Numerische Mathematik (2019) 141:253–288
https://doi.org/10.1007/s00211-018-0988-9

Numerische Mathematik

Pointwise rates of convergence for the Oliker–Prussner method for the Monge–Ampère equation

Ricardo H. Nochetto[1] · **Wujun Zhang**[2]

Received: 9 November 2016 / Revised: 28 May 2018 / Published online: 21 July 2018

Abstract
We study the Oliker–Prussner method exploiting its geometric nature. We derive discrete stability and continuous dependence estimates in the max-norm by using a discrete Alexandroff estimate and the Brunn–Minkowski inequality. We show that the method is exact for all convex quadratic polynomials provided the underlying set of nodes is translation invariant within the domain; nodes still conform to the domain boundary. . This gives a suitable notion of operator consistency which, combined with stability, leads to pointwise rates of convergence for classical and non-classical solutions of the Monge–Ampère equation.

Mathematics Subject Classification 65N12 · 65N15 · 65N30 · 35J96

1 Introduction

We consider the fully nonlinear Monge–Ampère equation

$$\det D^2 u = f \quad \text{in } \Omega \tag{1.1a}$$
$$u = g \quad \text{on } \partial\Omega, \tag{1.1b}$$

Both authors were partially supported by NSF Grants DMS-1109325 and DMS-1411808. The second author was also partially supported by the Brin Postdoctoral Fellowship of the University of Maryland and the start up fund of Rutgers University.

✉ Wujun Zhang
wujun@math.rutgers.edu

Ricardo H. Nochetto
rhn@math.umd.edu

[1] Department of Mathematics and Institute for Physical Science and Technology, University of Maryland, College Park, USA

[2] Department of Mathematics, Rutgers University, New Brunswick, USA

where Ω denotes a *uniformly convex* domain in $\mathbb{R}^d$ ($d \geq 2$), $f \in C(\overline{\Omega})$ satisfies $0 < \lambda_F \leq f(x) \leq \Lambda_F$ for all $x \in \Omega$, and $g \in C(\partial\Omega)$. Such an equation arises in differential geometry, optimal mass transport and several fields of science and engineering, and has received considerable attention in recent years.

In contrast to an extensive PDE literature [10,17,30], the numerical approximation is still under development. Convergence to the *viscosity solution* (see Definition 3.3), in the general framework of fully nonlinear elliptic PDEs, is studied in the early works [3,32,33,35] and hinges on operator stability, consistency and monotonicity properties. Oberman et al. designed several finite difference methods (wide stencil schemes) within this framework [5,27,28,41]. Benamou et. al. proposed recently a convergent finite difference method, which reduces the stencil size, and showed that the method is consistent for convex polynomials [4].

A geometric notion of generalized solution for the Monge–Ampère equation is the so called *Alexandroff solution* (see Definition 3.2). It hinges on the geometric interpretation of $\int_D \det D^2 u$ as the measure of the subdifferential $|\partial u(D)|$ of a convex function u for any Borel set D. This notion of solution is weaker than the viscosity solution. Oliker and Prussner developed a discrete counterpart and proved convergence of the ensuing geometric numerical method [43]. Even though this is a natural idea, the scheme has defied analysis ever since its conception. The purpose of this paper is to fill this gap upon deriving stability and decay rates in the max norm for the scheme of [43].

Other methods do exist such as the vanishing moment method of Feng and Neilan [24–26], the numerical moment method of Feng et al. [23], the penalty method of Brenner et al. [7,8], nonconforming elements and quadratic elements by Neilan [38, 39], standard finite element methods by Awanou [2], least squares and augmented Lagrangian methods of Dean and Glowinski [18–21,44], and the C^1 finite element method of Böhmer [6]. Error estimates in $H^1(\Omega)$ are established in [7,8] for solutions u with regularity H^3 and above.

The Oliker–Prussner method reads as follows [43]. Let $\mathcal{N}_h$ be a set of nodes with quasi-uniform spacing h and let $\mathcal{T}_h$ be a subordinate mesh which induces a computational domain $\Omega_h \subset \Omega$. Let $\mathcal{N}_h^0 \subset \Omega$ denote the interior nodes and $\mathcal{N}_h^\partial \subset \partial\Omega$ the boundary nodes. The discrete solution u_h is a *convex nodal function* defined on $\mathcal{N}_h$ satisfying $u_h(x_i) = g(x_i)$ for all $x_i \in \mathcal{N}_h^\partial$ and

$$|\partial u_h(x_i)| = f_i \qquad \forall\, x_i \in \mathcal{N}_h^0, \tag{1.2}$$

where $|\cdot|$ denotes the d-dimensional Lebesgue measure, $\partial u_h(x_i)$ is the subdifferential of u_h at x_i [see (2.10)], and $f_i = \int_\Omega f\phi_i$ is a suitable averaging of f against the canonical hat basis functions $\{\phi_i\}$ over $\mathcal{T}_h$ [43]. Our primary goal in this paper is to establish a bound for the error $u - u_h$ in the $L^\infty(\Omega_h)$-norm, which seems missing in the current literature for (1.2).

This endeavor entails developing suitable notions of stability, consistency, and monotonicity within the max-norm framework. We now outline the main ingredients. We first need to control the L^∞-norm of a function v by the size of its subdifferential. This is the celebrated *Alexandroff estimate* [30], whose discrete counterpart for a nodal

function v_h is established in [40]: if $\mathcal{C}_h^-(v_h)$ is the (lower) *contact set* of nodes, namely $x_i \in \mathcal{N}_h$ so that $v_h(x_i) = \Gamma(v_h)(x_i)$ with $\Gamma(v_h)$ the (lower) convex envelope of v_h, then

$$\max_{x_i \in \mathcal{N}_h} v_h^-(x_i) \leq C \left(\sum_{x_i \in \mathcal{C}_h^-(v_h)} |\partial v_h(x_i)| \right)^{1/d}, \tag{1.3}$$

where $v_h \geq 0$ on $\partial\Omega$ and $v_h^-(x) = \max\{-v_h(x), 0\}$ is the negative part of v_h. Note that the constant $C = C(d, \Omega)$ is proportional to the diameter of Ω and the nodal contact set is just a collection of nodes [40]. We also refer to [32–35] for discrete Alexandroff estimates similar to (1.3) and corresponding Alexandroff–Bakelman–Pucci maximum principles for general fully nonlinear elliptic problems.

Our concept of stability in the max-norm and second ingredient is the following *discrete continuous dependence* estimate, which is a refinement of (1.3) and is derived in Sect. 4. If v_h and w_h are two nodal functions over $\mathcal{N}_h$ and $v_h(x_i) \geq w_h(x_i)$ for all $x_i \in \mathcal{N}_h^\partial$, then

$$\max_{x_i \in \mathcal{N}_h} (v_h - w_h)^-(x_i) \leq C \left(\sum_{x_i \in \mathcal{C}_h^-(v_h - w_h)} \left(|\partial v_h(x_i)|^{1/d} - |\partial w_h(x_i)|^{1/d} \right)^d \right)^{1/d}.$$

where $C = C(d, \Omega)$. The proof of such an estimate relies on a combinination of two novel tools from analysis and geometry, namely the discrete Alexandroff estimate (1.3) and the Brunn–Minkowski inequality (see Lemma 4.2). The above estimate will be instrumental to compare the discrete solution u_h with the nodal function $N_h u$ associated with the exact solution u, which is defined as $N_h u(x_i) = u(x_i)$ for all $x_i \in \mathcal{N}_h$.

The third ingredient is *operator consistency*, which is a careful study of the discrepancy between u and $N_h u$ carried out in Sect. 5. We first show that the discrete operator is exact for convex quadratic polynomials p satisfying $0 < \lambda I \leq D^2 p \leq \Lambda I$, namely

$$|\partial N_h p(x_i)| = \int_\Omega \phi_i(x) \det D^2 p(x) dx$$

at interior nodes $x_i \in \mathcal{N}_h^0$ so that $\operatorname{dist}(x_i, \partial\Omega_h) \geq Rh$ with constant $R = R(\lambda, \Lambda)$, provided $\mathcal{N}_h^0$ is *translation invariant*. For $u \in C^{2,\alpha}(\overline{\Omega})$, we immediately deduce that the consistency error is of order $O(h^\alpha)$ for $0 < \alpha \leq 1$

$$\left| |\partial N_h u(x_i)| - \int_\Omega \phi_i(x) \det D^2 u(x) dx \right| \leq C h^\alpha |u|_{C^{2,\alpha}(\overline{B_i})} \int_\Omega \phi_i(x) dx,$$

where $B_i := B_{Rh}(x_i)$ is a ball centered at x_i with radius Rh. We can also measure the consistency error in Sobolev norms. If $u \in W_q^s(\Omega)$ with $\frac{d}{q} + 2 < s \leq 3$, then we exploit the Sobolev embedding $W_q^s(\Omega) \subset C^{2,\alpha}(\overline{\Omega})$ with $\alpha = s - 2 - d/q$ and thus replace $|u|_{C^{2,\alpha}(\overline{B_i})}$ by $|u|_{W_q^s(B_i)}$. Since the set of nodes $\mathcal{N}_h$ conforms to $\partial\Omega$, its

translation invariance structure breaks down for nodes close to $\partial\Omega$, and so do the consistency bounds which become of order $O(1)$. These nodes are handled differently via a discrete barrier argument discussed in Sect. 6.1.

Combining the consistency and stability estimates, together with the non-degeneracy assumption $f \geq \lambda_F > 0$, we derive the following pointwise convergence rate for $C^{2,\alpha}$-solutions in Sect. 6

$$\| u - \Gamma(u_h) \|_{L^\infty(\Omega_h)} \leq C h^\alpha,$$

where the constant $C = C(d, \Omega, \lambda, \Lambda, \lambda_F)\big(\| u \|_{C^{2,\alpha}(\overline{\Omega})} + |u|_{W^2_\infty(\Omega)}\big)$. We point out that the $C^{2,\alpha}$-regularity assumption of u is a consequence of suitable assumptions on f. In fact, if $0 < \lambda_F \leq f \leq \Lambda_F$, then $0 < \lambda \leq D^2 u \leq \Lambda$ for some constant λ, Λ, and if $f \in C^\alpha(\overline{\Omega})$ and Ω is of class $C^{2,\alpha}$, then $u \in C^{2,\alpha}(\overline{\Omega})$ [11], [30, Section 4.3]. We also stress that the discrete barrier argument for nodes close to $\partial\Omega$ is responsible for the semi-norm $|u|_{W^2_\infty(\Omega)}$ in the error estimate.

In addition, we prove in Sect. 6 the following pointwise error estimate for functions with W^s_q-regularity and $\frac{d}{q} + 2 < s \leq 3$

$$\| u - \Gamma(u_h) \|_{L^\infty(\Omega_h)} \leq C h^{s-2},$$

where the constant $C = C(d, \Omega, \lambda, \Lambda, \lambda_F)\big(|u|_{W^s_q(\Omega)} + |u|_{W^2_\infty(\Omega)}\big)$. This estimate shows that the discrete method (1.2) exhibits first order accuracy in the max-norm provided $u \in W^3_q(\Omega)$ with $q > d$. Since $u \in W^3_q(\Omega) \subset C^2(\overline{\Omega})$, the above error estimate is still in the realm of classical solutions. However, our results extend to solutions $u \in W^{2,\infty}(\overline{\Omega}) \backslash C^2(\overline{\Omega})$ whose Hessian $D^2 u$ is discontinuous across a set S of (box) dimension $n < d$:

$$\| u - \Gamma(u_h) \|_{L^\infty(\Omega_h)} \leq C h^{s-2} |u|_{W^s_q(\Omega \backslash S)} + C h^{\frac{d-n}{d}} |u|_{W^2_\infty(\Omega)}.$$

The rest of this paper is organized as follows. In Sect. 2 we discuss our notion of discrete convexity, a topic that has received considerable attention recently, and explore properties of subdifferentials. In Sect. 3, we introduce the Alexandroff and viscosity solution concepts for the Monge–Ampère equation (1.1) as well as the geometric method (1.2). We prove stability of (1.2) in Sect. 4 and consistency in Sect. 5. We conclude with three rates of convergence depending on solution regularity in Sect. 6.

2 Discrete convexity

Approximating convex solutions of the Monge–Ampère equation (1.1) entails two essential difficulties: dealing with discrete convexity and the fully nonlinear nature of (1.1). The former issue has been investigated in [1,13,16,36,42], and [16] shows that convex piecewise linear functions over a sequence of shape-regular meshes obtained by uniform refinement of a fixed mesh may not be dense in the set of convex functions. In fact, the Lagrange interpolant of a convex function may not even be convex. Several

notions of discrete convexity have been proposed in the literature: [13] deals with convex function interpolation on given meshes; [1] introduces finite element functions with positive weak Hessian; [42] imposes positive second order finite differences in all directions within a given stencil. In this paper, we deal with nodal functions and say they are convex if they admit a supporting hyperplane at every node. We make this explicit below.

