_h}{b_h}\Big[a_h^{(1)}\|u-u_{ch}\|_{Y^{(1)}} + a_h^{(2)}\|u-u_{ch}\|_{Y^{(2)}}^2\Big]\|w_h - v_h\|_{X_h}.\end{aligned}$$

The condition (4.6) then implies

$$\big\|\mathcal{M}_h w_h - \mathcal{M}_h v_h\big\|_{X^h} \le \tau_1\|w_h - v_h\|_{X_h}, \qquad \tau_1 \in (0,1). \tag{4.16}$$

Next, by (4.14), (4.1), (4.15), and (4.6), for $w_h \in \mathbb{B}_{\rho_1}(u_{ch})$,

$$\begin{aligned}\big\|u_{ch} - \mathcal{M}_h w_h\big\|_{X_h} &\le \frac{c_h}{b_h}\Big(\|u-u_{ch}\|_{Y^{(1)}}^2 + \big(a_h^{(1)}\rho_1\big)^2 \\ &\qquad + \|u-u_{ch}\|_{Y^{(2)}}^3 + \big(a_h^{(2)}\rho_1\big)^3\Big) \\ &\le \frac{2c_h}{b_h}\Big(\|u-u_{ch}\|_{Y^{(1)}}^2 + \|u-u_{ch}\|_{Y^{(2)}}^3\Big) \\ &\le \min\left\{\frac{1}{a_h^{(1)}}\|u-u_{ch}\|_{Y^{(1)}}, \frac{1}{a_h^{(2)}}\|u-u_{ch}\|_{Y^{(2)}}\right\} = \rho_1.\end{aligned} \tag{4.17}$$

It then follows from (4.16)–(4.17) that $\mathcal{M}_h$ has a fixed point u_h in $\mathbb{B}_{\rho_1}(u_{ch})$ which is a solution to (2.1).

To derive the error estimates (4.7)–(4.9), we use the triangle inequality, (4.15), and (a.1b) to get

$$\begin{aligned}\|u-u_h\|_{X_h} &\le \|u-u_{ch}\|_{X_h} + \rho_1 = \|u-u_{ch}\|_{X_h} \\ &\qquad + \min\left\{\frac{1}{a_h^{(1)}}\|u-u_{ch}\|_{Y^{(1)}}, \frac{1}{a_h^{(2)}}\|u-u_{ch}\|_{Y^{(2)}}\right\}, \\ \|u-u_h\|_{Y^{(1)}} &\le \|u-u_h\|_{Y^{(1)}} + a_h^{(1)}\rho_1 = 2\|u-u_{ch}\|_{Y^{(1)}}, \\ \|u-u_h\|_{Y^{(2)}} &\le \|u-u_h\|_{Y^{(2)}} + a_h^{(2)}\rho_1 \le 2\|u-u_{ch}\|_{Y^{(2)}}.\end{aligned}$$

To derive the L^2 estimate (4.10), we let ψ solve the auxiliary problem (2.21). Using similar arguments to that of the proof of Theorem 2.6, we have for any $\psi_h \in X_h$,

$$\begin{aligned}\|u-u_h\|_{L^2(\Omega)}^2 &= \big\langle \mathrm{L}(u-u_h), \psi-\psi_h\big\rangle + \big\langle \mathrm{L}(u-u_h), \psi_h\big\rangle \\ &= \big\langle \mathrm{L}(u-u_h), \psi-\psi_h\big\rangle + \big\langle \mathrm{L}(u-u_h) + \mathrm{F}(u_h), \psi_h\big\rangle \\ &\le M\|u-u_h\|_{X_h}\|\psi-\psi_h\|_{X_h} + \big\langle \mathrm{L}(u-u_h) + \mathrm{F}(u_h), \psi_h\big\rangle.\end{aligned} \tag{4.18}$$

Using the identity (4.12) with $w = u_h$, we obtain

$$\begin{aligned}\big\langle \mathrm{L}(u-u_h) + \mathrm{F}(u_h), \psi_h\big\rangle &= \frac{1}{2}\Big\langle \mathrm{D}^2\mathrm{F}^{(3)}[u](u_h-u, u_h-u), \psi_h\Big\rangle \\ &\qquad + \frac{1}{6}\Big\langle \mathrm{D}^2\mathrm{F}^{(3)}[u_h-u](u_h-u, u_h-u), \psi_h\Big\rangle,\end{aligned}$$

