BIT manuscript No.
(will be inserted by the editor)

Spline element method for Monge-Ampère equations

Gerard Awanou

Received: date / Accepted: date

Abstract We analyze the convergence of an iterative method for solving the nonlinear system resulting from a natural discretization of the Monge-Ampère equation with smooth approximations. We make the assumption, supported by numerical experiments for the two dimensional problem, that the discrete problem has a convex solution. The method we analyze is the discrete version of Newton's method in the vanishing moment methodology. Numerical experiments are given in the framework of the spline element method.

Keywords iterative methods · Monge-Ampère · C^1 conforming approximations

Mathematics Subject Classification (2000) 65N30 · 35J25

1 Introduction

This paper addresses the numerical solution of the Dirichlet problem for the Monge-Ampère equation

$$\det D^2u = f \,\text{in}\, \Omega, \quad u = g \,\text{on}\, \partial\Omega. \tag{1}$$

Here $D^2u = \Big((\partial^2 u)/(\partial x_i \partial x_j)\Big)_{i,j=1,\ldots,n}$ is the Hessian of u and f, g are given functions with $f \geq c_0 > 0$ for a constant c_0. The domain $\Omega \subset \mathbb{R}^n, n = 2, 3$ is assumed to be bounded and convex with a polygonal boundary and $\partial\Omega$ denotes its boundary.

Gerard Awanou
Department of Mathematics, Statistics, and Computer Science (M/C 249), University of Illinois at Chicago, Chicago, IL, 60607-7045
Tel.: +1-312-413-2167
Fax: +1-312-996-1491
E-mail: awanou@uic.edu

Let V_h denote a finite dimensional space of C^1 functions which are piecewise polynomials of degree d at least 2, and let us assume that $f \in L^1(\Omega)$. We consider the discrete problem: find $u_h \in V_h$ such that

$$\begin{aligned} \int_{\Omega} v_h \det D^2 u_h \, dx = \int_{\Omega} f v_h \, dx, \forall v_h \in V_h \cap H_0^1(\Omega) \\ u_h = g_h \text{on} \, \partial \Omega, \end{aligned} \tag{2}$$

where g_h is the natural interpolant in V_h of a smooth extension of g. In this paper, we make the assumption that (2) has a strictly convex solution u_h. We analyze the convergence of the following iterative method. Given an initial guess $u_h^0 \in V_h$ with $u_h^0 = g_h$ on $\partial\Omega$, find $u_h^{k+1} \in V_h$ such that $u_h^{k+1} = g_h$ on $\partial\Omega$ and such that for $\epsilon > 0$ we have $\forall v_h \in V_h \cap H_0^1(\Omega)$

$$\begin{aligned} \epsilon \int_{\Omega} \Delta u_h^{k+1} \Delta v_h \, dx + \int_{\Omega} [(\operatorname{cof} D^2 u_h^k) D u_h^{k+1}] \cdot D v_h \, dx = - \int_{\Omega} f v_h \, dx \\ + \epsilon^3 \int_{\partial\Omega} \frac{\partial v_h}{\partial n_{\partial\Omega}} \, ds + \frac{n-1}{n} \int_{\Omega} [(\operatorname{cof} D^2 u_h^k) D u_h^k] \cdot D v_h \, dx. \end{aligned} \tag{3}$$

We use the notation Dv to denote the gradient vector of the function v and recall that $\operatorname{cof} A$ denotes the matrix of cofactors of the matrix A. The main difficulties of the numerical resolution of (1) is that when it does not have a smooth solution, Newton's method (i.e. (3) with $\epsilon = 0$) breaks down.

In [5] we show that (2) is well defined and has a strictly convex solution when (1) has a smooth strictly convex solution. Less restrictive conditions under which (2) has a strictly convex solution are addressed in [6] in the framework of the Aleksandrov theory of the Monge-Ampère equation. The assumption of existence of a strictly convex solution of (2) is supported in this paper by numerical experiments in two dimension. We prove the convergence of the iterations (3) to a limit $u_{\epsilon,h}$ which solves a discrete variational problem. With that result, one may prove a quadratic convergence rate for (3) as an iterative method converging to $u_{\epsilon,h}$, using for example the techniques of [5]. That issue is not addressed in this paper since (3) is not a direct method for solving (2).

For C^1 conforming approximations, we use the spline element method [2, 8,9,13,37,3]. It uses piecewise polynomials of arbitrary degree and Lagrange multipliers to enforce continuity and smoothness conditions as well as constraints. However, unlike other methods which also use Lagrange multipliers, the constraints here are enforced exactly. More details are given in Section 4.1. An alternative to the spline element method is the Argyris finite element for the two dimensional problem or concepts from isogeometric analysis [50]. The study of C^1 conforming approximations provides a natural setting for presenting techniques for proving results on the numerical analysis of Monge-Ampère equations. These techniques may be extended to the setting of isogeometric analysis, mixed finite elements, Lagrange elements or the standard finite difference method.

1.1 Relation with other work

The first rigorous treatment of the numerical resolution of the Monge-Ampère equation was given in [46]. See also the references therein for some heuristic arguments previously proposed for the balance equation of dynamic meteorology, a Monge-Ampère type equation. The work of Oliker and Prussner in [46] is based on the notion of weak solution of (1) in the sense of Aleksandrov. Dean and Glowinski [23–25,34] suggested that the numerical resolution of (1), for the notion of weak solution in the viscosity sense, can be approached through standard discretizations of the finite element or finite difference type. It is known [35] that the notions of viscosity and Aleksandrov solutions of (1) are equivalent for $f > 0$ and continuous on $\overline{\Omega}$. The analysis of numerical methods for (1) under the assumption that the solution is smooth was first initiated in [17,15,18,19]. In particular, Böhmer analyzed the discretization (2). Böhmer proved the quadratic convergence of Newton's method for solving (2), [15, Theorem 9.1] and his method has been implemented only recently [22]. Oberman in [45] constructed finite difference schemes which satisfy the conditions of monotonicity, stability and consistency of convergence of numerical schemes to viscosity solutions. His approach through viscosity solutions was later generalized in [31,33]. Feng and Neilan proposed the notion of vanishing moment methodology [27,29,30,41]. The latter has been recently shown to be valid for strictly convex radial viscosity solutions [26]. The formal limit of the vanishing moment methodology turns out, in the case of the Monge-Ampère equation, to be the method recently proposed by Lakkis and Pryer in [39]. Neilan analyzed the method of Lakkis and Pryer for the two-dimensional problem under the assumption that the solution is smooth in [42]. Awanou and Li provided a unified analysis for both dimensions from a different point of view in [10]. Several other numerical methods for (1) have been proposed, e.g. by Mohammadi [40] and by Zheligovsky *et al* [51].

In [5,1,6,4] we started the study of the numerical resolution of (1) from the point of view of compatible discretizations. Our point of view is that, after regularization of the data, the discretization of (1) leads to a sequence of discrete problems, the solution of which are discrete strictly convex functions in a sense that has to be defined for each type of discretization. For (2), the notion of convexity used in [5] is the usual notion of convexity while in [1] we required the approximations to be piecewise strictly convex. For another example, in [4] for the (non monotone) standard finite discretization, we required a certain discrete Hessian to be positive definite. Our approach is detailed in [6] and is based on the notion of Aleksandrov solution of (1), the characterization of Aleksandrov solution based on approximation by smooth functions [49], and the technique of considering a smooth uniformly convex exhaustion of the domain. The conclusion is that for the numerical analysis of robust methods for (1), one only needs to understand how the methods perform when (1) is assumed to have a smooth strictly convex solution.