2.1 Nodes and meshes

In contrast to mesh-based methods, the discretization of (1.1) hinges on a collection of nodes $\mathcal{N}_h := \mathcal{N}_h^0 \cup \mathcal{N}_h^\partial$ so that

$$\mathcal{N}_h^0 := \{x_i\}_{i=1}^n \subset \Omega, \qquad \mathcal{N}_h^\partial := \{x_i\}_{i=n+1}^N \subset \partial\Omega,$$

and a collection of simplices (or elements) T with nodes x_i which form a conforming mesh $\mathcal{T}_h$ of Ω and determine the computational domain $\Omega_h := \cup_{T\in\mathcal{T}_h} T$; since Ω is convex we infer that $\Omega_h \subset \Omega$. For each element T, we denote by h_T the diameter of T and by ρ_T the diameter of the largest inscribed ball in T. For each node x_i, we define the local spacing at x_i as

$$h_i := \max\{h_T : x_i \in T\}.$$

We say that the nodal set $\mathcal{N}_h$ is *quasi-uniform* if there exist constants $0 < \gamma \leq 1$ and $h > 0$ such that $\gamma h \leq h_i \leq h$ for all $1 \leq i \leq n$. We define the shape-regularity constant of an element T to be $\sigma_T := \frac{h_T}{\rho_T}$ and we say that $\mathcal{N}_h$ is *shape-regular* if there exists $\sigma > 0$ such that $\sigma_T \leq \sigma$ for all elements T. We will assume throughout that $\mathcal{N}_h$ is quasi-uniform and shape-regular.

We say that $\mathcal{N}_h^0$ is *translation invariant* if there is a basis $\{e_j\}_{j=1}^d$ of $\mathbb{R}^d$ with $|e_j| \leq 1$ for all $1 \leq j \leq d$ so that

$$\mathcal{N}_h^0 = \left\{ x_i = h\sum_{j=1}^d k_j e_j : k_j \in \mathbb{Z} \right\} \cap \Omega. \tag{2.1}$$

For the boundary nodes $\mathcal{N}_h^\partial$ we only require that

$$\partial\Omega \subset \cup_{x_i\in\mathcal{N}_h^\partial} B_{h/2}(x_i).$$

Obviously, a cartesian lattice with spacing $\sqrt{d}h$ is translation invariant and $\{e_j\}_{j=1}^d$ are the canonical unit vectors in $\mathbb{R}^d$.

We denote by $\{\phi_i\}_{i=1}^n$ the canonical basis of piecewise linear functions associated with $\mathcal{N}_h^0$ over $\mathcal{T}_h$. We say that $\{\phi_i\}_{i=1}^n$ is *translation invariant* provided that for all $x_i, x_j \in \mathcal{N}_h^0$ such that $\operatorname{dist}(x_i, \partial\Omega_h), \operatorname{dist}(x_j, \partial\Omega_h) > h$ and for all $x \in \mathbb{R}^d$ we have

$$\phi_i(x + x_i) = \phi_j(x + x_j). \tag{2.2}$$

If $\mathcal{N}_h^0$ is translation invariant, then so is $\{\phi_i\}_{i=1}^n$ for a suitable mesh $\mathcal{T}_h$. Since the construction of such $\mathcal{T}_h$ is obvious for $d = 2$, we now examine the case $d = 3$. Take a node $z \in \mathcal{N}_h^0$ with $\operatorname{dist}(z, \partial\Omega_h) > \sqrt{d}h$ and consider the box

$$Q = \Big\{ x \in \Omega : \ x = z + \sum_{j=1}^{d} t_j e_j, \ 0 \le t_j \le 1 \Big\}.$$

We would like to divide that box into a set of six disjoint tetrahedra $\{T_i\}_{i=1}^6$ such that $\cup_{i=1}^6 T_i = Q$. To do so, we label the eight nodes of Q as follows:

$$v_0 = z, \quad v_1 = z + e_3, \quad v_2 = z + e_1 + e_3, \quad v_4 = z + e_2, \quad v_3 = z + e_1,$$
$$v_5 = z + e_2 + e_3, \quad v_6 = z + e_1 + e_2 + e_3, \quad v_7 = z + e_1 + e_2.$$

Let $\{T_i\}_{i=1}^6$ be the convex hulls of the given nodes

$$T_1 = \operatorname{hull}\{v_0, v_1, v_3, v_4\}, \ T_2 = \operatorname{hull}\{v_1, v_2, v_4, v_5\}, \ T_3 = \operatorname{hull}\{v_1, v_2, v_3, v_4\},$$
$$T_4 = \operatorname{hull}\{v_6, v_2, v_5, v_7\}, \ T_5 = \operatorname{hull}\{v_7, v_4, v_3, v_2\}, \ T_6 = \operatorname{hull}\{v_7, v_4, v_5, v_2\}.$$

We realize that opposite faces are cut into two compatible triangles, e.g. faces $\operatorname{hull}\{v_0, v_1, v_2, v_3\}$ and $\operatorname{hull}\{v_4, v_5, v_6, v_7\}$ are cut along the segments $\operatorname{hull}\{v_1, v_3\}$ and $\operatorname{hull}\{v_5, v_7\}$. Finally, for every node $x_i \in \mathcal{N}_h^0$, we build $Q_i = Q - z + x_i$ and corresponding six tetrahedra, and observe that this construction yields a conforming mesh $\mathcal{T}_h$ satisfying (2.2).

2.2 Convexity of nodal functions

We say that the nodal function $u_h : \mathcal{N}_h \to \mathbb{R}$ is *convex* if for all $x_i \in \mathcal{N}_h^0$ there is a supporting hyperplane L of u_h, that is

$$L(x_j) \le u_h(x_j) \text{ for all } x_j \in \mathcal{N}_h \text{ and } L(x_i) = u_h(x_i).$$

We define the convex envelope of a nodal function u_h to be

$$\Gamma(u_h)(x) = \sup_{L \text{ affine}} \{L(x) : \ L(x_i) \le u_h(x_i) \quad \forall x_i \in \mathcal{N}_h\} \quad \forall x \in \Omega_h. \tag{2.3}$$

If the nodal function u_h is convex, then we have

$$u_h(x_i) = \Gamma(u_h)(x_i) \quad \text{for all } x_i \in \mathcal{N}_h^0. \tag{2.4}$$

We regard $\Gamma(u_h)$ as a natural convex extension of u_h and call it the *convex interpolant* of u_h. On the other hand, given a continuous function $u : \overline{\Omega} \to \mathbb{R}$ we denote by $N_h u : \mathcal{N}_h \to \mathbb{R}$ the nodal function associated with u:

$$N_h u(x_i) = u(x_i) \quad \forall\, x_i \in \mathcal{N}_h. \tag{2.5}$$

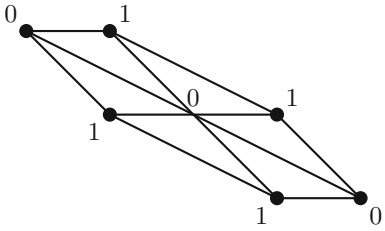

Fig. 1 Anisotropic star at $(0, 0)$ induced by the convex envelope $\Gamma(u_h)$ of the nodal function $u_h(x_i) = (x_i \cdot e)^2$ where $e = (1, 2)$. Note that $\Gamma(u_h)(x) = |x \cdot e|$ in the star

Since $\Gamma(u_h)$ is a piecewise linear function, it induces a mesh $\mathcal{T}_h$ which depends on u_h and is in general different from the original mesh satisfying (2.2); see Fig. 1. Given a convex nodal function u_h and a node $x_i \in \mathcal{N}_h^0$, we define the set $A_i(u_h)$ of *adjacent nodes* (or *adjacent set*) of x_i for u_h as a collection of nodes $x_j \in \mathcal{N}_h$ closest to x_i such that there exists a supporting hyperplane L of u_h at x_i and $L(x_j) = u_h(x_j)$. The set $A_i(u_h)$ is the collection of nodes in the star associated with x_i in the mesh $\mathcal{T}_h$ induced by $\Gamma(u_h)$. The following example illustrates that such a star could be quite elongated even for a cartesian lattice $\mathcal{N}_h^0$, especially if the Hessian D^2u is degenerate or nearly degenerate.

Example 2.1 (anisotropic star) Let $\Omega = \mathbb{R}^2$, $\mathcal{N}_h = \mathbb{Z}^2$ and $h = 1$. Let $u(x) = (x \cdot e)^2$ where $e = (1, m)$ for some integer $m \geq 1$ and $u_h(x_i) = u(x_i)$ for all $x_i \in \mathcal{N}_h$. Then the convex envelope $\Gamma(u_h)$ of u_h induces an anisotropic mesh $\mathcal{T}_h$; Fig. 1 displays the star associated to the origin for $m = 2$. The convex envelope $\Gamma(u_h)$ in such a star is $\Gamma(u_h)(x) = |x \cdot e|$.

Given a mesh $\mathcal{T}_h$ with nodes $\mathcal{N}_h$ we denote by $I_h(u_h)$ the Lagrange interpolant $I_h(u_h)$ of u_h over $\mathcal{T}_h$, namely the continuous piecewise linear function that interpolates the nodal values of u_h over $\mathcal{T}_h$. The following property is helpful to check discrete convexity: given a mesh $\mathcal{T}_h$ with nodes $\mathcal{N}_h$ and a nodal function u_h, $I_h(u_h)$ is convex if and only if it satisfies [40, Lemma 5.3]

$$[\![\nabla I_h(u_h)]\!]|_F \geq 0 \text{ for all faces } F \tag{2.6}$$

where $F = T^+ \cap T^-$ with $T^\pm \in \mathcal{T}_h$, the jump is given by

$$[\![\nabla I_h(u_h)]\!]|_F := -n_F^+ \cdot \nabla I_h(u_h)|_{T_+} - n_F^- \cdot \nabla I_h(u_h)|_{T_-} \tag{2.7}$$

and $n_F^\pm$ are the outer normal vectors of $T_\pm$ on face F. If $I_h(u_h)$ is convex, then $I_h(u_h) = \Gamma(u_h)$.

We finally let the (nodal lower) *contact set* of a nodal function u_h be

$$\mathcal{C}_h^-(u_h) := \{x_i \in \mathcal{N}_h^0 : \ \Gamma(u_h)(x_i) = u_h(x_i)\}. \tag{2.8}$$

Note that if u_h is convex, then $\mathcal{C}_h^-(u_h) = \mathcal{N}_h^0$; otherwise, $\mathcal{C}_h^-(u_h) \subset \mathcal{N}_h^0$.

2.3 Subdifferential

Let $u : \Omega \to \mathbb{R}$ be a convex function and $x_0 \in \Omega$. The subdifferential of u at x_0 is the set

$$\partial u(x_0) = \{v \in \mathbb{R}^d : u(x) \geq u(x_0) + v \cdot (x - x_0)\}. \tag{2.9}$$

Given a set $S \subset \Omega$, we define

$$\partial u(S) = \cup_{x \in S} \partial u(x).$$

Since u is a convex function, the subdifferential $\partial u(x)$ is non-empty and convex for all $x \in \Omega$.

Similarly, we define the subdifferential of nodal function u_h at node x_i by

$$\partial u_h(x_i) = \{v \in \mathbb{R}^d : \ u_h(x_j) \geq u_h(x_i) + v \cdot (x_j - x_i) \quad \forall x_j \in \mathcal{N}_h\}. \tag{2.10}$$

If the nodal function u_h is convex, then $\partial u_h(x_i)$ is non-empty for all $x_i \in \mathcal{N}_h^0$. The following lemma relates definitions (2.9) and (2.10).

Lemma 2.1 (discrete subdifferential) *If u_h is a convex nodal function, then $\partial u_h(x_i) = \partial \Gamma(u_h)(x_i)$ for all $x_i \in \mathcal{N}_h^0$.*

Proof Obviously, if $v \in \partial \Gamma(u_h)(x_i)$, that is,

$$\Gamma(u_h)(x) \geq \Gamma(u_h)(x_i) + v \cdot (x - x_i) \quad \forall x \in \Omega,$$

then $v \in \partial u_h(x_i)$ thanks to (2.4).