Therefore by applying the estimates stated in assumption (A.4), we have

$$\langle \mathrm{L}(u-u_h)+\mathrm{F}(u_h),\psi_h\rangle \le c_h\Big(\frac{1}{2}\|u-u_h\|^2_{Y^{(1)}}+\frac{1}{6}\|u-u_h\|^3_{Y^{(2)}}\Big)\|\psi_h\|_{X_h}.$$

Using this last estimate in (4.18), we have

$$\begin{aligned}\|u-u_h\|^2_{L^2(\Omega)} &\le M\|u-u_h\|_{X_h}\|\psi-\psi_h\|_{X_h}\\ &\qquad + c_h\Big(\frac{1}{2}\|u-u_h\|^2_{Y^{(1)}}+\frac{1}{6}\|u-u_h\|^3_{Y^{(2)}}\Big)\|\psi_h\|_{X_h}.\end{aligned}$$

Dividing by $\|u-u_h\|_{L^2(\Omega)}$ and using the elliptic regularity estimate (2.22), we obtain (4.10). The proof is complete. ∎

REMARK 4.3. It was recently shown in [11] that the C^0 finite element method (3.18) satisfies assumptions (A.1)–(A.5) with

$$\begin{aligned} a_h^{(1)} &= O\left(h^{-1}\right), & a_h^{(2)} &= O\left(h^{-\frac{3}{2}}\right),\\ b_h &= O(1), & c_h &= O\left(h^{-\frac{1}{2}}\right).\end{aligned}$$

As a result, the authors showed there exists a solution u_h to the method (3.18) in three dimensions and derived quasi-optimal error estimates provided that $u\in H^s(\Omega)$, $s>7/2$, and cubic polynomials or higher are used. We expect that similar results will hold for the C^1 finite element method (3.6) and the discontinuous Galerkin method (3.27) as well.

5. Some concluding remarks. In this paper, we have developed and analyzed various numerical methods for the two and three dimensional Monge-Ampère equation under a general framework. The key idea to build convergent numerical schemes is to construct discretizations such that the resulting discrete linearization is stable and consistent with the continuous linearization. With this in hand, and with a few more mild conditions, we proved existence of the numerical solution as well as some abstract error estimates using a simple fixed-point technique. We expect that the analysis presented here can be extended to general Monge-Ampère equations, in which the function f depends on ∇u and u, as well as parabolic Monge-Ampère equations. Furthermore, we conjecture that the abstract framework can be expanded so that other numerical methods including mixed finite element methods, local discontinuous Galerkin methods, and Petrov-Galerkin methods can naturally fit into the setting.

REFERENCES

[1] D. N. ARNOLD, F. BREZZI, B. COCKBURN, AND L. D. MARINI, *Unified analysis of discontinuous Galerkin methods for elliptic problems*, SIAM J. Numer. Anal., 39 (2002), pp. 1749–1779.

[2] F. E. BAGINSKI AND N. WHITAKER, *Numerical solutions of boundary value problems for $\mathcal{K}$-surfaces in $\mathbf{R}^3$*, Numer. Methods Partial Differential Equations, 12 (1996), pp. 525–546.

[3] G. BAKER, *Finite element methods for elliptic equations using nonconforming elements*, Math. Comp., 31 (1977), pp. 44–59.

[4] G. BARLES AND P. E. SOUGANIDIS, *Convergence of approximation schemes for fully nonlinear second order equations*, Asymptotic Anal., 4 (1991), pp. 271–283.

[5] J. D. BENAMOU, B. D. FROESE, AND A. M. OBERMAN, *Two numerical methods for the elliptic Monge-Ampère equation*, M2AN Math. Model. Numer. Anal., 44 (2010), pp. 737–758.

[6] C. BERNARDI, *Optimal finite element interpolation on curved domains*, SIAM J. Numer. Anal., 26 (1989), pp. 1212–1240.

[7] K. BÖHMER, *On finite element methods for fully nonlinear elliptic equations of second order*, SIAM J. Numer. Anal., 46 (2008), pp. 1212–1249.