To the best of our knowledge the theoretical convergence, even for smooth solutions, of the methods proposed by Dean and Glowinski, Mohammadi, and

Zheligovsky, is not understood. It is also not known whether the vanishing moment methodology approach is valid for non radial viscosity solutions of (1). We propose to analyze these methods from the new point of view we embraced in [5,1,6,4]. The goal is thus to understand how these methods perform when (1) is assumed to have a strictly convex smooth solution. There are several motivations for pursuing this line of investigation. For example, the method of Mohammadi is known to be robust even when the right hand side f is not positive. The vanishing moment methodology allows to use a Newton type iterative method for the resolution of (2). Although this feature is shared with the mixed method approach, a complete understanding of the vanishing moment methodology could lead to the development of even more robust algorithms. We recall that the vanishing moment methodology consists in the singular perturbation problem

$$-\epsilon \Delta^2 u_\epsilon + \det D^2 u_\epsilon = f, \text{ in } \Omega, \quad u_\epsilon = g, \ \Delta u_\epsilon = \epsilon^2 \text{ on } \partial\Omega, \epsilon > 0. \tag{4}$$

In [27,29,30,41], it was assumed that if $f > 0$, (4) has a unique strictly convex solution u_ϵ and u_ϵ converges uniformly on Ω to the unique convex viscosity solution of (1). Moreover, it was assumed that

$$\begin{aligned} ||u_\epsilon||_j &= O(\epsilon^{-\frac{j-1}{2}}), j = 2,3; ||u_\epsilon||_{2,\infty} = O(\epsilon^{-1}) \\ ||\operatorname{cof} D^2 u_\epsilon||_\infty &= O(\epsilon^{-1}); ||(D^2 u_\epsilon)_{i,j}||_0 = O(\epsilon^{-\frac{1}{2}}), i,j = 1,\ldots,n. \end{aligned} \tag{5}$$

In (5), we used standard Sobolev norms notation recalled in section 2. We recall that the above assumptions were proved for radial solutions in [26]. As a consequence of these assumptions, the problem: find $u_{\epsilon,h} \in V_h$ such that $u_{\epsilon,h} = g_h$ on $\partial\Omega$ and for all $v_h \in V_h \cap H_0^1(\Omega)$,

$$\epsilon \int_\Omega \Delta u_{\epsilon,h} \Delta v_h \, dx - \int_\Omega (\det D^2 u_{\epsilon,h}) v_h \, dx = - \int_\Omega f v_h \, dx + \epsilon^3 \int_{\partial\Omega} \frac{\partial v_h}{\partial n_{\partial\Omega}} \, ds, \tag{6}$$

can be shown to be well posed and error estimates were derived. Feng and Neilan mentioned the convexity of $u_{\epsilon,h}$ as a major open problem, [28, Remark 3.2].

It is not very difficult to see that if one replaces the nonlinear operator det in (4) by the Laplace operator, one obtains a singular perturbation problem similar to the one analyzed in [44]. It is therefore reasonable to expect that the techniques in [44] can be extended to the vanishing moment methodology. Since the ultimate goal of the methodology is to produce numerical approximations for (1), we analyze in this paper directly the iterative method (3) which is Newton's method applied to problem (6). It turns out that at the discrete level the strategies used in [44, Lemma 5.1] take a simpler form. We wish to address the proof of the assumptions made in [27,29,30,41] in the framework of the Aleksandrov theory of the Monge-Ampère equation [6], taking into account the above remarks, in a separate work.

The contributions of this paper are therefore

- the validation of the vanishing moment methodology for the numerical resolution of (1), an issue which has been open for more than 7 years, if one takes into account the work in [6]. It validates the method for smooth not necessarily radial solutions if one takes only into account the work in [5] or alternatively the work of Böhmer [15]
- the convexity of the numerical solution in the vanishing moment methodology is established
- an increase in the understanding of numerical methods for Monge-Ampère type equations. In particular this paper links the vanishing moment methodology to the unifying point of view presented in [5,1,6,4].

The interested reader may recover, using the strategies of this paper, the results of [29] from the ones in [42,10]. The analysis for non smooth solutions of the mixed methods discussed in [39,43,10] will be discussed in [7].

1.2 Organization of the paper

The paper is organized as follows: in the second section, we introduce some notation and give some preliminary results. In section 3 we prove the convergence of Newton's method in the vanishing moment methodology. The last section is devoted to numerical experiments.

2 Notation and Preliminaries

We use the usual notation $L^p(\Omega), 1 \leq p \leq \infty$ for the Lebesgue spaces and $W^{k,p}(\Omega)$ for the Sobolev spaces with norms $||.||_{k,p}$ and semi-norm $|.|_{k,p}$. In particular, $H^k(\Omega) = W^{k,2}(\Omega)$ and in this case, the norm and semi-norms will be denoted respectively by $||.||_k$ and $|.|_k$. For two $n \times n$ matrices A, B, we recall the Frobenius inner product $A : B = \sum_{i,j=1}^n A_{ij} B_{ij}$, where A_{ij} and B_{ij} refer to the entries of the corresponding matrices. For a matrix field A, we denote by $\operatorname{div} A$ the vector obtained by taking the divergence of each row. We will use the notation

$$||A||_\infty := \max_{i,j} |a_{ij}|,$$

for a matrix $A = (a_{ij})_{i,j=1,\ldots,n}$ and denote by $n_{\partial\Omega}$ the unit outward normal vector to $\partial\Omega$. We make the usual convention of denoting constants by C. Our results hold for h sufficiently small. We will thus state them for $h \leq h_0 \leq 1$ where h_0 is a constant which may change from occurrences.

We require our approximation spaces V_h to satisfy the following property: there exists an interpolation operator I_h mapping $W^{l+1,p}(\Omega)$ into the space V_h for $1 \leq p \leq \infty, 0 \leq l \leq d$ such that

$$||v - I_h v||_{k,p} \leq C h^{l+1-k} ||v||_{l+1,p}, \tag{7}$$

for $0 \leq k \leq l$ and the inverse estimates

$$||v||_{s,p} \leq C h^{l-s+\min(0,\frac{n}{p}-\frac{n}{q})} ||v||_{l,q}, \forall v \in V_h, \tag{8}$$

and for $0 \leq l \leq s, 1 \leq p, q \leq \infty$.

The above assumptions are known to be satisfied for standard finite element spaces [20]. For the spline spaces used in the computations, (7) is known to hold [38]. One may view (8) as a consequence of Markov inequality, [38, p. 2], and [16, section 4.2.6] for details.

It follows from (7) that

$$||I_h v||_{k,p} \leq C||v||_{k,p},$$

for $1 \leq p \leq \infty$ and $0 \leq k \leq d$.

We will need the following lemma whose proof can be found in [5].

Lemma 1 *We have*

$$\det D^2 v = \frac{1}{n}(\operatorname{cof} D^2 v) : D^2 v = \frac{1}{n} \operatorname{div} \big((\operatorname{cof} D^2 v) D v\big).$$

And for $F(v) = \det D^2 v$ we have

$$F'(v)(w) = (\operatorname{cof} D^2 v) : D^2 w = \operatorname{div} \big((\operatorname{cof} D^2 v) D w\big),$$

for v, w sufficiently smooth.

Let us denote by $\lambda_1(D^2 v)$ and $\lambda_n(D^2 v)$ the smallest and largest eigenvalues respectively of $D^2 v$, for v piecewise smooth. We make in this paper the following assumption

Assumption 1 We assume that (2) has a strictly convex solution u_h with $0 < 2C_0 \leq \lambda_1(D^2 u_h) \leq \lambda_n(D^2 u_h) \leq C_{00}/2$ for constants C_0 and C_{00} independent of h.

We define for $\rho > 0$

$$B_\rho(u_h) = \{\, v_h \in V_h, ||v_h - u_h||_1 \leq \rho \,\},$$

and we have

Lemma 2 *Under Assumption 1, there exists a constant C_{conv} such that for $\rho \leq C_{conv} h^{1+n/2}$ and for $v_h \in B_\rho(u_h)$, v_h is strictly convex with $\lambda_1(D^2 v_h) \geq C_0$.*

Proof By the continuity of the eigenvalues of a (symmetric) matrix as a function of its entries, [48] Appendix K, or [36], there exists $\delta > 0$ such that for $|v_h - u_h|_{2,\infty} \leq \delta$ we have $|\lambda_1(D^2 v_h(x)) - \lambda_1(D^2 u_h(x))| < C_0$ for all $x \in \Omega$.

By the inverse estimate (8), we have $|v_h - u_h|_{2,\infty} \leq C_{inv} h^{-1-n/2} ||v_h - u_h||_1$. Thus taking $C_{conv} = \delta/(2C_{inv})$ we obtain for $||v_h - u_h||_1 \leq C_{conv} h^{1+n/2}$, we get for all $x \in \Omega$, $|\lambda_1(D^2 v_h(x)) \geq \lambda_1(D^2 u_h(x)) - C_0 \geq C_0$. Since v_h is piecewise convex and $v_h \in C^1(\Omega)$, v_h is convex, [21, section 5]. This completes the proof.

Arguing as in the proof of Lemma 2, one shows that if the exact smooth solution u of (1) is strictly convex, i.e. for $f \geq c_0 > 0$, then $I_h u$ is also strictly convex and the solution u_h of (2) is also strictly convex. See [5] for details.