Conversely, let $\mathcal{T}_h$ be a mesh induced by the convex envelope $\Gamma(u_h)$, let $T \in \mathcal{T}_h$ be an element with vertices $\{x_j\}$. If $v \in \partial u_h(x_i)$, then $u_h(x_j) \geq u_h(x_i) + v \cdot (x_j - x_i)$ for all $x_j \in \mathcal{N}_h$, whence again thanks to (2.4), we have $\Gamma(u_h)(x_j) \geq \Gamma(u_h)(x_i) + v \cdot (x_j - x_i)$ for all vertices x_j of T. Since $\Gamma(u_h)$ is linear in T, we have $\Gamma(u_h)(x) \geq \Gamma(u_h)(x_i) + v \cdot (x - x_i)$ for any $x \in T$. This shows that $v \in \partial \Gamma(u_h)(x_i)$ as well. □

Lemma 2.2 (subdifferential monotonicity) *Let u_h and v_h be two convex nodal functions. If $u_h(x_i) \leq v_h(x_i)$ and $u_h(x_j) = v_h(x_j)$ for all $j \neq i$, then*

$$\partial v_h(x_i) \subset \partial u_h(x_i) \quad \textit{and} \quad \partial u_h(x_j) \subset \partial v_h(x_j) \textit{ for all } j \neq i.$$

Proof This follows directly from the definition of subdifferential (2.10). □

Given two compact sets A, B of $\mathbb{R}^d$, their *Minkowski sum* is given by

$$A + B := \{v + w \in \mathbb{R}^d : v \in A \text{ and } w \in B\}. \tag{2.11}$$

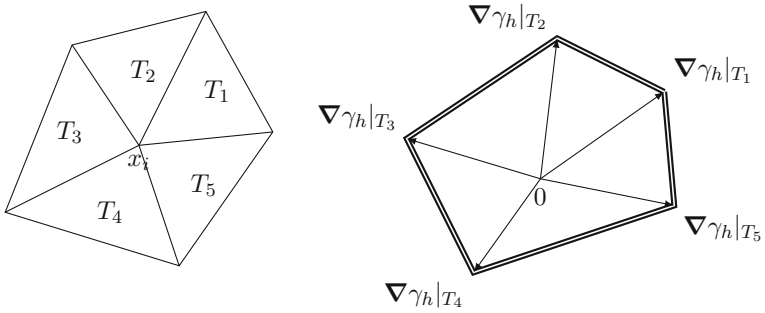

Fig. 2 Star centered at node x_i corresponding to the mesh $\mathcal{T}_h$ induced by the convex envelope $\gamma_h = \Gamma(u_h)$ of u_h and subdifferential $\partial u_h(x_i)$ of the convex nodal function u_h at node x_i such that $0 \in \partial u_h(x_i)$. The latter is the convex hull of the constant element gradients $\nabla\gamma_h|_{T_j}$ for $1 \le j \le 5$

Lemma 2.3 (addition of subdifferentials) *If u_h and v_h be two convex nodal functions, then*

$$\partial w_h(x_i) + \partial v_h(x_i) \subset \partial(w_h + v_h)(x_i) \text{ for all } x_i \in \mathcal{N}_h^0.$$

Proof We note that if $w \in \partial w_h(x_i)$ and $v \in \partial v_h(x_i)$, then

$$w_h(x_j) \ge w_h(x_i) + w \cdot (x_j - x_i) \quad \text{and} \quad v_h(x_j) \ge v_h(x_i) + v \cdot (x_j - x_i)$$

for all $x_j \in \mathcal{N}_h$. Adding both inequalities yields

$$w_h(x_j) + v_h(x_j) \ge w_h(x_i) + v_h(x_i) + (w + v) \cdot (x_j - x_i)$$

which implies $w + v \in \partial(w_h + v_h)$. □

Computing the subdifferential $\partial u_h(x_i)$ of a given convex nodal function u_h at $x_i \in \mathcal{N}_h^0$ is a nontrivial task because it is nonlocal. In fact, Lemma 2.1 shows that it involves computing the convex envelope $\Gamma(u_h)$ of u_h. The following lemma, proved in [40, Lemma 5.4], characterizes $\partial\Gamma(u_h)$.

Lemma 2.4 (characterization of subdifferential) *Let u_h be a convex nodal function and $\mathcal{T}_h$ be the mesh induced by its convex envelope $\Gamma(u_h)$. Then the subdifferential of u_h at $x_i \in \mathcal{N}_h^0$ is the convex hull of the constant gradients $\nabla\Gamma(u_h)|_T$ for all $T \in \mathcal{T}_h$ which contain x_i.*

Figure 2 depicts the subdifferential $\partial u_h(x_i)$ of the convex nodal function u_h at node x_i for $d = 2$. Computing $\partial u_h(x_i)$ is equivalent to finding a mesh $\mathcal{T}_h$ such that the jumps $[\![\nabla I_h(u_h)]\!]|_F$ of the Lagrange interpolant of $I_h(u_h)$ on faces F are nonnegative.

Unfortunately, the notion of discrete convexity is not sufficiently robust geometrically. Consider a pair of elements $T^{\pm} \in \mathcal{T}_h$ with common face $F = T^+ \cap T^-$, and assume $[\![\nabla I_h(u_h)]\!]|_F = 0$. Increasing the values of u_h at the nodes opposite to F increases the jump, and thus preserves convexity of $I_h(u_h)$ on $\mathcal{T}_h$, whereas decreasing the values of u_h violates convexity on $\mathcal{T}_h$.

Lemma 2.5 (geometric stability) *Let u_h be a nodal function. If the convex envelope $\Gamma(u_h)$ of u_h has positive jumps on all faces of the induced mesh $\mathcal{T}_h$, then $\mathcal{T}_h$ is invariant under small nodal perturbations of u_h.*

Proof Since the jumps $[\![\nabla \Gamma(u_h)]\!]|_F$ are continuous with respect to nodal variations of u_h, and they are positive, they remain positive under small nodal perturbations. We then apply (2.6) to deduce the assertion. □

3 Solution and approximation

There are two notions of weak solutions of (1.1): the Alexandroff solution hinges on a geometric interpretation of (1.1) and the viscosity solution relies on the comparison principle. We review these definitions now and discuss a geometric approximation of (1.1) due to Oliker and Prussner [43].

3.1 Alexandroff solution

To motivate this solution concept, suppose for the moment that $u \in C^2(\Omega)$ is a strictly convex function satisfying (1.1) in a classical sense. The convexity and C^2-regularity assumptions imply that $\partial u(x) = \nabla u(x)$ and that the subdifferential, viewed as a map $\partial u : \Omega \to \mathbb{R}^d$, is injective. Consequently, the change of variables $y = \nabla u(x)$ reveals that

$$\int_D f\,dx = \int_D \det D^2 u(x)\,dx = \int_{\partial u(D)} dy = |\partial u(D)|$$

for all Borel sets $D \subset \Omega$, where $\partial u(D) = \cup_{x\in D} \partial u(x)$ and $|\cdot|$ is the d-dimensional Lebesgue measure. Since ∂u is well-defined for non-smooth convex functions, the above identity allows one to widen the class of admissible solutions [30, Section 1.2].

Definition 3.1 (Monge–Ampère measure) We define the Monge–Ampère measure associated with a convex function $u \in C(\Omega)$ as

$$Mu(D) = |\partial u(D)|$$

for any Borel set D, where $|\cdot|$ denotes the d-dimensional Lebesgue measure.

Definition 3.2 (Alexandroff solution) Let μ be a Borel measure defined in Ω, an open and convex subset of $\mathbb{R}^d$. We say a convex function $u \in C(\Omega)$ is an Alexandroff solution to the Monge–Ampère equation

$$\det D^2 u = \mu$$

if the Monge–Ampère measure Mu associated with u equals μ.

The Alexandroff solution is closed under uniform convergence. This is stated in the lemma below and its proof is given in [30, Lemma 1.2.3].

Lemma 3.1 (weak convergence of Monge–Ampère measures) *If u_k are convex functions in Ω such that $u_k \to u$ as $k \to \infty$ uniformly on compact subsets of Ω, then the associated Monge–Ampère measures Mu_k tend to Mu weakly, that is*

$$\int_\Omega \phi(x) dMu_k(x) \to \int_\Omega \phi(x) dMu(x) \quad as\ k \to \infty,$$

for every ϕ continuous with compact support in Ω.

3.2 Viscosity solution

This solution concept hinges on the comparison principle.

Definition 3.3 (*viscosity solution*) Let $u \in C(\Omega)$ be a convex function and $f \in C(\Omega)$, $f \geq 0$. The function u is a viscosity sub-solution (super-solution) of the Monge–Ampère equation in Ω if whenever a convex function $\phi \in C^2(\Omega)$ and $x_0 \in \Omega$ are such that $(u - \phi)(x) \leq (\geq)(u - \phi)(x_0)$ for all x in a neighborhood of x_0, then we must have

$$\det D^2\phi(x_0) \geq (\leq) f(x_0).$$

If $u \in C(\Omega)$ is an Alexandroff solution with $Mu = f$ and $f \in C(\Omega)$, then u is a viscosity solution [30, Proposition 1.3.4]. Conversely, if $f \in C(\overline{\Omega})$ and $f > 0$, then the viscosity solution is also the Alexandroff solution [30, Proposition 1.7.1].

3.3 Examples of weak solutions

We show examples of Alexandroff and viscosity solutions, which are not classical solutions, namely $u \notin C^2(\overline{\Omega})$.

Example 3.1 (Alexandroff solution) Let $\Omega = B_1(0) \subset \mathbb{R}^2$ and

$$u(x) = |x| - 1.$$

The function $u(x)$ is an Alexandroff solution of the Monge–Ampère equation

$$\det D^2 u = \pi \delta_{(x=0)},$$

where $\delta_{(x=0)}$ is the Dirac measure at the origin; u is not a viscosity solution.

Example 3.2 (viscosity solution) Let $\Omega = B_2(0) \subset \mathbb{R}^2$ and [27]

$$u(x) = \begin{cases} \frac{1}{2}|x|^2 & \text{in } |x| \leq 1, \\ \frac{1}{2}|x|^2 + \frac{1}{2}\big(|x| - 1\big)^2 & \text{in } 1 \leq |x| \leq 2. \end{cases} \tag{3.1}$$

The function u is a viscosity solution of the Monge–Ampère equation

$$\det D^2 u(x) = \begin{cases} 1 & \text{in } |x| \le 1, \\ 4 - 2|x|^{-1} & \text{in } 1 \le |x| \le 2. \end{cases}$$

We note that $u \in C^{1,1}(\overline{\Omega}) \backslash C^2(\overline{\Omega})$ and the Hessian of $D^2 u$ exhibits a jump discontinuity across $\partial B_1(0)$.

3.4 Oliker–Prussner method

Following [43] we can approximate (1.1) exploiting its geometric interpretation. Given a quasi-uniform and shape regular nodal set $\mathcal{N}_h$, corresponding conforming mesh $\mathcal{T}_h$ and canonical basis functions $\{\phi_i\}_{i=1}^n$ associated with $\mathcal{N}_h^0$, for any function $f \ge 0$ we define the nodal function $f_h : \mathcal{N}_h \to \mathbb{R}$ to be

$$f_i := \int_{\omega_i} f(x)\phi_i(x)dx \quad \forall x_i \in \mathcal{N}_h^0, \tag{3.2}$$

with $\omega_i = \text{supp}\,(\phi_i)$ being the star corresponding to x_i.

The discretization of (1.1) reads as follows: we seek a convex nodal function u_h satisfying

$$|\partial u_h(x_i)| = f_i \quad \forall\, x_i \in \mathcal{N}_h^0. \tag{3.3}$$

We refer to [43] for a proof of existence and uniqueness of (3.3) for $d = 2$. Below we give a proof of existence for $d \ge 2$ that ties the concepts developed so far together. We observe that definition (3.2) is different from that in [43], which replaces the functions ϕ_i by characteristic functions of disjoint sets containing the nodes x_i, and that $\{\phi_i\}_{i=1}^n$ need not be translation invariant for u_h to exist. This property is instrumental later to derive consistency.

3.5 Existence: discrete Perron's method

We construct a monotone sequence $\{u_h^k\}_{k=0}^\infty$ of convex nodal functions, namely

$$u_h^{k+1}(x_i) \ge u_h^k(x_i) \quad \forall\, x_i \in \mathcal{N}_h^0,$$

which converges to a solution of (3.3) as $k \to \infty$. For each $k \ge 0$ there is a mesh $\mathcal{T}_h^k$ with nodes $\mathcal{N}_h$ but possibly different connectivity than $\mathcal{T}_h^j$ for $j < k$ and the property

$$I_h^k(u_h^k) = \Gamma(u_h^k).$$

Therefore, the interpolant $I_h^k(u_h^k)$ of u_h^k over $\mathcal{T}_h^k$ is convex. We illustrate how these meshes $\mathcal{T}_h^k$ change with the iteration counter k.