ETNA
Kent State University
http://etna.math.kent.edu

[8] ———, *Numerical Methods for Nonlinear Elliptic Differential Equations. A Synopsis*, Oxford University Press, Oxford, 2010.

[9] S. C. BRENNER, *Discrete Sobolev and Poincaré inequalities for piecewise polynomial functions*, Electron. Trans. Numer. Anal., 18 (2004), pp. 42–48.
http://etna.mcs.kent.edu/vol.18.2004/pp42-48.dir

[10] S. C. BRENNER, T. GUDI, M. NEILAN, AND L.-Y. SUNG, *C^0 penalty methods for the fully nonlinear Monge-Ampère equation*, Math. Comp., 80 (2011), pp. 1979–1995.

[11] S. C. BRENNER AND M. NEILAN, *Finite element approximations of the three dimensional Monge-Ampère equation*, ESAIM Math. Model. Numer. Anal., 46 (2012), pp. 979–1001.

[12] S. C. BRENNER AND L. R. SCOTT, *The Mathematical Theory of Finite Element Methods*, 3rd ed., Springer, New York, 2008.

[13] S. C. BRENNER AND L.-Y. SUNG, *C^0 interior penalty methods for fourth order elliptic boundary value problems on polygonal domains*, J. Sci. Comput., 22/23 (2005), pp. 83–118.

[14] X. CABRÉ AND L. A. CAFFARELLI, *Fully Nonlinear Elliptic Equations*, AMS, Providence, 1995.

[15] L. A. CAFFARELLI AND M. MILMAN (eds.), *Monge-Ampère Equation: Applications to Geometry and Optimization*, AMS, Providence, 1999.

[16] L. A. CAFFARELLI, L. NIRENBERG, AND J. SPRUCK, *The Dirichlet problem for nonlinear second-order elliptic equations I. Monge-Ampère equation*, Comm. Pure Appl. Math., 37 (1984), pp. 369–402.

[17] P. G. CIARLET, *The Finite Element Method for Elliptic Problems*, North-Holland, Amsterdam, 1978.

[18] M. G. CRANDALL, H. ISHII, AND P. L. LIONS, *User's guide to viscosity solutions of second order partial differential equations*, Bull. Amer. Math. Soc., 27 (1992), pp. 1–67.

[19] E. J. DEAN AND R. GLOWINSKI, *Numerical methods for fully nonlinear elliptic equations of the Monge-Ampère type*, Comput. Methods Appl. Mech. Engrg., 195 (2006), pp. 1344–1386.

[20] G. L. DELZANNO, L. CHACÓN, J. M. FINN, Y. CHUNG, AND G. LAPENTA, *An optimal robust equidistribution method for two-dimensional grid adaptation based on Monge-Kantorovich optimization*, J. Comput. Phys., 227 (2008), pp. 9841–9864.

[21] J. DOUGLAS AND T. DUPONT, *Interior Penalty Procedures for Elliptic and Parabolic Galerkin Methods*, Lecture Notes in Phys. 58, Springer, Berlin, 1976.

[22] G. ENGEL, K. GARIKIPATI, T. HUGHES, M. LARSON, L. MAZZEI, AND R. TAYLOR, *Continuous/discontinuous finite element approximations of fourth-order elliptic problems in structural and continuum mechanics with applications to thin beams and plates, and strain gradient elasticity*, Comput. Methods Appl. Mech. Eng., 191 (2002), pp. 3669–3750.

[23] L. C. EVANS, *Partial Differential Equations*, AMS, Providence, 1998.

[24] X. FENG AND M. NEILAN, *Mixed finite element methods for the fully nonlinear Monge-Ampère equation based on the vanishing moment method*, SIAM J. Numer. Anal., 47 (2009), pp. 1226–1250.

[25] ———, *Vanishing moment method and moment solutions for second order fully nonlinear partial differential equations*, J. Sci. Comput., 38 (2009), pp. 74–98.

[26] B. D. FROESE AND A. M. OBERMAN, *Fast finite difference solvers for singular solutions of the elliptic Monge-Ampère equation*, J. Comput. Phys., 230 (2011), pp. 818–834.