3 Convergence of the discrete vanishing moment methodology

In this section, we make the assumption that

$$\rho \leq C_{conv} h^{1+n/2}.$$

By Lemma 1, we have for $w_h \in B_\rho(u_h)$ and $v_h \in V_h \cap H_0^1(\Omega)$,

$$\begin{aligned}\int_\Omega [(\operatorname{cof} D^2 w_h) D w_h] \cdot D v_h \, dx &= - \int_\Omega \operatorname{div}[(\operatorname{cof} D^2 w_h) D w_h] v_h \, dx \\ &= -n \int_\Omega (\det D^2 w_h) v_h \, dx.\end{aligned} \tag{9}$$

Thus, we can rewrite (3) as

$$\begin{aligned}\epsilon \int_\Omega \Delta u_h^{k+1} \Delta v_h \, dx + \int_\Omega [(\operatorname{cof} D^2 u_h^k) D u_h^{k+1}] \cdot D v_h \, dx = \epsilon^3 \int_{\partial\Omega} \frac{\partial v_h}{\partial n_{\partial\Omega}} \, ds \\ + \int_\Omega p_h^k v_h \, dx,\end{aligned} \tag{10}$$

for all $v_h \in V_h \cap H_0^1(\Omega)$ with

$$p_h^k = -f - (n-1) \det D^2 u_h^k.$$

Given $u_h^k \in B_\rho(u_h)$, with $u_h^k = g_h$ on $\partial\Omega$, let $\hat{u}_h^{k+1}$ satisfy $\hat{u}_h^{k+1} = g_h$ on $\partial\Omega$ and for all $v_h \in V_h \cap H_0^1(\Omega)$,

$$\int_\Omega [(\operatorname{cof} D^2 u_h^k) D \hat{u}_h^{k+1}] \cdot D v_h \, dx = \int_\Omega p_h^k v_h \, dx. \tag{11}$$

We note that since u_h^k is strictly convex, the existence of $\hat{u}_h^{k+1}$ follows from the Lax-Milgram lemma. The following theorem identifies $\hat{u}_h^{k+1}$ as the result of one step of Newton's method applied to (2) starting with u_h^k.

Theorem 1 *Given $u_h^k \in B_\rho(u_h), \rho \leq C_{conv} h^{1+n/2}$, the solution $\hat{u}_h^{k+1}$ of (11) solves*

$$\begin{aligned}\int_\Omega [\operatorname{div}((\operatorname{cof} D^2 u_h^k) D(\hat{u}_h^{k+1} - u_h^k)] v_h \, dx = - \int_\Omega (\det D^2 u_h^k - f) v_h \, dx \\ \hat{u}_h^{k+1} = g_h \, on \, \partial\Omega,\end{aligned} \tag{12}$$

$\forall v_h \in V_h \cap H_0^1(\Omega)$. Moreover there exists $h_0 \leq 1$ such that for $h \leq h_0$

$$||\hat{u}_h^{k+1} - u_h||_1 \leq C ||u_h^k - u_h||_1^2. \tag{13}$$

Proof The result follows from integration by parts and taking into account (9). Explicitly, using the expression of the Fréchet derivative of the determinant of Lemma 1, we obtain (12) as one step of Newton's method applied to (2). We then obtain by an integration by parts

$$-\int_{\Omega}[(\operatorname{cof} D^2u_h^k)D\hat{u}_h^{k+1}]\cdot Dv_h\,dx+\int_{\Omega}[(\operatorname{cof} D^2u_h^k)Du_h^k]\cdot Dv_h\,dx \\ =-\int_{\Omega}(\det D^2u_h^k-f)v_h\,dx.$$

By (9), we obtain

$$-\int_{\Omega}[(\operatorname{cof} D^2u_h^k)D\hat{u}_h^{k+1}]\cdot Dv_h\,dx-n\int_{\Omega}(\det D^2u_h^k)v_h\,dx \\ =-\int_{\Omega}(\det D^2u_h^k-f)v_h\,dx,$$

from which (11) follows.

The inequality (13) is nothing but a consequence of a Newton's step for C^1 conforming approximations of the Monge-Ampère equation. The quadratic convergence rate of Newton's method follows for example from [15, Theorem 9.1]. For the two dimensional problem, (13) can also be inferred from [43].

Remark 1 In a more general setting, we analyzed in [5] the pseudo-transient iterative method, one step of which, starting with u_h^k, is given by

$$-\nu\int_{\Omega}(D\hat{u}_h^{k+1}-Du_h^k)\cdot Dv_h\,dx \\ +\int_{\Omega}[\operatorname{div}((\operatorname{cof} D^2u_h^k)D(\hat{u}_h^{k+1}-u_h^k)]v_h\,dx=-\int_{\Omega}(\det D^2u_h^k-f)v_h\,dx \\ \hat{u}_h^{k+1}=g_h\text{on}\,\partial\Omega,$$

for $\nu\geq 0$. For $\nu=0$, we recover one step of Newton's method, the quadratic convergence rate of which follows from [5, 3.10]. See [5, Remark 3.2]. In both cases discussed above, the rate of convergence is mesh dependent. In particular, [5, Remark 3.3], we have

$$||\hat{u}_h^{k+1}-u_h||_1\leq C_{newton}h^{-1-\frac{n}{2}}||u_h^k-u_h||_1^2. \tag{14}$$

We will use (14) in the remaining part of this paper.

We define

$$a=\min\left\{C_{conv},\frac{1}{2C_{newton}}\right\}.$$

Corollary 1 *There exists a constant $C_1<1$ such that given $u_h^k\in B_{\rho}(u_h),\rho\leq a\,h^{1+n/2}$, we have for $h\leq h_0$ for a constant $h_0\leq 1$*

$$||\hat{u}_h^{k+1}-u_h||_1\leq C_1||u_h^k-u_h||_1. \tag{15}$$

Proof From (14), we obtain

$$\begin{aligned} ||\hat{u}_h^{k+1} - u_h||_1 &\leq C_{newton} h^{-1-\frac{n}{2}} ||u_h^k - u_h||_1^2 \\ &\leq C_{newton} h^{-1-\frac{n}{2}} \rho \, ||u_h^k - u_h||_1 \\ &\leq \frac{1}{2} ||u_h^k - u_h||_1, \end{aligned}$$

and we obtain the result.

Lemma 3 *There exists $h_0 \leq 1$ such that for $h \leq h_0$, $u_h^k \in B_\rho(u_h)$ and $\rho \leq a\, h^{1+n/2}$ we have*

$$||u_h^{k+1} - \hat{u}_h^{k+1}||_1 \leq C_3 h^{-1} \epsilon^3 + C_4 \epsilon h^{-2} (\rho + ||u_h||_1), \tag{16}$$

for positive constants C_3 and C_4. Moreover, for ϵ satisfying

$$\epsilon \leq \min \left\{ \left(\frac{(1-C_1) h \rho}{3 C_3} \right)^{\frac{1}{3}}, \frac{(1-C_1) h^2}{3 C_4}, \frac{(1-C_1) h^2 \rho}{3 C_4 ||u_h||_1} \right\}, \tag{17}$$

we have

$$||\hat{u}_h^{k+1} - u_h^{k+1}||_1 \leq (1 - C_1) \rho. \tag{18}$$

Proof We view this step as a correction of a Newton step with the regularization. Substituting (11) into (10), we obtain

$$\begin{aligned} \epsilon \int_\Omega (\Delta u_h^{k+1} - \Delta \hat{u}_h^{k+1}) \Delta v_h \, dx + \int_\Omega [(\operatorname{cof} D^2 u_h^k) D(u_h^{k+1} - \hat{u}_h^{k+1})] \cdot D v_h \, dx \\ = \epsilon^3 \int_{\partial\Omega} \frac{\partial v_h}{\partial n_{\partial\Omega}} \, ds - \epsilon \int_\Omega \Delta \hat{u}_h^{k+1} \Delta v_h \, dx. \end{aligned}$$

Substituting $v_h = u_h^{k+1} - \hat{u}_h^{k+1}$ in the above equation, and using the strict convexity of u_h^k, we obtain using a trace estimate and inverse inequalities

$$\begin{aligned} \epsilon ||\Delta(u_h^{k+1} - \hat{u}_h^{k+1})||_0^2 + C_2 |u_h^{k+1} - \hat{u}_h^{k+1}|_1^2 &\leq C \epsilon^3 \left(\int_{\partial\Omega} \left| \frac{\partial v_h}{\partial n_{\partial\Omega}} \right|^2 ds \right)^{\frac{1}{2}} \\ &\quad + \epsilon ||\Delta \hat{u}_h^{k+1}||_0 ||\Delta v_h||_0 \\ &\leq C \epsilon^3 ||v_h||_2 + C \epsilon ||\Delta \hat{u}_h^{k+1}||_0 ||\Delta v_h||_0 \\ &\leq C h^{-1} \epsilon^3 ||v_h||_1 \\ &\quad + C \epsilon h^{-2} ||\hat{u}_h^{k+1}||_1 ||v_h||_1. \end{aligned}$$

We conclude that

$$||u_h^{k+1} - \hat{u}_h^{k+1}||_1 \leq C_3 h^{-1} \epsilon^3 + C_4 \epsilon h^{-2} (\rho + ||u_h||_1).$$

We obtain (18) if we choose ϵ such that (17) is satisfied.