Construction of u_h. We first initialize the iteration. We assume that Ω is contained in a ball $B_R(0)$ of radius R centered at the origin and $d \geq 2$. We consider the quadratic polynomial $p(x) := \frac{\Lambda^{1/d}}{2}\left(|x|^2 - 2R^2\right)$ with $\Lambda > 0$ to be specified later. Then

$$\det D^2 p(x) = \Lambda \quad \forall\, x \in \Omega, \qquad p(x) \leq g(x) \quad \forall\, x \in \partial\Omega,$$

provided $-\frac{\Lambda^{1/d}}{2}R^2 \leq g(x)$ for all $x \in \partial\Omega$. We define $u_h^0 = N_h p$, namely

$$u_h^0(x_i) := p(x_i) \quad \forall\, x_i \in \mathcal{N}_h.$$

Let $\mathcal{T}_h^0$ be a Delaunay triangulation associated with $\mathcal{N}_h$, which exists according to [15, Theorem 2.3]. For such a mesh, the Lagrange interpolant of u_h^0 is convex [14,15,22], and thus coincides with the convex envelope of u_h^0:

$$I_h^0(u_h^0) = \Gamma(u_h^0).$$

We stress that $\mathcal{T}_h^0$ is in general different from the mesh $\mathcal{T}_h$ used to set (3.2) and (3.3). We assert that u_h^0 is a subsolution of (3.3). To see this we need to estimate the measure of the subdifferential $|\partial u_h^0(x_i)|$ at every node $x_i \in \mathcal{N}_h^0$. Since $\mathcal{T}_h^0$ is shape-regular, the star ω_i^0 of $\mathcal{T}_h^0$ around x_i has a diameter proportional to h and contains a ball $B_{\alpha h}(x_i)$ of radius αh with $\alpha > 0$ only dependent on shape regularity. Upon subtracting the affine function $L(x) = p(x_i) + \nabla p(x_i)\cdot(x - x_i)$ from p, we can assume that p and its gradient vanish at $x = x_i$ without changing $|\partial u_h^0(x_i)|$. Since $p(x) = \frac{\Lambda^{1/d}}{2}|x - x_i|^2$, we see that $p(x) \geq \frac{\Lambda^{1/d}}{2}\alpha^2 h^2$ on $\partial\omega_i^0$ and we can apply the discrete Alexandroff estimate (1.3) to $u_h^0 - \frac{\Lambda^{1/d}}{2}\alpha^2 h^2$ on ω_i^0 to deduce

$$|\partial u_h^0(x_i)|^{1/d} \geq Ch^{-1} \sup_{\omega_i}\left(u_h^0 - \frac{\Lambda^{1/d}}{2}\alpha^2 h^2\right)^- \geq C\Lambda^{1/d} h,$$

whence

$$|\partial u_h^0(x_i)| \geq C\Lambda h^d = C\Lambda|\omega_i^0|.$$

Recalling from (3.2) that $f_i = \int_{\omega_i} f\phi_i \leq \Lambda_F|\omega_i|$, and realizing that $|\omega_i|$ and $|\omega_i^0|$ are comparable, we can choose $\Lambda > 0$ sufficiently large so that

$$|\partial u_h^0(x_i)| \geq f_i \quad \forall x_i \in \mathcal{N}_h^0. \tag{3.4}$$

Moreover, in view of $u_h^0 \leq I_h^0 g$ on $\partial\Omega_h$ we infer that u_h^0 is a subsolution of (3.3), as asserted.

To construct the iterates u_h^{k+1} for $k \geq 0$ we recall Lemma 2.2 (subdifferential monotonicity): if the nodal value $u_h^k(x_i)$ increases, then $|\partial u_h^k(x_i)|$ decreases and $|\partial u_h^k(x_j)|$

increases for all other nodes x_j ($j \neq i$). With this in mind, we first set the boundary values one at a time

$$u_h^1(x_i) = g(x_i) \quad \forall\, x_i \in \mathcal{N}_h^\partial,$$

which preserves (3.4) and the convexity of u_h^1. If a convex nodal function u_h^k has been computed, we now describe how to construct u_h^{k+1} upon increasing the internal nodal values $u_h^k(x_i)$ one at a time. Having determined $u_h^{k+1}(x_j)$ for $j < i$, we thus define $u_h^{k+1}(x_i)$ to be the largest value so that

$$|\partial u_h^{k+1}(x_i)| = f_i \quad \forall\, x_i \in \mathcal{N}_h^0,$$

provided $u_h^{k+1}(x_j) = u_h^k(x_j)$ for $j > i$, which in turn implies for $j \neq i$

$$|\partial u_h^{k+1}(x_j)| \geq f_j \quad \forall\, x_j \in \mathcal{N}_h^0.$$

This is always doable because $|\partial u_h^{k+1}(x_i)|$ decreases continuously with increasing $u_h^{k+1}(x_i)$ until it vanishes, which corresponds to having a supporting hyperplane at x_i that touches u_h^{k+1} at $d+1$ nodes distinct from x_i and not lying in one hyperplane. In addition, since $f_i \geq 0$ for all $x_i \in \mathcal{N}_h^0$, the intermediate iterates leading to u_h^{k+1} are always convex by definition. This process creates a monotone sequence $\{u_h^k\}_{k\in\mathbb{N}}$ of convex nodal functions, namely $u_h^{k+1}(x_i) \geq u_h^k(x_i)$ for all $x_i \in \mathcal{N}_h^0$. Since this sequence has a uniform upper bound, namely $u_h^k(x_i) \leq \max_{x_j \in \mathcal{N}_h^\partial} g(x_j)$ for all $x_i \in \mathcal{N}_h^0$ and all k, as a consequence of Corollary 4.4 (maximum principle) below, we deduce that the sequence converges to a nodal function u_h. Next, we show that the limit u_h satisfies (3.3).

Taking $\phi(x)$ in Lemma 3.1 (weak convergence of Monge–Ampère measures) as the hat function ϕ_i associated with $x_i \in \mathcal{N}_h^0$ in definition (3.2), we deduce that $|\partial u_h^k(x_i)|$ converges and

$$\begin{aligned} |\partial u_h(x_i)| &= \int_\Omega \phi_i(x) dMu_h(x) \\ &= \lim_{k\to\infty} \int_\Omega \phi_i(x) dMu_h^k(x) = \lim_{k\to\infty} |\partial u_h^k(x_i)| \geq f_i, \end{aligned}$$

because Mu_h^k is a sum of Dirac measures with mass $|\partial u_h^k(x_i)|$ supported at $x_i \in \mathcal{N}_h^0$. We argue by contradiction. If u_h does not satisfy (3.3), then there exists a node $x_i \in \mathcal{N}_h^0$ such that $|\partial u_h(x_i)| > f_i$ and there exists $\delta > 0$ such that increasing $u_h(x_i)$ by δ, the subdifferential at x_i equals f_i. On the other hand, given any $\epsilon > 0$, there exists k_ϵ such that $0 \leq u_h(x_j) - u_h^k(x_j) \leq \epsilon$ for all $k \geq k_\epsilon$, $x_j \in \mathcal{N}_h^0$. We define the auxiliary function $\widetilde{u}_h^k(x_j) := u_h^{k+1}(x_j)$ if $j < i$, $\widetilde{u}_h^k(x_j) := u_h^k(x_j)$ if $j > i$ and $\widetilde{u}_h^k(x_i) := u_h^k(x_i) + \delta - \epsilon$. We note that

$$\widetilde{u}_h^k(x_j) + \epsilon \geq u_h(x_j) \quad j \neq i, \qquad \widetilde{u}_h^k(x_i) + \epsilon = u_h^k(x_i) + \delta \leq u_h(x_i) + \delta,$$

whence, applying Lemma 2.2 to $\widetilde{u}_h^k + \epsilon$ and u_h perturbed by δ at x_i yields

$$|\partial \widetilde{u}_h^k(x_i)| \geq f_i.$$

Therefore, since $u_h^{k+1}(x_i)$ is the largest value satisfying $|\partial u_h^{k+1}(x_i)| = f_i$ we deduce $u_h^{k+1}(x_i) \geq \widetilde{u}_h^k(x_i) \geq u_h(x_i) + \delta - 2\epsilon > u_h(x_i)$ provided that $\epsilon < \delta/2$. This contradicts the fact that $u_h^{k+1}(x_i) \leq u_h(x_i)$ and proves (3.3).

Computation of subdifferentials Computing $|\partial u_h^k(x_i)|$ is a key step of the algorithm, which reduces to computing the constant gradients $\nabla \Gamma(u_h^k)|_T$ of the convex envelope $\Gamma(u_h^k)$ of u_h^k for each element T of the induced mesh $\mathcal{T}_h^k$. Lemma 2.4 yields

$$|\partial u_h^k(x_i)| = \text{measure of polygon with vertices } \{\nabla \Gamma(u_h^k)|_T\} \text{ and } x_i \in T.$$

During the iteration the underlying mesh $\mathcal{T}_h^k$ changes, starting from the Delaunay mesh $\mathcal{T}_h^0$ for $\mathcal{N}_h$, first for $d = 2$ and next for $d = 3$. We describe now how these changes occur and can be implemented.

Case $d = 2$. We consider two triangles $T_1, T_2 \in \mathcal{T}_h^k$ with vertices z_1, z_2, z_4 and z_2, z_3, z_4, namely they are the convex hulls

$$T_1 = \text{hull}\{z_1, z_2, z_4\}, \quad T_2 = \text{hull}\{z_2, z_3, z_4\},$$

and

$$z_1 = (h, 0), \quad z_2 = (0, -h), \quad z_3 = (-h, 0), \quad z_4 = (0, h).$$

For $t \in [-1, 1]$, we consider the one-parameter convex nodal function

$$u_h^k(z_i) = 0 \quad i = 1, 2, 3, \qquad u_h^k(z_4) = t,$$

and observe that its Lagrange interpolant $I_h^k(u_h^k)$ is convex for $-1 \leq t \leq 0$ and concave for $0 \leq t \leq 1$; hence $\Gamma(u_h^k) = I_h^k(u_h^k)$ only for $-1 \leq t \leq 0$. The constant gradients in T_i are $t\nabla\phi_4$ for $i = 1, 2$ and the jump on the edge $F_1 = [z_2, z_4]$ is given by

$$[\![\nabla I_h^k(u_h^k)]\!]|_{F_1} = -\frac{t}{h}.$$

We see that it is non-negative only for $-1 \leq t \leq 0$, which means that $I_h^k(u_h^k)$ is convex according to (2.6). When $t = 0$ the function $I_h^k(u_h^k)$ is linear in $T_1 \cup T_2$ and for $0 \leq t \leq 1$ we have to flip the edge F_1 to $F_2 = [z_1, z_3]$ and consider a new mesh $\mathcal{T}_h^{k+1}$ upon replacing T_1, T_2 with the two new triangles

$$T_3 = \text{hull}\{z_1, z_2, z_3\}, \quad T_4 = \text{hull}\{z_1, z_3, z_4\},$$

for which $[\![\nabla I_h^{k+1}(u_h^k)]\!]|_{F_2} = \frac{t}{h} \geq 0$ and $I_h^{k+1}(u_h^k)$ is convex. This example reveals that the connectivity of the underlying mesh $\mathcal{T}_h^k$ may change as we increase nodal

values $u_h^k(x_i)$. Since increasing $u_h^k(x_i)$ is equivalent to adding a multiple of ϕ_i^k to u_h^k, we realize that changes are local and restricted to the star ω_i^k. They can be monitored on edges within ω_i^k by simply checking the signs of jumps and flipping edges accordingly. Jumps on the boundary edges of ω_i increase, which is consistent with Lemma 2.2, and require no attention whatsoever. We further see that the edge flipping process is similar to the construction of Delaunay meshes for $d = 2$.

Case $d = 3$. The change of mesh connectivity is more complicated for $d = 3$, but it is still local (within the star ω_i^k) and thus trackable [31]. To describe this process, we consider the following setting with five nodes

$$z_0 = (0, 0, -1),\ z_1 = (1, 0, 0),\ z_2 = (1, 1, 0),\ z_3 = (0, 1, 0),\ z_4 = (0, 0, 1),$$

and two configurations for the convex hull of $\{z_0, z_1, z_2, z_3, z_4\}$; see Fig. 3. The first configuration has two tetrahedra T_1, T_2 with one common face F

$$T_1 = \text{hull}\{z_0, z_1, z_2, z_4\} \quad T_2 = \text{hull}\{z_0, z_2, z_3, z_4\} \quad F = \text{hull}\{z_0, z_2, z_4\}.$$

The second configuration has three tetrahedra T_1^*, T_2^*, T_3^* and three common faces F_1^*, F_2^*, F_3^*

$$\begin{aligned} &T_1^* = \text{hull}\{z_4, z_1, z_2, z_3\},\ T_2^* = \text{hull}\{z_0, z_1, z_2, z_3\},\ T_3^* = \text{hull}\{z_0, z_1, z_3, z_4\},\\ &F_1^* = \text{hull}\{z_1, z_2, z_3\},\ F_2^* = \text{hull}\{z_0, z_1, z_3\},\ F_3^* = \text{hull}\{z_1, z_3, z_4\}, \end{aligned}$$

whence $F_1^* = T_1^* \cap T_2^*$, $F_2^* = T_2^* \cap T_3^*$, $F_3^* = T_3^* \cap T_1^*$. We will see that perturbing the values of a convex nodal function u_h^k, one configuration switches to the other to keep convexity of $I_h^k(u_h^k)$.