[27] E. H. GEORGOULIS AND P. HOUSTON, *Discontinuous Galerkin methods for the biharmonic problem*, IMA J. Numer. Anal., 29 (2009), pp. 573–594.

[28] D. GILBARG AND N.S. TRUDINGER, *Elliptic Partial Differential Equations of Second Order*, Springer, Berlin, 2001.

[29] B. GUAN, *The Dirichlet problem for Monge-Ampère equations in non-convex domains and spacelike hypersurfaces of constant Gauss curvature*, Trans. Amer. Math. Soc., 350 (1998), pp. 4955–4971.

[30] B. GUAN AND J. SPRUCK, *Boundary-value problems on S^n for surfaces of constant Gauss curvature*, Ann. of Math. (2), 138 (1993), pp. 601–624.

[31] ———, *The existence of hypersurfaces of constant Gauss curvature with prescribed boundary*, J. Differential Geom., 62 (2002), pp. 259–287.

[32] C. E. GUTIERREZ, *The Monge-Ampère Equation*, Birkhäuser, Boston, 2001.

[33] N. M. IVOCHKINA, *Solution of the Dirichlet problem for certain equations of Monge-Ampère type*, Mat. Sb. (N.S.), 128 (1985), pp. 403–415.

[34] H. J. KUO AND N. S. TRUDINGER, *Discrete methods for fully nonlinear elliptic equations*, SIAM J. Numer. Anal., 29 (1992), pp. 123–135.

[35] ———, *Schauder estimates for fully nonlinear elliptic difference operators*, Proc. Roy. Soc. Edinburgh Sect. A, 132 (2002), pp. 1395–1406.

[36] I. MOZOLEVSKI, E. SÜLI, AND P. R. BÖSING, *hp-version a priori error analysis of interior penalty discontinuous Galerkin finite element approximations to the biharmonic equation*, J. Sci. Comput., 30 (2007), pp. 465–491.

[37] J. A. NITSCHE, *Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unteworfen sind*, Abh. Math. Sem. Univ. Hamburg, 36 (1971), pp. 9–15.

[38] A. M. OBERMAN, *Wide stencil finite difference schemes for the elliptic Monge-Ampère equation and functions of the eigenvalues of the Hessian*, Discrete Contin. Dyn. Syst. Ser. B, 10 (2008), pp. 221–238.

[39] V. I. OLIKER AND L. D. PRUSSNER, *On the numerical solution of the equation* $\frac{\partial^2 z}{\partial x^2}\frac{\partial^2 z}{\partial y^2} - (\frac{\partial^2 z}{\partial x \partial y})^2 = f$ *and its discretizations*, Numer. Math., 54 (1988), pp. 271–293.

[40] B. RIVIÉRE, *Discontinuous Galerkin Methods for Solving Elliptic and Parabolic Equations: Theory and Implementation*, SIAM, Philadelphia, 2008.

[41] B. RIVIÉRE, M. F. WHEELER, AND V. GIRAULT, *Improved energy estimates for interior penalty, constrained and discontinuous Galerkin methods for elliptic problems I.*, Comput. Geosci., 3 (1999), pp. 337–360.

[42] M. V. SAFONOV, *Classical solution of second-order nonlinear elliptic equations*, Izv. Akad. Nauk SSSR Ser. Mat., 52 (1988), pp. 1272–1287.

[43] D. C. SORENSEN AND R. GLOWINSKI, *A quadratically constrained minimization problem arising from PDE of Monge-Ampère type*, Numer. Algorithms, 53 (2010), pp. 53–66.

[44] N. S. TRUDINGER AND X.-J. WANG, *The Monge-Ampère equation and its geometric applications*, in Handbook of Geometric Analysis Vol. I, L. Ji, P. Li, R. Schoen, and L. Simon, (eds.), Advanced Lectures in Mathematics, 7, International Press, Somerville, pp. 467–524.

[45] C. VILLANI, *Topics in Optimal Transportation*, AMS, Providence, 2003.

[46] V. ZHELIGOVSKY, O. PODVIGINA, AND U. FRISCH, *The Monge-Ampère equation: various forms and numerical solution*, J. Comput. Phys., 229 (2010), pp. 5043–5061.