We can now state the main result of this paper.

Theorem 2 *There exists $h_0 \leq 1$ such that the sequence u_h^k defined by* (3) *converges to the solution u_h of* (2) *as $k \to \infty$ and $\epsilon \to 0$, for $h \leq h_0$ and an initial guess $u_h^0 \in B_\rho(u_h), \rho \leq a\, h^{1+n/2}$. Moreover, as $k \to \infty$, the sequence u_h^k converges to the unique convex solution $u_{\epsilon,h}$ of* (6) *in $B_\rho(u_h)$ for ϵ satisfying condition* (17). *We also have $||u_{\epsilon,h} - u_h||_1 \to 0$ as $\epsilon \to 0$.*

Proof By (15) and (18) we have

$$||u_h^{k+1} - u_h||_1 \leq ||\hat{u}_h^{k+1} - u_h^{k+1}||_1 + ||\hat{u}_h^{k+1} - u_h||_1 \leq (\rho - C_1\rho) + C_1\rho \leq \rho.$$

We conclude that given an initial guess u_h^0 in $B_\rho(u_h)$, we have $u_h^k \in B_\rho(u_h)$ for all k. Therefore, there exists a subsequence, which is also denoted u_h^k, which converges to an element $u_{\epsilon,h}$ of $B_\rho(u_h)$. By Lemma 2, $u_{\epsilon,h} \in V_h$ is a convex function.

We first show that $u_{\epsilon,h} = g_h$ on $\partial\Omega$ and for all $v_h \in V_h \cap H_0^1(\Omega)$, (6) holds, i.e.

$$\epsilon \int_\Omega \Delta u_{\epsilon,h} \Delta v_h \, dx - \int_\Omega (\det D^2 u_{\epsilon,h}) v_h \, dx = - \int_\Omega f v_h \, dx + \epsilon^3 \int_{\partial\Omega} \frac{\partial v_h}{\partial n_{\partial\Omega}} \, ds.$$

We then prove that the above problem has a unique solution in $B_\rho(u_h)$. Therefore the whole sequence u_h^k must converge to $u_{\epsilon,h}$. Finally we prove the convergence of u_h^k to u_h as $k \to \infty$ and $\epsilon \to 0$ and the convergence of $u_{\epsilon,h}$ to u_h as $\epsilon \to 0$.

Step 1: Passage to the limit in (3). By an inverse estimate or the equivalence of norms in a finite dimensional space, the sequence u_h^k is also bounded in $W^{2,n}(\Omega)$ and hence converges (up to a subsequence) in $W^{2,n}(\Omega)$ to a limit $u_{\epsilon,h}$. Passing in the limit in (3), we obtain (6) as follows. For $v_h \in V_h \cap H_0^1(\Omega)$, we have

$$\begin{aligned}\left| \int_\Omega (\Delta u_{\epsilon,h} - \Delta u_h^{k+1}) \Delta v_h \, dx \right| &\leq ||\Delta u_{\epsilon,h} - \Delta u_h^{k+1}||_0 ||\Delta v_h||_0 \\ &\leq C ||u_{\epsilon,h} - u_h^{k+1}||_2 ||v_h||_2 \\ &\to 0 \,\text{as}\, k \to \infty.\end{aligned}$$

Put

$$A_1 = \int_\Omega [(\text{cof}\, D^2 u_h^k - \text{cof}\, D^2 u_{\epsilon,h}) D u_h^{k+1}] \cdot D v_h \, dx,$$

and

$$A_2 = \int_\Omega [(\text{cof}\, D^2 u_{\epsilon,h})(D u_h^{k+1} - D u_{\epsilon,h})] \cdot D v_h \, dx.$$

We have by Cauchy-Schwarz inequality and the inverse estimate (8)

$$\begin{aligned}|A_2| &\leq C ||u_{\epsilon,h}||_{2,\infty}^{n-1} ||u_h^{k+1} - u_{\epsilon,h}||_1 ||v_h||_1 \\ &\leq C h^{-(n-1)(2+\frac{n}{2})} ||u_{\epsilon,h}||_2 ||u_h^{k+1} - u_{\epsilon,h}||_1 ||v_h||_1 \\ &\to 0 \,\text{as}\, k \to \infty.\end{aligned}$$

Let us denote by $(\mathrm{cof})'$ the Fréchet derivative of the mapping $A \to \mathrm{cof}\, A$. Since $(\mathrm{cof})'(A)(B)$ is the sum of terms which are products of $n-2$ components of A and is linear in the components of B, we have

$$||(\mathrm{cof})'(D^2 v)(D^2 w)||_{0,\infty} \leq C||D^2 v||_{2,\infty}^{n-2}||D^2 w||_{2,\infty}.$$

It follows that

$$\begin{aligned} |A_1| &\leq C \sum_{K \in \mathcal{T}_h} ||u_h^k - u_{\epsilon,h}||_{2,\infty}||u_h^{k+1}||_{1,K}||v_h||_{1,K} \\ &\leq C||u_h^k - u_{\epsilon,h}||_{2,\infty}||u_h^{k+1}||_1||v_h||_1 \\ &\leq Ch^{-(2+\frac{n}{2})}||u_h^k - u_{\epsilon,h}||_2||u_h^{k+1}||_1||v_h||_1 \\ &\to 0 \,\mathrm{as}\, k \to \infty, \end{aligned}$$

since the convergent sequence $||u_h^{k+1}||_1$ is bounded. Finally

$$\left| \int_\Omega [(\mathrm{cof}\, D^2 u_h^k) D u_h^{k+1}] \cdot D v_h \, dx - \int_\Omega [(\mathrm{cof}\, D^2 u_{\epsilon,h}) D u_{\epsilon,h}] \cdot D v_h \, dx \right| = |A_1 + A_2| \\ \to 0 \,\mathrm{as}\, k \to \infty.$$

Passing in the limit in (3), we have

$$\begin{aligned} \epsilon \int_\Omega \Delta u_{\epsilon,h} \Delta v_h \, dx + \int_\Omega [(\mathrm{cof}\, D^2 u_{\epsilon,h}) D u_{\epsilon,h}] \cdot D v_h \, dx = -\int_\Omega f v_h \, dx \\ + \epsilon^3 \int_{\partial\Omega} \frac{\partial v_h}{\partial n_{\partial\Omega}} \, ds + \frac{n-1}{n} \int_\Omega [(\mathrm{cof}\, D^2 u_{\epsilon,h}) D u_{\epsilon,h}] \cdot D v_h \, dx. \end{aligned}$$

By (9) we obtain (6).

Step 2: Pointwise convergence of boundary data. Since $u_h^{k+1} = g_h$ on $\partial\Omega$, it follows that u_h^{k+1} is bounded on $\partial\Omega$. Passing to a subsequence, we conclude that $u_{\epsilon,h} = g_h$ on $\partial\Omega$ as well.

Step 3: Unicity of the solution of (6) in $B_\rho(u_h)$. By Assumption 1 and Lemma 2, for $v_h \in B_\rho(u_h), \rho \leq C_{conv} h^{1+n/2}$, we have $\lambda_1(D^2 v_h) \geq C_0$. Again, by the continuity of the eigenvalues of a matrix as a function of its entries, we have, if necessary by taking h smaller, $|\lambda_n(D^2 v_h(x)) - \lambda_n(D^2 u_h(x))| < C_{00}/2$ for all $x \in \Omega$. It follows that $\lambda_n(D^2 v_h) \leq C_{00}$. Using the definition of $\lambda_1(D^2 v_h)$ and $\lambda_n(D^2 v_h)$ through the Rayleigh quotient [5], we get for each element K and $w \in H^1(K)$

$$C_0 |w|_{1,K}^2 \leq \int_K [(\mathrm{cof}\ D^2 v_h(x)) D w(x)] \cdot D w(x)\ dx \leq C_{00} |w|_{1,K}^2. \tag{19}$$

We recall that the constants C_0 and C_{00} are independent of h. Let $u_{\epsilon,h}$ and $v_{\epsilon,h}$ be two solutions of (6) in $B_\rho(u_h)$. For all $t \in [0,1]$, $t u_{\epsilon,h} + (1-t) v_{\epsilon,h} \in B_\rho(u_h)$. Thus with $w_h = u_{\epsilon,h} - v_{\epsilon,h}$, we obtain

$$\epsilon ||\Delta w_h||_0^2 - \int_\Omega (\det D^2 u_{\epsilon,h} - \det D^2 v_{\epsilon,h}) w_h \, dx = 0.$$

Thus by the mean value theorem, we have for some $t \in [0,1]$

$$\epsilon ||\Delta w_h||_0^2 - \int_\Omega [\operatorname{div}((\operatorname{cof}(tD^2 u_{\epsilon,h} + (1-t)D^2 v_{\epsilon,h}))Dw_h(x))] \cdot Dw_h(x)\, dx = 0$$

$$\epsilon ||\Delta w_h||_0^2 + \int_\Omega [\operatorname{cof}(tD^2 u_{\epsilon,h} + (1-t)D^2 v_{\epsilon,h})Dw_h(x)] \cdot Dw_h(x)\, dx = 0.$$

Using (19), we obtain

$$\begin{aligned} 0 &= \epsilon ||\Delta w_h||_0^2 + \int_\Omega [(\operatorname{cof} D^2 w_h(x))Dw_h(x)] \cdot Dw_h(x)\, dx \\ &\geq \epsilon ||\Delta w_h||_0^2 + C_0 |w_h|_1^2. \end{aligned}$$

Thus $|w_h|_1 = 0$ and since $w_h = 0$ on $\partial\Omega$, we obtain $w_h = 0$, the uniqueness of the discrete solution and the proof of the claim.