Case I. We first describe a transition from the first to the second configuration. Let u_h^k be the following nodal function

$$u_h^k(z_0) = 0, \quad u_h^k(z_1) = u_h^k(z_2) = u_h^k(z_3) = 1, \quad u_h^k(z_4) = 2 + t.$$

A simple computation yields

$$\nabla I_h^k(u_h^k)|_{T_1} = (0, 0, 1) + (-1, 0, 1)\frac{t}{2}, \quad \nabla I_h^k(u_h^k)|_{T_2} = (0, 0, 1) + (0, -1, 1)\frac{t}{2}.$$

Since the unit normal vector to F is $\boldsymbol{n}|_{T_1} = -\boldsymbol{n}|_{T_2} = \frac{\sqrt{2}}{2}(-1, 1, 0)$, the jump on F reads $[\![\nabla I_h^k(u_h^k)]\!]|_F = -\frac{\sqrt{2}}{2}t$ according to (2.7), and changes sign as t increases from -1 to 1. To preserve the convexity of $I_h^k(u_h^k)$ for $-1 \le t < 0$, we switch to the second configuration for $0 < t \le 1$. We thus get gradients

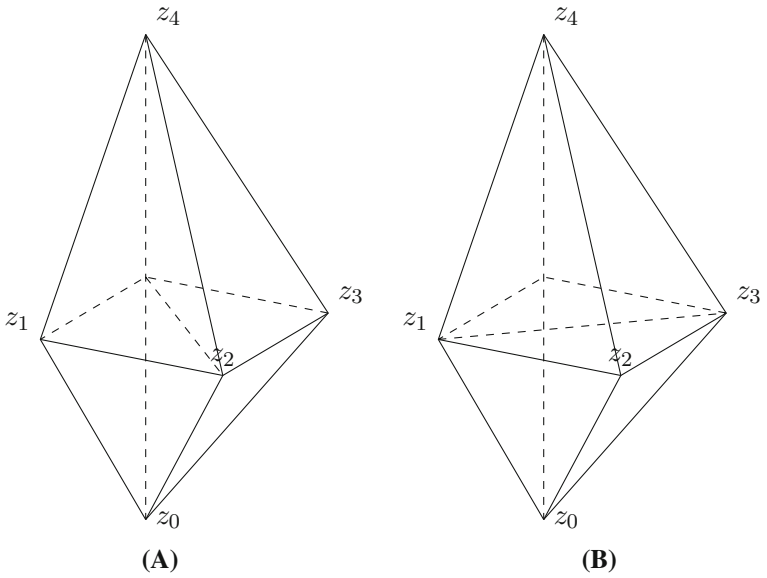

Fig. 3 Two conforming partitions of the convex hull of z_0, z_1, z_2, z_3, z_4: the first configuration contains two tetrahedra $T_1 = \text{hull}\{z_0, z_1, z_2, z_4\}$ and $T_2 = \text{hull}\{z_0, z_2, z_3, z_4\}$, whereas the second one consists of three tetrahedra $T_1^* = \text{hull}\{z_4, z_1, z_2, z_3\}$, $T_2^* = \text{hull}\{z_0, z_1, z_2, z_3\}$, and $T_3^* = \text{hull}\{z_0, z_1, z_3, z_4\}$

$$\begin{aligned}
\nabla I_h^k(u_h^k)|_{T_1^*} &= (0,0,1) + (0,0,1)t,\\
\nabla I_h^k(u_h^k)|_{T_2^*} &= (0,0,1),\\
\nabla I_h^k(u_h^k)|_{T_3^*} &= (0,0,1) + (-1,-1,1)\frac{t}{2},
\end{aligned}$$

unit normals

$$\boldsymbol{n}_1|_{F_1^*} = (0,0,-1),\ \boldsymbol{n}_2|_{F_2^*} = \frac{1}{\sqrt{3}}(-1,-1,1),\ \boldsymbol{n}_3|_{F_3^*} = \frac{1}{\sqrt{3}}(1,1,1),$$

and jumps

$$[\![\nabla I_h^k(u_h^k)]\!]|_{F_1^*} = t, \quad [\![\nabla I_h^k(u_h^k)]\!]|_{F_2^*} = \frac{\sqrt{3}}{2}t, \quad [\![\nabla I_h^k(u_h^k)]\!]|_{F_3^*} = \frac{\sqrt{3}}{2}t.$$

Therefore, the function $I_h^k(u_h^k)$ is convex in $T_1 \cup T_2$ for $t < 0$ and $T_1^* \cup T_2^* \cup T_3^*$ for $t > 0$. The function $I_h^k(u_h^k)$ is linear for $t = 0$.

Case II. We next describe a transition from the second to the first configuration upon increasing one nodal value. If

$$u_h(z_0) = 0, \quad u_h(z_1) = u_h(z_2) = 2, \quad u_h(z_3) = 2 + t, \quad u_h(z_4) = 4,$$

then it is easy to check that

$$\begin{aligned}\nabla I_h^k(u_h^k)|_{T_1^*} &= (0,0,2)+(-1,0,-1)t\\ \nabla I_h^k(u_h^k)|_{T_2^*} &= (0,0,2)+(-1,0,1)t\\ \nabla I_h^k(u_h^k)|_{T_3^*} &= (0,0,2)+(0,1,0)t\end{aligned}$$

and

$$[\![\nabla I_h^k(u_h^k)]\!]|_{F_1^*} = -2t,\ \ [\![\nabla I_h^k(u_h^k)]\!]|_{F_2^*} = -\sqrt{3}t,\ \ [\![\nabla I_h^k(u_h^k)]\!]|_{F_3^*} = -\sqrt{3}t.$$

Therefore, $I_h^k(u_h^k)$ is convex in $T_1^* \cup T_2^* \cup T_3^*$ for $t < 0$ but not for $t > 0$. On the other hand, we see that

$$\nabla I_h^k(u_h^k)|_{T_1} = (0,0,2),\quad \nabla I_h^k(u_h^k)|_{T_2} = (0,0,2)+(-1,1,0)t,$$

and the jump of the interelement face F is

$$[\![\nabla I_h^k(u_h^k)]\!]|_F = \sqrt{2}t,$$

whence $I_h^k(u_h^k)$ is convex in $T_1 \cup T_2$ for $t > 0$. This concludes the discussion.

In the rest of the paper, we focus on estimating the rates of convergence. Such an estimate relies on the stability and consistency of (3.3), which we address in Sects. 4 and 5, respectively.

4 Stability

Given two nodal functions v_h and w_h we control the L^∞-norm of $(v_h - w_h)^-$ in terms of the discrepancy of their subdifferential measures. This is the content of Proposition 4.3 and its proof is the chief goal of this section. This result hinges on two important estimates, the discrete Alexandroff estimate and the Brunn–Minkowski inequality, which we discuss next.

4.1 Discrete Alexandroff estimate

The Alexandroff estimate for a continuous piecewise affine function v_h states that the L^∞-norm of v_h is controlled by the Lebesgue measure of its subdifferential. We refer to [40] for a complete proof, and also to [32–35] for similar estimates and for discrete Alexandroff Bakelman Pucci estimates for general fully nonlinear elliptic problems.

Lemma 4.1 (discrete Alexandroff estimate) *Let v_h be a nodal function and $v_h(x_i) \geq 0$ at all $x_j \in \mathcal{N}_h^\partial$. Then*

$$\sup_\Omega v_h^- \leq C \left(\sum_{x_i \in \mathcal{C}_h^-(v_h)} |\partial v_h(x_i)| \right)^{1/d}, \tag{4.1}$$

where $C = C(d, \Omega)$ is proportional to the diameter of Ω and $\mathcal{C}_h^-(v_h)$ is the contact set defined in (2.8).

4.2 Brunn–Minkowski inequality

The second main tool to prove Proposition 4.3 is the celebrated Brunn–Minkowski inequality [29]. This inequality relates the Lebesgue measures of compact subsets A, B of Euclidean space $\mathbb{R}^d$ with that of their Minkowski sum $A + B$ defined in (2.11).

Lemma 4.2 (Brunn–Minkowski inequality) *Let A and B be two nonempty compact subsets of $\mathbb{R}^d$ for $d \geq 1$. Then the following inequality holds:*

$$|A + B|^{1/d} \geq |A|^{1/d} + |B|^{1/d}.$$

Since $(a + b)^t \leq a^t + b^t$ for $a, b \geq 0$ and $0 < t < 1$, we deduce the following immediate consequence of Lemma 4.2

$$|A + B| \geq |A| + |B|. \tag{4.2}$$

4.3 Continuous dependence

We now compare two arbitrary nodal functions in terms of their subdifferentials. This is instrumental for the error analysis.

Proposition 4.3 (continuous dependence) *Let v_h and w_h be two nodal functions associated with nodes $\mathcal{N}_h$ and $v_h \geq w_h$ at all $x_i \in \mathcal{N}_h^\partial$. Then*

$$\sup(v_h - w_h)^- \leq C \left(\sum_{x_i \in \mathcal{C}_h^-(v_h - w_h)} \left(|\partial v_h(x_i)|^{1/d} - |\partial w_h(x_i)|^{1/d} \right)^d \right)^{1/d},$$

where $C = C(d, \Omega)$ is proportional to the diameter of Ω.

Proof Let v_h, w_h be two arbitrary nodal functions. We consider the convex envelope $\Gamma(v_h - w_h)$ defined in (2.3) and the nodal contact set $\mathcal{C}_h^-(v_h - w_h)$ defined in (2.8).

Lemma 4.1 (discrete Alexandroff estimate) yields

$$\sup_{\Omega}(v_h - w_h)^- \le C \left(\sum_{x_i \in \mathcal{C}_h^-(v_h - w_h)} |\partial\Gamma(v_h - w_h)(x_i)| \right)^{1/d}, \tag{4.3}$$

whence we only need to estimate $|\partial\Gamma(v_h - w_h)(x_i)|$ for all $x_i \in \mathcal{C}_h^-(v_h - w_h)$. For these nodes, we easily see that

$$\partial\Gamma(v_h - w_h)(x_i) \subset \partial(v_h - w_h)(x_i).$$

Consequently, Lemma 2.3 (addition of subdifferentials) gives

$$\partial w_h(x_i) + \partial\Gamma(v_h - w_h)(x_i) \subset \partial v_h(x_i) \quad \forall\, x_i \in \mathcal{C}_h^-(v_h - w_h).$$

Applying Lemma 4.2 (Brunn–Minkowski inequality), we obtain

$$\begin{aligned} &|\partial w_h(x_i)|^{1/d} + |\partial\Gamma(v_h - w_h)(x_i)|^{1/d} \\ &\qquad \le |\partial w_h(x_i) + \partial\Gamma(v_h - w_h)(x_i)|^{1/d} \le |\partial v_h(x_i)|^{1/d}, \end{aligned}$$

whence

$$|\partial\Gamma(v_h - w_h)(x_i)| \le \left(|\partial v_h(x_i)|^{1/d} - |\partial w_h(x_i)|^{1/d} \right)^d.$$

This inequality gives us the desired estimate for $|\partial\Gamma(v_h - w_h)(x_i)|$. In view of (4.3), adding over all $x_i \in \mathcal{C}_h^-(v_h - w_h)$ concludes the proof. □

A direct consequence of this stability result is the maximum principle for nodal functions, which we state next.

Corollary 4.4 (discrete maximum principle) *Let v_h and w_h be two nodal functions associated with nodes $\mathcal{N}_h$. If $v_h(x_i) \ge w_h(x_i)$ at all $x_i \in \mathcal{N}_h^\partial$ and $|\partial v_h(x_i)| \le |\partial w_h(x_i)|$ at all $x_i \in \mathcal{N}_h^0$, then*

$$w_h(x_i) \le v_h(x_i) \quad \forall x_i \in \mathcal{N}_h.$$

Proof For any node $x_i \in \mathcal{C}_h^-(v_h - w_h)$, we have

$$\partial w_h(x_i) \subset \partial v_h(x_i).$$

Since $|\partial v_h(x_i)| \le |\partial w_h(x_i)|$ for all $x_i \in \mathcal{N}_h^0$, we deduce $|\partial v_h(x_i)| = |\partial w_h(x_i)|$ for all $x_i \in \mathcal{C}_h^-(v_h - w_h)$. Consequently, Proposition 4.3 (continuous dependence) implies

$$\sup(v_h - w_h)^- = 0,$$

whence $v_h - w_h \ge 0$. This completes the proof. □

Remark 4.1 (uniqueness) If two nodal functions u_h and w_h are both solutions of (3.3), then by Corollary 4.4 (discrete maximum principle),

$$\sup(u_h - w_h)^- = 0 \quad \text{and} \quad \sup(w_h - u_h)^- = 0.$$

Hence, we infer that $u_h = w_h$ at all nodes. This shows uniqueness.

5 Consistency

In this section, we examine the consistency of the Oliker–Prussner method (3.3). In general, this method is consistent in the sense that the right hand side of the (3.3) can be written equivalently as $\sum_{x_i \in \mathcal{N}_h} f_i \delta_{x_i}$ and this converges to f in measure. However, such a concept of convergence is too weak to derive rates of convergence. Fortunately, we realize that if internal nodes are translation invariant, then a reasonable notion of operator consistency holds for any convex quadratic polynomial; see Lemma 5.3. Such property is shown in [4,37] for Cartesian nodes. In contrast, we give here an alternative proof of consistency based on the geometric interpretation of subdifferentials of convex quadratic polynomials in the interior of the domain, extend the results to $C^{2,\alpha}$ and W^s_p functions, and further investigate the consistency error in the region close to the boundary.

Lemma 2.4 (characterization of subdifferential) states that the subdifferential of a convex nodal function p_h at node $x_i \in \mathcal{N}_h^0$ is the convex hull of piecewise constant gradients of its convex envelope $\Gamma(p_h)$, which in turn is determined by the nodal values $p_h(x_j)$ in the adjacent set $A_i(p_h)$ of x_i for p_h. The following lemma gives an estimate of the size of $A_i(p_h)$.