Since (6) has a unique solution in $B_\rho(u_h)$, we conclude that the whole sequence defined by (3) converges to the unique local solution of (6).

Step 4: Convergence of u_h^k to u_h as $k \to \infty$ and $\epsilon \to 0$ and of $u_{\epsilon,h}$ to u_h as $\epsilon \to 0$. We have by (15) and (16)

$$\begin{aligned} ||u_h^{k+1} - u_h||_1 &\leq ||\hat{u}_h^{k+1} - u_h^{k+1}||_1 + ||\hat{u}_h^{k+1} - u_h||_1 \\ &\leq C_3 h^{-1}\epsilon^3 + C_4 \epsilon h^{-2}(\rho + ||u_h||_1) + C_1 ||u_h^k - u_h||_1. \end{aligned} \tag{20}$$

Taking the limit as $\epsilon \to 0$, we obtain

$$||u_h^{k+1} - u_h||_1 \leq C_1 ||u_h^k - u_h||_1,$$

and we recall that $C_1 < 1$. It follows that u_h^k converges to u_h as $k \to \infty$ and $\epsilon \to 0$. Finally, since by definition $||u_h^k - u_{\epsilon,h}||_1 \to 0$ as $k \to \infty$, we obtain from (20)

$$||u_{\epsilon,h} - u_h||_1 \leq C_3 h^{-1}\epsilon^3 + C_4 \epsilon h^{-2}(\rho + ||u_h||_1) + C_1 ||u_{\epsilon,h} - u_h||_1.$$

It follows that $||u_{\epsilon,h} - u_h||_1 \leq 1/(1-C_1)(C_3 h^{-1}\epsilon^3 + C_4 \epsilon h^{-2}(\rho + ||u_h||_1))$ and we conclude that $||u_{\epsilon,h} - u_h||_1 \to 0$ as $\epsilon \to 0$. This completes the proof.

Remark 2 The convexity of the solution of the discrete variational problem obtained in the vanishing moment methodology, namely (6), has long been an open problem, [28, Remark 3.2].

4 Numerical results

The iterative method (3) depends on a parameter ϵ which has to be carefully chosen. It takes about 5 iterations to converge. As an alternative to (3), we present numerical results for a parameter independent iterative method. The latter is delicate to analyze and one can only expect a linear convergence rate. The numerical results are presented in order to illustrate some open problems in the numerical resolution of Monge-Ampère equations.

4.1 Spline element method

The spline element method has been described in [2,8,9,13,37] under different names and more recently in [3]. It can be described as a conforming discretization implementation with Lagrange multipliers. We first outline the main steps of the method, discuss its advantages and possible disadvantages. We then give more details of this approach but refer to the above references for explicit formulas.

First, start with a representation of a piecewise discontinuous polynomial as a vector in $\mathbb{R}^N$, for some integer $N > 0$. Then express boundary conditions and constraints including global continuity or smoothness conditions as linear relations. In our work, we use the Bernstein basis representation, [2,3] which is very convenient to express smoothness conditions and very popular in computer aided geometric design. Hence the term "spline" in the name of the method. Splines are piecewise polynomials with smoothness properties. One then writes a discrete version of the equation along with a discrete version of the spaces of trial and test functions. The boundary conditions and constraints are enforced using Lagrange multipliers. We are lead to saddle point problems which are solved by an augmented Lagrangian algorithm (sequences of linear equations with size $N \times N$). The approach here should be contrasted with other approaches where Lagrange multipliers are introduced before discretization, i.e. the approach in [12] or the discontinuous Galerkin methods.

The spline element method, stands out as a robust, flexible, efficient and accurate method. It can be applied to a wide range of PDEs in science and engineering in both two and three dimensions; constraints and smoothness are enforced exactly and there is no need to implement basis functions with the required properties; it is particularly suitable for fourth order PDEs; no inf-sup condition is needed for the approximation of Lagrange multipliers which arise due to the constraints, e.g. the pressure term in the Navier-Stokes equations; one gets in a single implementation approximations of variable order. Other advantages of the method include the flexibility of using polynomials of different degrees on different elements [37], the facility of implementing boundary conditions and the simplicity of a posteriori error estimates since the method is conforming for many problems. A possible disadvantage of this approach is the high number of degrees of freedom and the need to solve saddle point problems.

For illustration, we consider a general variational problem: Find $u \in W$ such that

$$a(u,v) = \langle l, v \rangle \quad \text{for all } v \in V,$$

where W and V are respectively the space of trial and test functions. We will assume that the form l is bounded and linear and a is a continuous mapping in some sense on $W \times V$ which is linear in the argument v.

Let W_h and V_h be conforming subspaces of W and V respectively. We can write

$$W_h = \{c \in \mathbf{R}^N, Rc = G\}, \ V_h = \{c \in \mathbf{R}^N, Rc = 0\},$$

for a suitable vector G and a suitable matrix R which encodes the constraints on the solution, e.g. smoothness and boundary conditions.

The condition $a(u,v)=\langle l,v\rangle$ for all $v\in V$ translates to

$$K(c)d = L^T d \quad \forall d \in V_h, \text{ that is for all } d \text{ with } Rd = 0,$$

for a suitable matrix $K(c)$ which depends on c and L is a vector of coefficients associated to the linear form l. If for example $\langle l,v\rangle = \int_\Omega fv$, then $L^T d = d^T MF$ where M is a mass matrix and F a vector of coefficients associated to the spline interpolant of f. In the linear case $K(c)$ can be written $c^T K$.

Introducing a Lagrange multiplier λ, the functional

$$K(c)d - L^T d + \lambda^T Rd,$$

vanishes identically on V_h. The stronger condition

$$K(c) + \lambda^T R = L^T,$$

along with the side condition $Rc = G$ form the discrete equations to be solved.

By a slight abuse of notation, after linearization by Newton's method, the above nonlinear equation leads to solving systems of type

$$c^T K + \lambda^T R = L^T.$$

The approximation c of $u \in W$ is thus a limit of a sequence of solutions of systems of type

$$\begin{bmatrix} K^T & R^T \\ R & 0 \end{bmatrix} \begin{bmatrix} c \\ \lambda \end{bmatrix} = \begin{bmatrix} L \\ G \end{bmatrix}.$$

It is therefore enough to consider the linear case. If we assume for simplicity that $V = W$ and that the form a is bilinear, symmetric, continuous and V-elliptic, existence of a discrete solution follows from Lax-Milgram lemma. On the other hand, the ellipticity assures uniqueness of the component c which can be retrieved by a least squares solution of the above system [2]. The Lagrange multiplier λ may not be unique. To avoid systems of large size, a variant of the augmented Lagrangian algorithm is used. For this, we consider the sequence of problems

$$\begin{pmatrix} K^T & R^T \\ R & -\mu M \end{pmatrix} \begin{bmatrix} \mathbf{c}^{(l+1)} \\ \lambda^{(l+1)} \end{bmatrix} = \begin{bmatrix} L \\ G - \mu M \lambda^{(l)} \end{bmatrix},$$

where $\lambda^{(0)}$ is a suitable initial guess for example $\lambda^{(0)} = 0$, M is a suitable matrix and $\mu > 0$ is a small parameter taken in practice in the order of 10^{-5}. It is possible to solve for $\mathbf{c}^{(l+1)}$ in terms of $\mathbf{c}^{(l)}$. A uniform convergence rate in μ for this algorithm was shown in [11].

4.2 Subharmonicity preserving iterations

We also give numerical results for the following iterative method. Given an initial guess $u_h^0 \in V_h$ with $u_h^0 = g_h$ on $\partial\Omega$, find $u_h^{k+1} \in V_h$ such that $u_h^{k+1} = g_h$ on $\partial\Omega$ and $\forall v_h \in V_h \cap H_0^1(\Omega)$

$$\int_\Omega Du_h^{k+1} \cdot Dv_h \, dx = -\int_\Omega ((\Delta u_h^k)^n + n^n(f - \det D^2 u_h^k))^{\frac{1}{n}} v_h \, dx. \tag{21}$$

The iterative method (21) is the discrete analogue of the iterative method

$$\Delta u^{k+1} = ((\Delta u^k)^n + n^n(f - \det D^2 u^k))^{\frac{1}{n}} \text{ in } \Omega, u^{k+1} = g \text{ on } \partial\Omega. \tag{22}$$

Since

$$\det D^2 u^k \leq \frac{1}{n^n}(\Delta u^k)^n, \tag{23}$$

it follows from (22) that $\Delta u^{k+1} \geq 0$. Hence, starting with an initial guess u^0 with $\Delta u^0 \geq 0$, (22) preserves subharmonicity. At the formal limit, $\det D^2 u = f \geq 0$. Thus convexity is enforced for the two dimensional problem (at the continuous level). The iterative method (22) generalizes the method

$$\Delta u_{k+1} = ((\Delta u_k)^2 + 2(f - \det D^2 u_k))^{\frac{1}{2}} \text{ in } \Omega, u_{k+1} = g \text{ on } \partial\Omega. \tag{24}$$

proposed in [14]. In [31,32], the following generalization was proposed

$$\Delta u_{k+1} = ((\Delta u_k)^n + n!(f - \det D^2 u_k))^{\frac{1}{2}} \text{ in } \Omega, u_{k+1} = g \text{ on } \partial\Omega. \tag{25}$$

It is clear that (22) and (25) are different. Moreover, (22) is better since (25) may not converge for a class of smooth functions as we now show. For $n > 2$, the method (25) can only converge for solutions of (1) which also satisfies $(\Delta u)^2 = (\Delta u)^n$. Thus even for smooth solutions, the generalization we propose is better.

Let a be such that $0 < a \leq n^n$. Then by (23), we have

$$a \det D^2 v \leq (\Delta v)^n,$$

and we can equally consider the iterative method

$$\Delta u^{k+1} = ((\Delta u^k)^n + a(f - \det D^2 u^k))^{\frac{1}{n}},$$

For $n = 2$ and $a = 2$ we get the one used in [14]. It will be referred to as the BFO iterative method.

In three dimension, we can also consider

$$\Delta u_{k+1} = ((\Delta u_k)^3 + 9(f - \det D^2 u_k))^{\frac{1}{3}}, \tag{26}$$

corresponding to $a = 9$.

However, the formulation (22), which shall henceforth be referred to as natural iterative method, appears to be better and this is supported numerically by a 2D example.

h	n_{it}	L^2 norm	rate	H^1 norm	rate
$1/2^1$	35	$1.3558\ 10^{-5}$		$1.1212\ 10^{-4}$	
$1/2^2$	36	$9.2704\ 10^{-7}$	3.87	$5.5654\ 10^{-6}$	4.33
$1/2^3$	35	$5.8359\ 10^{-8}$	3.99	$3.0329\ 10^{-7}$	4.20
$1/2^4$	35	$3.6861\ 10^{-9}$	3.98	$1.8180\ 10^{-8}$	4.06

Table 1 BFO iterative method (24) for Test 1, Lagrange elements $d = 5$

h	n_{it}	L^2 norm	rate	H^1 norm	rate
$1/2^1$	14	$3.4383\ 10^{-6}$		$8.8363\ 10^{-5}$	
$1/2^2$	17	$1.1022\ 10^{-7}$	4.96	$3.1305\ 10^{-6}$	4.82
$1/2^3$	18	$7.5096\ 10^{-9}$	3.87	$1.0762\ 10^{-7}$	4.86
$1/2^4$	18	$4.9561\ 10^{-10}$	3.92	$4.1682\ 10^{-9}$	4.69

Table 2 Natural iterative method (22) for Test 1, Lagrange elements $d = 5$

4.3 Initial guess for the iterative methods

The initial guess for the subharmonicity preserving iterations is taken as the spline approximation of the solution of the Poisson equation $\Delta u = n^n f^{1/n}$, $n = 2, 3 \text{ in } \Omega$, $u = g \text{ on } \partial\Omega$. The initial guess for the discrete vanishing moment methodology is taken as the spline approximation of the biharmonic regularization of a Poisson equation, $-\epsilon\Delta^2 u + \Delta u = n f^{1/n}\ n = 2, 3 \text{ in } \Omega$, $u = g, \Delta u = \epsilon^2 \text{ on } \partial\Omega$.

4.4 Two dimensional computational results

The computational domain is the unit square $[0, 1]^2$ which is first divided into squares of side length h. Then each square is divided into two triangles by the diagonal with negative slope. We recall that d refers to the local degree of the piecewise polynomial used.

For $g = 0$, equation (1) admits both a concave solution and a convex solution. Approximating concave solutions can be done by either changing the initial guess or the structure of the approximations.

1. Newton's method: initial guess $\pm u_0$,
2. Iterative method (24): $u_{k+1} = \pm\sqrt{(\Delta u_k)^2 + 2(f - \det D^2 u_k)}$.

We consider the following test cases:

Test 1: A smooth solution $u(x, y) = e^{(x^2+y^2)/2}$ so that $f(x, y) = (1 + x^2 + y^2)e^{(x^2+y^2)}$ and $g(x, y) = e^{(x^2+y^2)/2}$ on $\partial\Omega$.

The subharmonicity preserving iterations can be used with C^0 approximations. We compare the performance of (22) and (24) with Lagrange finite elements in Tables 1 and 2.

Test 2: A solution not in $H^2(\Omega)$, $u(x, y) = -\sqrt{2 - x^2 - y^2}$ so that $f(x, y) = 2/(2 - x^2 - y^2)^2$ and $g(x, y) = -\sqrt{2 - x^2 - y^2}$ on $\partial\Omega$.

h	L^2 norm	H^1 norm
$1/2^1$	$7.6680\ 10^{-3}$	$7.4491\ 10^{-2}$
$1/2^2$	$1.4536\ 10^{-3}$	$3.9244\ 10^{-2}$
$1/2^3$	$9.8727\ 10^{-3}$	$2.5112\ 10^{-1}$
$1/2^4$	$5.6819\ 10^{-3}$	$2.4927\ 10^{-1}$
$1/2^5$	$1.9830\ 10^{+4}$	$1.1812\ 10^{+6}$

h	L^2 norm	H^1 norm
$1/2^1$	$7.8254\ 10^{-3}$	$9.3184\ 10^{-2}$
$1/2^2$	$1.0646\ 10^{-2}$	$9.5201\ 10^{-2}$
$1/2^3$	$1.1306\ 10^{-2}$	$9.6154\ 10^{-2}$
$1/2^4$	$1.1500\ 10^{-2}$	$9.1336\ 10^{-2}$
$1/2^5$	$1.1625\ 10^{-2}$	$8.7785\ 10^{-2}$
$1/2^6$	$1.1681\ 10^{-2}$	$8.5632\ 10^{-2}$

Table 3 Vanishing moment Test 2 $\epsilon = 10^{-3}$ and $\epsilon = 10^{-2}$, $d = 5$

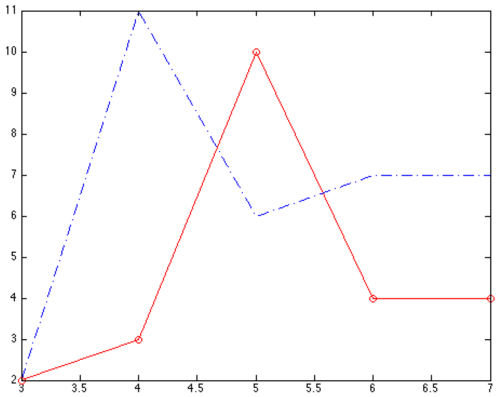

Fig. 1 Number of iterations as a function of $d = 3, \dots, 7$ with $- \circ\ \epsilon = 10^{-3}$, $- -\ \epsilon = 10^{-2}$ for Test 2 with $h = 1/2$, $d = 3$.