Lemma 5.1 (size of adjacent sets) *Let the nodal set $\mathcal{N}_h$ be quasi-uniform and shape-regular with constant σ. Let p be a C^2 convex function defined in Ω. If $\lambda I \le D^2 p \le \Lambda I$ in Ω for some constants $\lambda, \Lambda > 0$ and $p_h := N_h p$ is the nodal function associated with p defined in (2.5), then the adjacent set of nodes $A_i(p_h)$ at x_i for p_h satisfies*

$$A_i(p_h) \subset B_{Rh}(x_i)$$

where $R = \frac{\Lambda}{\lambda}\sigma^2$ and $B_{Rh}(x_i)$ is the ball centered at x_i with radius Rh.

Proof Let z be an adjacent node of x_i such that

$$|z - x_i| = \max\{|x_j - x_i| : x_j \in A_i(p_h)\}.$$

Without loss of generality, we may assume that $x_i = 0$, $p(x_i) = 0$ and $\nabla p(x_i) = 0$ and set $x_0 = x_i$. Let ω_0 be a star at x_0 in mesh $\mathcal{T}_h$ associated with nodal set $\mathcal{N}_h$. If $z \in \omega_0$, then the assertion is trivial because $R \ge 1$.

If $z \notin \omega_0$, then we may assume that there is a constant $R \ge 1$ such that $R^{-1}z \in T$ for some element $T \subset \omega_0$, which implies that $|z| \le Rh_T$ and $R^{-1}|z| \ge \rho_T$. If $\{x_k\}_{k=0}^d$ are the vertices of simplex T, then we write

$$z = R\sum_{k=0}^{d}\alpha_k x_k, \qquad \alpha_k \geq 0, \qquad \sum_{k=0}^{d}\alpha_k = 1.$$

We next note that $p(x_k) \leq \frac{1}{2}\Lambda h_T^2$ for all $1 \leq k \leq d$ because $D^2 p \leq \Lambda I$ and $|x_k| \leq h_T$. Since $z \in A_i(p_h)$, there exists a supporting hyperplane L at x_0 such that

$$L(z) = p_h(z), \qquad L(x_k) \leq p_h(x_k) \leq \frac{1}{2}\Lambda h_T^2.$$

Exploiting that L is linear yields

$$p_h(z) = R\sum_{k=0}^{d}\alpha_k L(x_k) \leq \frac{1}{2}\Lambda h_T^2 R$$

On the other hand, since $D^2 p \geq \lambda I$ and $|z| \geq R\rho_T = R h_T \sigma_T^{-1}$, we have

$$p_h(z) = p(z) \geq \frac{\lambda}{2}|z|^2 \geq \frac{\lambda}{2}R^2\sigma_T^{-2}h_T^2.$$

Combining the last two inequalities implies

$$R \leq \frac{\Lambda}{\lambda}\sigma_T^2 \leq \frac{\Lambda}{\lambda}\sigma^2.$$

This completes the proof. □

Lemma 5.1 (size of adjacent sets) shows that $\partial u_h(x_i)$ does not depend on values of u_h on the boundary $\partial\Omega$ for nodes x_i such that $\mathrm{dist}(x_i, \partial\Omega_h) \geq Rh$. We now consider nodes which are Rh away from $\partial\Omega_h$ and gather several properties of subdifferentials of convex quadratic polynomials.

Lemma 5.2 (properties of convex quadratic polynomials) *Let p be a convex quadratic polynomial such that $\lambda I \leq D^2 p \leq \Lambda I$ and $p_h := N_h p$ be the nodal function defined in (2.5). If $R = \frac{\Lambda}{\lambda}\sigma^2$, then the following properties hold:*

- *The subdifferential $\partial p_h(x_i)$ is a non-empty set for all $x_i \in \mathcal{N}_h$.*
- *If the nodal set $\mathcal{N}_h^0$ is translation invariant and $\mathrm{dist}(x_i, \partial\Omega_h) \geq Rh$ for $x_i \in \mathcal{N}_h^0$, then a uniform refinement $\mathcal{N}_{h/2}$ of $\mathcal{N}_h$ satisfies*

$$|\partial p_h(x_i)| = 2^d |\partial p_{\frac{h}{2}}(x_i)|.$$

- *If the nodal set $\mathcal{N}_h^0$ is translation invariant and $\mathrm{dist}(x_i, \partial\Omega_h) \geq Rh$ and $\mathrm{dist}(x_j, \partial\Omega_h) \geq Rh$ for $x_i, x_j \in \mathcal{N}_h^0$, then $A_i(p_h) = (x_i - x_j) + A_j(p_h)$ and*

$$|\partial p_h(x_i)| = |\partial p_h(x_j)|.$$

Proof Take $x_i = 0$ for simplicity. To prove the first assertion, we just observe that if L is the tangent plane touching p from below at 0, then L is a lower supporting hyperplane of p_h at 0. This implies that $\nabla L = \nabla p(0)$ is in the subdifferential of p_h at 0.

To prove the second assertion, we consider the auxiliary polynomial

$$q^i(x) := p(x) - \nabla p(0) \cdot x - p(0),$$

obtained by subtracting the tangent plane of p at 0. Since adding an affine function does not change the measure of the subdifferential, we have $|\partial p_h(0)| = |\partial q_h^i(0)|$. Using that q^i is homogeneous of degree 2 yields

$$q^i(x) = 4q^i\Big(\frac{x}{2}\Big).$$

Since

$$4q^i_{\frac{h}{2}}\Big(\frac{x_j}{2}\Big) = 4q^i\Big(\frac{x_j}{2}\Big) = q^i(x_j) = q_h^i(x_j) \geq v \cdot x_j = 2v \cdot \frac{x_j}{2}$$

for all $x_j \in \mathcal{N}_h^0$ and $v \in \partial q_h^i(0)$, we deduce

$$\partial q_h^i(0) = 2\partial q^i_{\frac{h}{2}}(0),$$

whence $|\partial q_h^i(0)| = 2^d |\partial q^i_{\frac{h}{2}}(0)|$ and the second assertion follows.

To prove the third assertion, we write $q^i(x) = (x - x_i)^t Q(x - x_i)$ for a suitable positive definite constant matrix Q. This implies

$$q^i(x) = q^j(x + x_j - x_i) \qquad \forall\, x \in \mathbb{R}^d$$

along with

$$A_i(q_h^i) = (x_i - x_j) + A_j(q_h^j), \qquad |\partial q_h^i(x_i)| = |\partial q_h^j(x_j)|.$$

Since $A_i(p_h) = A_i(q_h^i)$ and $\partial p_h(x_i) = \partial q_h^i(x_i)$, this concludes the proof. □

Now we are ready to prove the consistency of (3.3).

Lemma 5.3 (consistency of the discrete Monge–Ampère measure) *Let p be a convex quadratic polynomial such that $\lambda I \leq D^2 p \leq \Lambda I$ and $p_h := N_h p$ be the corresponding convex nodal function defined in (2.5). Let $\mathcal{N}_h^0$ be translation invariant and $\mathcal{T}_h$ be a mesh with nodes $\mathcal{N}_h$ and translation invariant basis of piecewise linear functions $\{\phi_i\}_{i=1}^n$. Then*

$$|\partial p_h(x_i)| = \int_\Omega \phi_i(x) \det D^2 p(x) dx$$

for any point $x_i \in \mathcal{N}_h^0$ *such that* $\operatorname{dist}(x_i, \partial\Omega_h) \geq Rh$ *and* $R = \frac{\Lambda}{\lambda}\sigma^2$.

Proof We consider a sequence of uniform refinements $\mathcal{T}_k$ of $\mathcal{T}_h$ and corresponding nodes $\mathcal{N}_k$ with $h_k = 2^{-k}h$ for $k \geq 1$. Let $p_k = N_{h_k} p$ be the nodal function of p associated with the set $\mathcal{N}_k$ and let $\gamma_k = \Gamma(p_k)$ be its convex envelope. Since ϕ_i is compactly supported in $\Omega_h \subset \Omega$ and $\gamma_k \to p$ uniformly as $k \to \infty$, Lemma 3.1 (weak convergence of Monge–Ampère measures) yields

$$\int_\Omega \phi_i(x) dM\gamma_k(x) \to \int_\Omega \phi_i(x) dMp(x) \quad \text{as } k \to \infty,$$

or equivalently

$$\sum_{x_j \in \mathcal{N}_k} \phi_i(x_j) |\partial p_k(x_j)| \to \int_\Omega \phi_i(x) \det D^2 p(x) dx \quad \text{as } k \to \infty.$$

Therefore, we only need to prove

$$\sum_{x_j \in \mathcal{N}_k} \phi_i(x_j) |\partial p_k(x_j)| = |\partial p_h(x_i)|,$$

which is independent of k. Since $\int_\Omega \phi_i^k = 2^{-kd} \int_\Omega \phi_i$ and $\operatorname{dist}(x_i, \partial\Omega_h) \geq Rh$, Lemma 5.2 (properties of convex quadratic polynomials) leads to

$$|\partial p_k(x_j)| = |\partial p_k(x_i)| = |\partial p_h(x_i)| \frac{\int_\Omega \phi_i^k}{\int_\Omega \phi_i} \qquad \forall\, k \geq 1,$$

where $\{\phi_j^k\}$ is the basis of hat functions over $\mathcal{T}_k$. Consequently, we obtain

$$\sum_{x_j \in \mathcal{N}_k} \phi_i(x_j) |\partial p_k(x_j)| = \frac{|\partial p_h(x_i)|}{\int_\Omega \phi_i} \sum_{x_j \in \mathcal{N}_k} \phi_i(x_j) \int_\Omega \phi_j^k = |\partial p_h(x_i)|$$

because $\int_\Omega \phi_i^k = \int_\Omega \phi_j^k$ according to (2.2) and $\sum_{x_j \in \mathcal{N}_k} \phi_i(x_j) \phi_j^k = \phi_i$, or equivalently $\phi_i \in \operatorname{span}\{\phi_j^k\}_{x_j \in \mathcal{N}_k}$. This completes the proof. □

Since $M = \det D^2 p(x)$ is constant for all $x \in \Omega$, Lemma 5.3 (consistency of the discrete Monge–Ampère measure) implies that $\int_\Omega \phi_i(x) dx = M^{-1} |\partial p_h(x_i)|$ for $x_i \in \mathcal{N}_h^0$ is independent of the mesh $\mathcal{T}_h$ supporting the translation invariant basis $\{\phi_i\}_{i=1}^n$, and so is the notion of consistency in Lemma 5.3. This is critical because both $\mathcal{T}_h$ and $\{\phi_i\}_{i=1}^n$ might not be unique. The following two local and quantitative consistency estimates of Proposition 5.4 are a consequence of Lemma 5.3 for nodes away from $\partial\Omega$.

Proposition 5.4 (interior consistency) *Let $\mathcal{N}_h^0$ be a translation invariant set of nodes, $\mathcal{T}_h$ be a mesh with nodes $\mathcal{N}_h$ and translation invariant basis of piecewise linear functions $\{\phi_i\}$. Let $u \in C^{2,\alpha}(\overline{B_i})$ be a convex function so that $\lambda I \le D^2 u \le \Lambda I$ in the ball $B_i := B_{Rh}(x_i)$ centered at node $x_i \in \mathcal{N}_h^0$ and radius Rh with $R = \frac{\Lambda}{\lambda}\sigma^2$. If $x_i \in \mathcal{N}_h^0$ satisfies* $\operatorname{dist}(x_i, \partial\Omega_h) \ge Rh$, *then*

$$\left| |\partial N_h u(x_i)| - \int_\Omega \phi_i(x) \det D^2 u(x) dx \right| \le Ch^\alpha |u|_{C^{2,\alpha}(\overline{B_i})} \int_\Omega \phi_i(x) dx,$$

where $C = C(d, \lambda, \Lambda)$ and $N_h u$ denotes the convex nodal function associated with u defined in (2.5). If instead $u \in W_q^s(B_i)$ with $s - \frac{d}{q} > 2$, $s \le 3$, then there is again $C = C(d, \lambda, \Lambda)$ such that

$$\left| |\partial N_h u(x_i)| - \int_\Omega \phi_i(x) \det D^2 u(x) dx \right| \le Ch^{s-2-\frac{d}{q}} |u|_{W_q^s(B_i)} \int_\Omega \phi_i(x) dx.$$

Proof To prove the first estimate we only need to show the inequality

$$|\partial N_h u(x_i)| \le \int_\Omega \phi_i(x) \det D^2 u(x) dx + Ch^\alpha |u|_{C^{2,\alpha}(\overline{B_i})} \int_\Omega \phi_i(x) dx,$$

because the reverse inequality can be derived similarly.

Since $u \in C^{2,\alpha}(\overline{B_i})$, we estimate u by a quadratic polynomial p so that

$$u(x) \le p(x) \quad \forall x \in B_{Rh}(x_i),$$

where $p(x_i) = u(x_i)$, $\nabla p(x_i) = \nabla u(x_i)$ and $D^2 p = D^2 u(x_i) + Ch^\alpha |u|_{C^{2,\alpha}(\overline{B_i})} I$ with universal constant C. If $p_h = N_h p$, then Lemma 2.2 (subdifferential monotonicity) yields

$$|\partial N_h u(x_i)| \le |\partial p_h(x_i)|.$$

If ϕ_i is the hat function over $\mathcal{T}_h$ associated with x_i, it remains to show

$$|\partial p_h(x_i)| \le \int_\Omega \phi_i(x) \det D^2 u(x) dx + Ch^\alpha |u|_{C^{2,\alpha}(\overline{B_i})} \int_\Omega \phi_i(x) dx.$$

Since $(\lambda + Ch^\alpha) I \le D^2 p \le (\Lambda + Ch^\alpha) I$ and

$$\frac{\Lambda + Ch^\alpha}{\lambda + Ch^\alpha} \le \frac{\Lambda}{\lambda} \quad \text{because } \Lambda \ge \lambda,$$

invoking Lemma 5.3 (consistency of the discrete Monge–Ampère measure), we obtain

$$|\partial p_h(x_i)| = \int_\Omega \phi_i(x) \det D^2 p(x) dx$$

because this holds for any mesh $\mathcal{T}_h$ with nodes $\mathcal{N}_h$ and translation invariant basis $\{\phi_i\}_{i=1}^n$ provided $\operatorname{dist}(x_i, \partial\Omega_h) \geq Rh$. Recalling that $u \in C^{2,\alpha}(\overline{B_i})$, we can write $D^2 p = D^2 u(x) + E(x)$ for all $x \in \overline{B_i}$, where $|E(x)| \leq C|u|_{C^{2,\alpha}(\overline{B_i})} h^\alpha$. Writing $\det D^2 p = \det D^2 u(x) \det(I + E(x) D^2 u(x)^{-1})$ and using Taylor expansion yields

$$|\partial p_h(x_i)| \leq \int_\Omega \phi_i(x) \det D^2 u(x) dx + Ch^\alpha |u|_{C^{2,\alpha}(\overline{B_i})} \int_\Omega \phi_i(x) dx,$$

and concludes the proof of the Hölder estimate. Finally, if $u \in W^s_q(B_i)$ with $s - \frac{d}{q} > 2$, then we resort to the Sobolev embedding $W^s_q(B_i) \subset C^{2,\alpha}(\overline{B_i})$ with $0 < \alpha = s - 2 - d/q < 1$ and apply the preceding Hölder estimate. □

For nodes close to the boundary, we can no longer exploit the node translation invariance and we thus get an error of order 1. We express this fact as follows.

Lemma 5.5 (boundary consistency) *Let $\mathcal{T}_h$ be a mesh with nodes $\mathcal{N}_h$ and $\{\phi_i\}$ be a basis of piecewise linear hat functions over $\mathcal{T}_h$. Let $u \in W^2_\infty(B_i)$ be a convex function with $\lambda I \leq D^2 u \leq \Lambda I$ in the set $B_i := B_{Rh}(x_i) \cap \Omega$ with $R = \frac{\Lambda}{\lambda}\sigma^2$, and let $N_h u$ be the convex nodal function associated with u. If $x_i \in \mathcal{N}_h^0$ is a node with $\operatorname{dist}(x_i, \partial\Omega_h) \leq Rh$, then*

$$\left| |\partial N_h u(x_i)| - \int_\Omega \phi_i(x) \det D^2 u(x) dx \right| \leq C|u|^d_{W^2_\infty(B_i)} \int_\Omega \phi_i(x) dx,$$

where the constant $C = C(d, \lambda, \Lambda)$. This estimate is valid for any $x_i \in \mathcal{N}_h^0$.

Proof We proceed as in Proposition 5.4 and only show

$$|\partial N_h u(x_i)| \leq \int_\Omega \phi_i(x) \det D^2 u(x) dx + C|u|^d_{W^2_\infty(B_i)} \int_\Omega \phi_i(x) dx.$$

Since $\lambda I \leq D^2 u \leq \Lambda I$, we have $A_i(N_h u) \subset B_{Rh}(x_i) \cap \Omega$ according to Lemma 5.1 (size of adjacent sets). The W^2_∞-regularity assumption of u gives the following estimate for the piecewise constant gradient of the convex envelope $\Gamma(N_h u)$ of $N_h u$ over each element T of the mesh induced by $\Gamma(N_h u)$, not necessarily $\mathcal{T}_h$, and contained in $B_{Rh}(x_i)$

$$\nabla \Gamma(N_h u)|_T = \nabla u(x_i) + v_T, \qquad |v_T| \leq Ch|u|_{W^2_\infty(B_i)}.$$

Applying Lemma 2.4 (characterization of subdifferential) we deduce that the convex hull of all $\nabla \Gamma(N_h u)|_T$ for $T \ni x_i$, whence $\partial N_h u(x_i)$, can be bounded by a ball of radius $Ch|u|_{W^2_\infty(B_i)}$ centered at $\nabla u(x_i)$. Hence, we arrive at

$$\begin{aligned} |\partial N_h u(x_i)| &\leq C|u|^d_{W^2_\infty(B_i)} \int_\Omega \phi_i(x) dx \\ &\leq \int_\Omega \phi_i(x) \det D^2 u(x) dx + C|u|^d_{W^2_\infty(B_i)} \int_\Omega \phi_i(x) dx, \end{aligned}$$

because $\det D^2 u(x) \geq 0$ a.e. $x \in \Omega$. This completes the proof. □

6 Rates of convergence

Our goal in this section is to establish rates of convergence in the max-norm for the approximation (3.3) of the Monge–Ampère equation (1.1). We first deal with classical solutions $u \in C^2(\overline{\Omega})$ and next with non-classical solutions $u \in C^{1,1}(\overline{\Omega}) \backslash C^2(\overline{\Omega})$. Interior error estimates result from combining the stability and consistency estimates derived in Sects. 4 and 5. Boundary error estimates entail a different approach involving discrete barrier functions, which we discuss next.

6.1 Discrete Barrier function

Since the consistency estimates of Proposition 5.4 are valid in the interior, we need to treat the boundary layer

$$\{x \in \Omega : \operatorname{dist}(x, \partial\Omega_h) \leq Rh\}$$

differently. We exploit that $N_h u - u_h = 0$ on $\partial\Omega_h$ together with the fact that $N_h u - u_h$ cannot grow too fast from $\partial\Omega_h$. This is a consequence of the next result.

Lemma 6.1 (discrete barrier) *Let Ω be uniformly convex and $\mathcal{N}_h^0$ be translation invariant. Given a constant $E > 0$, for each node $x_i \in \mathcal{N}_h^0$ with $\operatorname{dist}(x_i, \partial\Omega_h) \leq Rh$, $R = \frac{\Lambda}{\lambda}\sigma^2$, there exists a nodal function b_h^i such that $|\partial b_h^i(x_j)| \geq E \int_\Omega \phi_j(x)\, dx$ for all $x_j \in \mathcal{N}_h^0$, $b_h^i(x_j) \leq 0$ at $x_j \in \mathcal{N}_h^\partial$ and*

$$|b_h^i(x_i)| \leq C R E^{1/d} h,$$

provided that h is sufficiently small.

Proof We proceed in three steps. We denote $z = x_i$ for convenience.

Step 1. We first construct a nodal function p_h such that

$$|\partial p_h(x_j)| \geq E \int_\Omega \phi_j(x) dx \quad \forall\, x_j \in \mathcal{N}_h^0.$$

Let $z_0 \in \partial\Omega$ be such that $|z - z_0| = \operatorname{dist}(z, \partial\Omega)$. We introduce a coordinate system with origin at z_0 and $z = (0, \ldots, 0, |z - z_0|)$ and the domain Ω lying within the sphere S_r given by

$$x_1^2 + x_2^2 + \ldots + x_{d-1}^2 + (x_d - r)^2 \leq r^2,$$

where the radius r is a lower bound for the curvature of the boundary $\partial\Omega$ which is strictly positive. Let $p(x)$ be the convex quadratic polynomial

$$p(x) = \frac{E^{1/d}}{2}\left\{x_1^2 + x_2^2 + \ldots + x_{d-1}^2 + (x_d - r)^2 - r^2\right\},$$

which is ≤ 0 in Ω. We consider a extension $\mathcal{N}_h^+$ of the nodal set $\mathcal{N}_h^0$ to $\mathbb{R}^d$

$$\mathcal{N}_h^+ = \left\{ z = h\sum_{j=1}^{d} k_j e_j : k_j \in \mathbb{Z} \right\}$$

where $\{e_j\}$ is the basis spanning the translation invariant set $\mathcal{N}_h^0$. Let $p_h^+ = N_h p$ be the nodal function associated with p over $\mathcal{N}_h^+$, namely $p_h^+(x_j) = p(x_j)$ for all $x_j \in \mathcal{N}_h^+$. Since $\mathcal{N}_h^+$ is translation invariant, Lemma 5.3 (consistency of the discrete Monge–Ampère measure) yields

$$|\partial p_h^+(x_j)| = \int_{\omega_j} \phi_j(x) \det D^2 p(x) dx = E\int_{\omega_j} \phi_j(x) dx \quad \forall x_j \in \mathcal{N}_h^+,$$

where $\omega_j = \mathrm{supp}(\phi_j)$. To define the nodal function p_h on $\mathcal{N}_h$, we set

$$p_h(x_j) := p_h^+(x_j) \quad \forall x_j \in \mathcal{N}_h^0.$$

To define p_h at boundary nodes $x_j \in \mathcal{N}_h^\partial$, which may not belong to $\mathcal{N}_h^+$, we regard the convex envelope $\Gamma(p_h^+)$ of p_h^+ in $\mathbb{R}^d$ as a natural extension and we assign $p_h(x_j) := \Gamma(p_h^+)(x_j)$ for all $x_j \in \mathcal{N}_h^\partial$. In view of Lemma 2.1 (discrete subdifferential), we realize that $\partial p_h(x_j) = \partial\Gamma(p_h^+)(x_j)$ for all $x_j \in \mathcal{N}_h^0$.

Step 2. We assert that

$$p_h(x_j) \leq C E^{1/d} h \quad \forall x_j \in \mathcal{N}_h.$$

Since $p_h(x_j) \leq 0$ for all $x_j \in \mathcal{N}_h^0$, we only need to show this for $x_j \in \partial\Omega$. For each such node, due to the convexity of p_h^+, we have

$$p_h(x_j) = \Gamma(p_h^+)(x_j) \leq \max\{p_h^+(x_k) : x_k \in A_j(p_h^+)\},$$

where $A_j(p_h^+)$ is the adjacent set of x_j for p_h^+. By Lemma 5.1 (size of adjacent sets), $A_j(p_h^+)$ is contained in a ball $B_j = B_{Rh}(x_j)$. We thus deduce

$$p_h(x_j) \leq \max\{p_h^+(x_k) : x_k \in \Omega + B_{Rh}(0)\}.$$

Since Ω is contained in the ball S_r, we infer that

$$p_h(x_j) \leq E^{1/d}\left\{(r + Rh)^2 - r^2\right\} \leq 3E^{1/d} r Rh$$

for h sufficiently small.

Step 3. Finally, if we set

$$b_h^i(x) := p_h(x) - 3E^{1/d} r Rh,$$

then we have $b_h^i(x_j) \le 0$ for all $x_j \in \partial\mathcal{N}_h^\partial$. Moreover, we have

$$|b_h^i(z)| = |p(z) - 3E^{1/d} r Rh| \le CE^{1/d} r Rh.$$

This completes the proof. □

6.2 Rates of convergence for classical solutions

We prove rates of convergence for C^2 classical solutions of (1.1), assuming either Hölder or Sobolev regularity of D^2u. This is the content of Theorems 6.1 and 6.2.

Theorem 6.1 (rate of convergence for $C^{2,\alpha}$ solutions) *Let Ω be uniformly convex and $\mathcal{N}_h^0$ be translation invariant. Let u be the convex solution of the Monge–Ampère equation (1.1) with $\lambda I \le D^2u \le \Lambda I$ in Ω and u_h be the solution of (3.3) with right-hand sides $\{f_i\}_{i=1}^n$ defined over a mesh $\mathcal{T}_h$ with nodes $\mathcal{N}_h$ and translation invariant basis $\{\phi_i\}_{i=1}^n$. If $f(x) \ge \lambda_F > 0$ for all $x \in \Omega$ and $u \in C^{2,\alpha}(\overline{\Omega})$, then*

$$\| u - \Gamma(u_h) \|_{L^\infty(\Omega_h)} \le Ch^\alpha$$

where the constant $C = C(d, \Omega, \lambda, \Lambda, \lambda_F)\big(|u|_{C^{2,\alpha}(\overline{\Omega})} + |u|_{W^2_\infty(\Omega)}\big)$.