h	n_{it}	L^2 norm	H^1 norm
$1/2^1$	6	$2.1954\ 10^{-2}$	$1.6409\ 10^{-1}$
$1/2^2$	5	$3.6097\ 10^{-3}$	$6.1405\ 10^{-2}$
$1/2^3$	6	$1.0685\ 10^{-3}$	$4.0978\ 10^{-2}$
$1/2^4$	3	$5.0838\ 10^{-3}$	$2.8048\ 10^{-1}$
$1/2^5$	2	$2.5797\ 10^{+3}$	$2.2688\ 10^{+5}$
$1/2^6$	1	$1.8452\ 10^{+4}$	$3.5922\ 10^{+6}$

h	n_{it}	L^2 norm	H^1 norm
$1/2^1$	50	$2.3921\ 10^{-1}$	1.1900
$1/2^2$	159	$1.2585\ 10^{-1}$	$7.1292\ 10^{-1}$
$1/2^3$	151	$1.0341\ 10^{-1}$	$6.4299\ 10^{-1}$
$1/2^4$	160	$9.6031\ 10^{-2}$	$6.2088\ 10^{-1}$
$1/2^5$	199	$9.4551\ 10^{-2}$	$6.2453\ 10^{-1}$
$1/2^6$	8	$1.6977\ 10^{-2}$	$2.2925\ 10^{-1}$

Table 4 Newton's method and BFO iterative method (24) for Test 2, $d = 3$

For this non smooth solution, it is essential to adapt the choice of the parameter ϵ in the discrete vanishing methodology as a function of h, c.f. Table 3. This point was already made in [30]. We also recall that for this problem, Newton's method diverges and note that the subharmonicity preserving iterations are robust, c.f. Table 4.

d	n_{it}	L^2 norm	H^1 norm	H^2 norm
3	1	$1.2338\ 10^{-2}$	$7.6984\ 10^{-2}$	$4.4411\ 10^{-1}$
4	3	$1.6289\ 10^{-3}$	$1.4719\ 10^{-2}$	$1.3983\ 10^{-1}$
5	4	$1.5333\ 10^{-3}$	$8.7312\ 10^{-3}$	$6.0412\ 10^{-2}$
6	5	$1.2324\ 10^{-4}$	$9.7171\ 10^{-4}$	$1.0584\ 10^{-2}$
Rate		$0.18\ 0.25^{d-1}$	$4.58\ 0.25^{d}$	$59.96\ 0.3^{d+1}$

Table 5 Newton's method Test 3, Domain 1 on $\mathcal{T}_1, h = 1$

d	n_{it}	L^2 norm	H^1 norm	H^2 norm
3	1	$3.1739\ 10^{-3}$	$2.3005\ 10^{-2}$	$2.4496\ 10^{-1}$
4	7	$3.2786\ 10^{-4}$	$3.5626\ 10^{-3}$	$5.2079\ 10^{-2}$
5	5	$2.4027\ 10^{-5}$	$3.9210\ 10^{-4}$	$8.8868\ 10^{-3}$
6	6	$1.3821\ 10^{-6}$	$2.2369\ 10^{-5}$	$6.0918\ 10^{-4}$
Rate		$0.65\ 0.075^{d-1}$	$28.96\ 0.1^{d}$	$849.85\ 0.14^{d+1}$

Table 6 Newton's method Test 3, Domain 1 on $\mathcal{T}_2, h = 1/2$

4.5 Three dimensional computational results

Böhmer [15] and Feng and Neilan [30] have discussed the possibility of using C^1 finite elements in three dimension but no numerical evidence was given. This can be addressed with the spline element method. We used two computational domains both on the unit cube $[0,1]^3$ which is first divided into six tetrahedra (Domain 1) or twelve tetrahedra (Domain 2) forming a tetrahedral partition $\mathcal{T}_1$. This partition is uniformly refined following a strategy introduced in [2] similar to the one of [47] resulting in successive level of refinements $\mathcal{T}_k$, $k = 2, 3, \ldots$

We consider the following test cases

Test 3: $u(x,y,z) = e^{(x^2+y^2+z^2)/3}$ so that $f(x,y,z) = 8/81(3+2(x^2+y^2+z^2))e^{(x^2+y^2+z^2)}$ and $g(x,y,z) = e^{(x^2+y^2+z^2)/3}$ on $\partial\Omega$.

Since the solution is smooth, it is enough to use Newton's method, c.f. Tables 5 and 6. To emphasize this point, we numerically show in Table 7 the convergence as $\epsilon \to 0$ of the solution of (6) to the solution of (2). We also plot in Figure 2 the number of iterations as a function of ϵ.

Test 4: $f(x,y,z) = 0$ and $g(x,y,z) = |x - 1/2|$. For the degenerate case of this test, we did not capture the convexity of the discrete solution by discretizing (26) with the standard finite difference method. Surprisingly, with C^1 splines, we were able to capture a C^1 function which appears to approximate well the solution after 4 iterations, c.f. Figure 3.

For non smooth solutions in three dimension, it would be better to have iterative methods which preserve explicitly convexity. In some cases, i.e. for $f = 1$ and $g = 0$, we were able to capture the correct solution. Our methods did not work for the 3D analogue of Test 2.

ϵ	L^2 norm	H^1 norm	H^2 norm
10^{-1}	$6.6870\ 10^{-2}$	$3.9292\ 10^{-1}$	2.8852
10^{-2}	$1.8832\ 10^{-2}$	$1.3137\ 10^{-1}$	1.5882
10^{-3}	$2.4237\ 10^{-3}$	$2.5273\ 10^{-2}$	$5.3206\ 10^{-1}$
10^{-4}	$2.5661\ 10^{-4}$	$3.2633\ 10^{-3}$	$7.9936\ 10^{-2}$
10^{-5}	$3.1058\ 10^{-5}$	$5.0367\ 10^{-4}$	$1.2543\ 10^{-2}$
10^{-6}	$2.3519\ 10^{-5}$	$3.9165\ 10^{-4}$	$8.9744\ 10^{-3}$
10^{-7}	$2.3964\ 10^{-5}$	$3.9193\ 10^{-4}$	$8.8921\ 10^{-3}$
10^{-10}	$2.4027\ 10^{-5}$	$3.9210\ 10^{-4}$	$8.8868\ 10^{-3}$
0	$2.4027\ 10^{-5}$	$3.9210\ 10^{-4}$	$8.8868\ 10^{-3}$

Table 7 3D numerical robustness Test 3, Domain 1 on $\mathcal{T}_2, h = 1/2,\ d = 5$

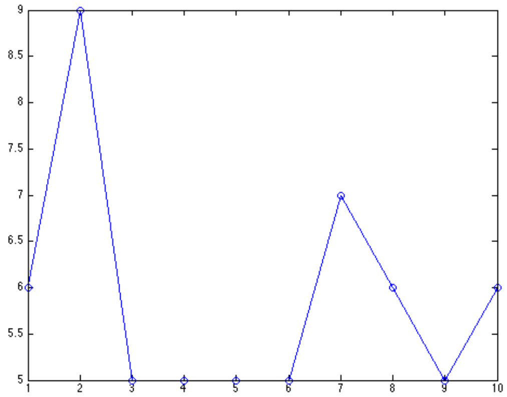

Fig. 2 Number of iterations as a function of $j = 1, \ldots, 10$ with $\epsilon = 10^{-j}$ for Test 3 with $h = 1/2,\ d = 5$ on Domain 1.

Acknowledgements

The author would like to thank the referees for a careful reading of the paper and their suggestions which lead to a better paper. This work began when the author was supported in part by a 2009-2013 Sloan Foundation Fellowship and continued while the author was in residence at the Mathematical Sciences Research Institute (MSRI) in Berkeley, California, Fall 2013. The MSRI receives major funding from the National Science Foundation under Grant No. 0932078 000. The author was partially supported by NSF DMS grant No 1319640.

References

1. G. Awanou, *Standard finite elements for the numerical resolution of the elliptic Monge-Ampère equation: classical solutions.* IMA J Numer Anal (2014) doi:

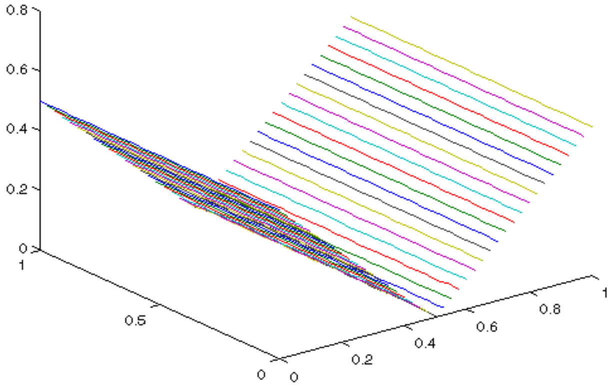

Fig. 3 Method (26) for Test 4 on Domain 2 and $\mathcal{I}_3$, $h = 1/8, d = 5, r = 1$, plane $z = 0$.

10.1093/imanum/dru028.

2. G. Awanou, *Energy methods in 3D spline approximations of the Navier-Stokes equations*, Ph.D. Dissertation, University of Georgia, Athens, Ga, 2003.
3. G. Awanou, *Robustness of a spline element method with constraints*, J. Sci. Comput., 36 (2008), pp. 421–432.
4. G. Awanou, *On standard finite difference discretizations of the elliptic Monge-Ampère equation.* Submitted, 2014.
5. ———, *Pseudo transient continuation and time marching methods for Monge-Ampère type equations.* http://arxiv.org/pdf/1301.5891v4.pdf, 2014.
6. ———, *Standard finite elements for the numerical resolution of the elliptic Monge-Ampère equation: Aleksandrov solutions.* http://arxiv.org/pdf/1310.4568v3.pdf, 2014.
7. ———, *Standard finite elements for the numerical resolution of the elliptic Monge-Ampère equation: mixed methods.* http://arxiv.org/pdf/1406.5666v1.pdf, 2014.
8. G. Awanou and M.-J. Lai, *Trivariate spline approximations of 3D Navier-Stokes equations*, Math. Comp., 74 (2005), pp. 585–601 (electronic).
9. G. Awanou, M.-J. Lai, and P. Wenston, *The multivariate spline method for scattered data fitting and numerical solution of partial differential equations*, in Wavelets and splines: Athens 2005, Mod. Methods Math., Nashboro Press, Brentwood, TN, 2006, pp. 24–74.
10. Awanou, G., Li, H.: Error analysis of a mixed finite element method for the Monge-Ampère equation. Int. J. Num. Analysis and Modeling **11**, 745–761 (2014)
11. G. M. Awanou and M. J. Lai, *On convergence rate of the augmented Lagrangian algorithm for nonsymmetric saddle point problems*, Appl. Numer. Math., 54 (2005), pp. 122–134.
12. I. Babuška, *The finite element method with Lagrangian multipliers*, Numer. Math., 20 (1972/73), pp. 179–192.
13. V. Baramidze and M.-J. Lai, *Spherical spline solution to a PDE on the sphere*, in Wavelets and splines: Athens 2005, Mod. Methods Math., Nashboro Press, Brentwood, TN, 2006, pp. 75–92.
14. 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.

15. K. Böhmer, *On finite element methods for fully nonlinear elliptic equations of second order*, SIAM J. Numer. Anal., 46 (2008), pp. 1212–1249.
16. K. Bohmer, *Numerical methods for nonlinear elliptic differential equations: a synopsis*, Oxford University Press, USA, 2010.
17. M. Bouchiba and F. B. Belgacem, *Numerical solution of Monge-Ampere equation*, Math. Balkanica (N.S.), 20 (2006), pp. 369–378.
18. 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.
19. 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.
20. S. C. Brenner and L. R. Scott, *The mathematical theory of finite element methods*, vol. 15 of Texts in Applied Mathematics, Springer-Verlag, New York, second ed., 2002.
21. W. Dahmen, *Convexity and Bernstein-Bézier polynomials*, in Curves and surfaces (Chamonix-Mont-Blanc, 1990), Academic Press, Boston, MA, 1991, pp. 107–134.
22. Davydov, O., Saeed, A.: Numerical solution of fully nonlinear elliptic equations by Böhmer's method. J. Comput. Appl. Math. **254**, 43–54 (2013)
23. E. J. Dean and R. Glowinski, *Numerical solution of the two-dimensional elliptic Monge-Ampère equation with Dirichlet boundary conditions: an augmented Lagrangian approach*, C. R. Math. Acad. Sci. Paris, 336 (2003), pp. 779–784.
24. ———, *Numerical solution of the two-dimensional elliptic Monge-Ampère equation with Dirichlet boundary conditions: a least-squares approach*, C. R. Math. Acad. Sci. Paris, 339 (2004), pp. 887–892.
25. 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.
26. Feng, X., Neilan, M.: Convergence of a fourth-order singular perturbation of the n-dimensional radially symmetric Monge–Ampère equation. Appl. Anal. **93**(8), 1626–1646 (2014)
27. X. Feng and M. Neilan, *Error analysis for mixed finite element approximations of the fully nonlinear Monge-Ampère equation based on the vanishing moment method*, SIAM J. Numer. Anal., 47 (2009), pp. 1226–1250.
28. X. Feng and M. Neilan, *A modified characteristic finite element method for a fully nonlinear formulation of the semigeostrophic flow equations*, SIAM J. Numer. Anal., 47 (2009), pp. 2952–2981.
29. X. Feng and M. Neilan, *Vanishing moment method and moment solutions for second order fully nonlinear partial differential equations*, J. Sci. Comput., 38 (2009), pp. 74–98.
30. X. Feng and M. Neilan, *Analysis of Galerkin methods for the fully nonlinear Monge-Ampère equation*, J. Sci. Comput., 47 (2011), pp. 303–327.
31. B. Froese and A. Oberman, *Convergent finite difference solvers for viscosity solutions of the elliptic Monge-Ampère equation in dimensions two and higher*, SIAM J. Numer. Anal., 49 (2011), pp. 1692–1714.
32. 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.
33. B. D. Froese and A. M. Oberman, *Convergent filtered schemes for the Monge-Ampère partial differential equation*, SIAM J. Numer. Anal., 51 (2013), pp. 423–444.
34. R. Glowinski, *Numerical methods for fully nonlinear elliptic equations*, in ICIAM 07—6th International Congress on Industrial and Applied Mathematics, Eur. Math. Soc., Zürich, 2009, pp. 155–192.
35. C. E. Gutiérrez, *The Monge-Ampère equation*, Progress in Nonlinear Differential Equations and their Applications, 44, Birkhäuser Boston Inc., Boston, MA, 2001.
36. G. Harris and C. Martin, *The roots of a polynomial vary continuously as a function of the coefficients*, Proc. Amer. Math. Soc., 100 (1987), pp. 390–392.
37. X.-L. Hu, D.-F. Han, and M.-J. Lai, *Bivariate splines of various degrees for numerical solution of partial differential equations*, SIAM J. Sci. Comput., 29 (2007), pp. 1338–1354 (electronic).
38. M.-J. Lai and L. L. Schumaker, *Spline functions on triangulations*, vol. 110 of Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 2007.

39. O. Lakkis and T. Pryer, *A finite element method for nonlinear elliptic problems*, SIAM J. Sci. Comput., 35 (2013), pp. A2025–A2045.
40. B. Mohammadi, *Optimal transport, shape optimization and global minimization*, C. R. Math. Acad. Sci. Paris, 344 (2007), pp. 591–596.
41. M. Neilan, *A nonconforming Morley finite element method for the fully nonlinear Monge-Ampère equation*, Numer. Math., 115 (2010), pp. 371–394.
42. Neilan, M.: Finite element methods for fully nonlinear second order PDEs based on a discrete Hessian with applications to the Monge–Ampère equation. J. Comput. Appl. Math. **263**, 351–369 (2014)
43. M. Neilan, *Quadratic finite element approximations of the Monge-Ampère equation*, J. Sci. Comput., 54 (2013), pp. 200–226.
44. T. K. Nilssen, X.-C. Tai, and R. Winther, *A robust nonconforming H^2-element*, Math. Comp., 70 (2001), pp. 489–505.
45. 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.
46. V. I. Oliker and L. D. Prussner, *On the numerical solution of the equation $(\partial^2 z/\partial x^2)(\partial^2 z/\partial y^2) - ((\partial^2 z/\partial x\partial y))^2 = f$ and its discretizations. I*, Numer. Math., 54 (1988), pp. 271–293.
47. M. E. G. Ong, *Uniform refinement of a tetrahedron*, SIAM J. Sci. Comput., 15 (1994), pp. 1134–1144.
48. A. M. Ostrowski, *Solution of equations and systems of equations*, Pure and Applied Mathematics, Vol. IX. Academic Press, New York-London, 1960.
49. J. Rauch and B. A. Taylor, *The Dirichlet problem for the multidimensional Monge-Ampère equation*, Rocky Mountain J. Math., 7 (1977), pp. 345–364.
50. A.-V. Vuong, C. Heinrich, and B. Simeon, *ISOGAT: a 2D tutorial MATLAB code for isogeometric analysis*, Comput. Aided Geom. Design, 27 (2010), pp. 644–655.
51. V. Zheligovsky, O. Podvigina, and U. Frisch, *The Monge-Ampère equation: various forms and numerical solution*, J. Comput. Phys., 229 (2010), pp. 5043–5061.