Proof Let $\mathcal{T}_h$ be the mesh with nodes $\mathcal{N}_h$ where we define the nodal values f_i according to (3.2), and let N_hu be the nodal function associated with u. Lemma 5.1 (size of adjacent sets) gives $A_i(N_hu) \subset B_{Rh}(x_i)$ for all $x_i \in \mathcal{N}_h^0$. Classical interpolation theory thus yields [9]

$$\| u - \Gamma(N_hu) \|_{L^\infty(\Omega_h)} \le Ch^2 |u|_{W^2_\infty(\Omega)}.$$

Therefore, we only need to prove that $|(N_hu - u_h)(x_i)| \le Ch^\alpha$ for all $x_i \in \mathcal{N}_h^0$, or equivalently the one-sided estimate

$$\sup_{x_i \in \mathcal{N}_h} \big(N_hu - u_h\big)^-(x_i) \le Ch^\alpha \tag{6.1}$$

because a corresponding inequality for $(N_hu - u_h)^+$ can be derived similarly.

Step 1 (boundary estimate). We first show that for all $x_i \in \mathcal{N}_h^0$ such that $\operatorname{dist}(x_i, \partial\Omega_h) \le Rh$ with $R = \frac{\Lambda}{\lambda}\sigma^2$

$$\big(u_h - N_hu\big)(x_i) \le C|u|_{W^2_\infty(B_i)} h.$$

Fix $x_i \in \mathcal{N}_h^0$ and let b_h^i be the discrete barrier defined in Lemma 6.1 (discrete barrier) with free parameter E. We consider the nodal function $u_h + b_h^i$, which satisfies

$$\partial u_h(x_j) + \partial b_h^i(x_j) \subset \partial(u_h + b_h^i)(x_j)$$

due to Lemma 2.3 (addition of subdifferentials). Applying (4.2), which is a consequence of Lemma 4.2 (Brunn–Minkowski inequality), implies

$$|\partial(u_h + b_h^i)(x_j)| \geq |\partial u_h(x_j)| + |\partial b_h^i(x_j)|.$$

Since $\partial u_h(x_j) = f_j$ according to (3.3), invoking Lemma 6.1 (discrete barrier) and Lemma 5.5 (boundary consistency) yields

$$|\partial(u_h + b_h^i)(x_j)| \geq f_j + E \int_\Omega \phi_j(x)dx \geq |\partial N_h u(x_j)| \quad \forall\, x_j \in \mathcal{N}_h^0,$$

provided $E \geq C|u|^d_{W^2_\infty(B_j)}$. Moreover, $b_h^i(x_j) \leq 0$ and $u_h(x_j) = N_h u(x_j) = g(x_j)$ for all $x_j \in \mathcal{N}_h^\partial$ imply $u_h + b_h^i \leq N_h u$ on $\partial\Omega_h$, whence Corollary 4.4 (discrete maximum principle) gives

$$u_h(x_j) + b_h^i(x_j) \leq N_h u(x_j) \quad \text{for all } x_j \in \mathcal{N}_h^0.$$

Finally, the estimate for $b_h^i(x_i)$ of Lemma 6.1 (discrete barrier) yields

$$u_h(x_i) - N_h u(x_i) \leq -b_h^i(x_i) \leq C|u|_{W^2_\infty(B_i)} h. \tag{6.2}$$

Step 2 (interior estimate). We intend to apply Proposition 4.3 (continuous dependence) to the nodal function $N_h u - u_h$ for all nodes $x_i \in \mathcal{N}_h^0$ with $\operatorname{dist}(x_i, \partial\Omega) \geq Rh$, for which we need to compare subdifferentials and verify boundary conditions. To deal with the former, we restate Proposition 5.4 (interior consistency) with the help of (1.1a) and (3.3), namely $\int_\Omega \phi_i(x) \det D^2 u(x)dx = f_i = |\partial u_h(x_i)|$:

$$|\partial N_h u(x_i)| \leq |\partial u_h(x_i)| + Ch^\alpha |u|_{C^{2,\alpha}(\overline{\Omega})} \int_\Omega \phi_i(x)dx.$$

Setting $\epsilon := Ch^\alpha |u|_{C^{2,\alpha}(\overline{\Omega})} \int_\Omega \phi_i(x)dx$, we readily see that

$$|\partial N_h u(x_i)|^{1/d} - |\partial u_h(x_i)|^{1/d} \leq (f_i + \epsilon)^{1/d} - f_i^{1/d}$$

for $x_i \in \mathcal{C}_h^-(N_h u - u_h)$. Since the function $\psi(t) = t^{1/d}$ is concave for $t > 0$, we deduce that $\psi(t + \epsilon) - \psi(t) \leq d^{-1} t^{1/d-1} \epsilon$, whence

$$|\partial N_h u(x_i)|^{1/d} - |\partial u_h(x_i)|^{1/d} \leq Ch^\alpha |u|_{C^{2,\alpha}(\overline{\Omega})} f_i^{1/d-1} \int_\Omega \phi_i(x)dx.$$

Exploiting now the lower bound of f, namely $f(x) \geq \lambda_F > 0$ for all $x \in \Omega$, we estimate f_i from below

$$f_i = \int_\Omega f(x)\phi_i(x)dx \geq \lambda_F \int_\Omega \phi_i(x)dx,$$

and insert this bound back into the preceding expression to obtain

$$\left(|\partial N_h u(x_i)|^{1/d} - |\partial u_h(x_i)|^{1/d}\right)^d \leq C\lambda_F^{1-d} h^{\alpha d} |u|^d_{C^{2,\alpha}(\overline{\Omega})} \int_\Omega \phi_i(x)dx.$$

In order to apply Proposition 4.3 (continuous dependence) to the nodal function $N_h u - u_h$ it remains to check boundary conditions on the smaller domain

$$\Omega_h^0 := \{x \in \Omega : \operatorname{dist}(x, \partial\Omega_h) \geq Rh\},$$

where the above calculation is valid. Since $\big(u_h - N_h u\big)(x_i) \leq C|u|_{W^2_\infty(B_i)} h$ for $x_i \in \mathcal{N}_h^0 \backslash \Omega_h^0$, according to (6.2), Proposition 4.3 leads to

$$\sup_{x_i \in \Omega_h^0 \cap \mathcal{N}_h^0} \big(N_h u - u_h + C|u|_{W^2_\infty(B_i)} h\big)^-(x_i) \leq C\lambda_F^{1/d-1} |\Omega|^{1/d} h^\alpha |u|_{C^{2,\alpha}(\overline{\Omega})},$$

or equivalently to

$$\big(u_h - N_h u\big)(x_i) \leq C|u|_{W^2_\infty(B_i)} h + C\lambda_F^{1/d-1} |\Omega|^{1/d} h^\alpha |u|_{C^{2,\alpha}(\overline{\Omega})} \tag{6.3}$$

for all $x_i \in \Omega_h^0 \cap \mathcal{N}_h^0$. Combining (6.2) with (6.3) proves the desired estimate (6.1) and concludes the proof. □

Remark 6.1 (non-degeneracy) The lower bound λ_F in the non-degeneracy assumption $f(x) \geq \lambda_F > 0$ of Theorem 6.1 may be viewed as a stability constant because $\lambda_F^{1/d-1}$ blows up as $\lambda_F \to 0$. If f vanishes somewhere, then $\lambda_F = 0$ and we have a reduced convergence rate $h^{\alpha/d}$ because $(f_i + \epsilon)^{1/d} - f_i^{1/d} \leq \epsilon^{1/d}$ for all $f_i \geq 0$.

Remark 6.2 ($C^{2,\alpha}$-regularity) It is worth observing that the assumptions on u in Theorem 6.1 can be verified from assumptions on f. In fact, if $0 < \lambda_F \leq f(x) \leq \Lambda_F$, then $0 < \lambda \leq D^2 u(x) \leq \Lambda$ for some constants λ, Λ [12], [30, Section 4.1]. Moreover, if $f(x) \in C^\alpha(\overline{\Omega})$ and Ω is of class $C^{2,\alpha}$, then $u \in C^{2,\alpha}(\overline{\Omega})$ [11], [30, Section 4.3].

Theorem 6.2 (rate of convergence for W^s_q solutions) *Let Ω be uniformly convex and $\mathcal{N}_h^0$ be translation invariant. Let u be the convex solution of the Monge–Ampère equation (1.1) with $\lambda I \leq D^2 u \leq \Lambda I$ in Ω and u_h be the solution of (3.3) with right-hand side $\{f_i\}_{i=1}^n$ defined in (3.2) over a mesh $\mathcal{T}_h$ with nodes $\mathcal{N}_h$ and translation invariant basis $\{\phi_i\}_{i=1}^n$. If $f(x) \geq \lambda_F > 0$ for all $x \in \Omega$ and $u \in W^s_q(\Omega)$ with $s > 2 + \frac{d}{q}$, $s \leq 3$, $d < q \leq \infty$, then*

$$\| u - \Gamma(u_h) \|_{L^\infty(\Omega_h)} \leq Ch^{s-2}$$

where the constant $C = C(d, \Omega, \lambda, \Lambda, \lambda_F)\big(|u|_{W^s_q(\Omega)} + |u|_{W^2_\infty(\Omega)}\big)$.

Proof We just employ the second estimate of Proposition 5.4 (interior consistency) in Step 2 of the proof of Theorem 6.1 (rate of convergence for $C^{2,\alpha}$ solutions). The key difference is the estimate of the sum

$$S := h^{s-2-\frac{d}{q}} \left(\sum_{x_i \in \mathcal{N}_h^0} |u|^d_{W^s_q(B_i)} \int_\Omega \phi_i(x) dx \right)^{\frac{1}{d}}$$

before applying Proposition 4.3 (continuous dependence); recall that $B_i = B(x_i, Rh)$. To exploit the ℓ^q-summability of $\{|u|_{W^s_q(B_i)}\}_{x_i \in \mathcal{N}_h^0}$ we utilize Hölder inequality with exponents $t = q/d$ and $t^* = q/(q-d)$ to arrive at

$$S \le h^{s-2-\frac{d}{q}} \left(\sum_{x_i \in \mathcal{N}_h^0} |u|^q_{W^s_q(B_i)} \right)^{\frac{1}{q}} \left(\sum_{x_i \in \mathcal{N}_h^0} \Big(\int_\Omega \phi_i \Big)^{\frac{q}{q-d}} \right)^{\frac{q-d}{dq}}.$$

We note that the balls B_i have a finite overlapping property. In fact, if $B_i \cap B_j \neq \emptyset$, then $\operatorname{dist}(x_i, x_j) \le 2Rh$ and the box

$$\omega_j := \Big\{ x = x_j + \sum_{k=1}^d t e_k : -\frac{h}{2} \le t \le \frac{h}{2} \Big\} \tag{6.4}$$

is contained in $B_{Ch}(x_i)$ where $C = 2R + \frac{1}{2}d$. Since $|B_{Ch}(x_i)| \le Ch^d$ and $|\omega_j| \ge ch^d$, with $c = c(\sigma)$, we deduce that balls $B_{Ch}(x_i)$ contain a fixed number of nodes x_j and B_i overlaps with a fixed number of B_j's. Hence

$$\left(\sum_{x_i \in \mathcal{N}_h^0} \Big(\int_\Omega \phi_i \Big)^{\frac{q}{q-d}} \right)^{\frac{q-d}{dq}} \le C|\Omega|^{\frac{q-d}{dq}} h^{\frac{d}{q}},$$

and we deduce that $S \le Ch^{s-2}|u|_{W^s_q(\Omega)}$ to conclude the proof. □

We stress that Theorem 6.2 (rate of convergence for W^s_q solutions) provides a linear rate for solutions $u \in W^3_q(\Omega)$ with $q > d$, which is much weaker than the requirement $u \in W^3_\infty(\Omega)$ of Theorem 6.1 (rate of convergence for $C^{2,\alpha}$ solutions) for a similar rate. We explore this further below.

6.3 Rates of convergence for piecewise smooth solutions

We now study the case where the consistency error may be large in a small region, for instance when the Hessian D^2u jumps across a surface within Ω, but is small

otherwise. We therefore give up the assumption $u \in C^2(\overline{\Omega})$. We exploit the structure of the estimate in Proposition 4.3 (continuous dependence), namely the fact that its right-hand side accumulates in ℓ_d.

Before stating our result, we introduce the Minkowski-Bouligand dimension of a subset ω of Ω. Given a translation invariant set of nodes $\mathcal{N}_h = \{x_i\}$, let $\{\omega_i\}_{x_i \in \mathcal{N}_h}$ be the translation invariant partition covering Ω defined in (6.4). Let $m(h)$ be the number of ω_i's required to cover ω. We define the (Minkowski-Bouligand or box) dimension of ω to be

$$\dim \omega := -\lim_{h \to 0} \frac{\log m(h)}{\log h}.$$

For example, it is easy to check that ∂B_1, the discontinuity set of $D^2 u$ in example (3.1), is of dimension one. In addition, the solution u satisfies $u \in W^2_\infty(B_2(0)) \backslash C^2(\overline{B_2(0)})$ and $u \in C^3(\overline{B_1(0)}), C^3(\overline{B_2(0)} \backslash B_1(0))$. The following theorem explores situations such as this one.

Theorem 6.3 (convergence rates for piecewise smooth Hessians) *Let Ω be uniformly convex and $\mathcal{N}_h$ be a translation invariant set of nodes. Let $u \in W^2_\infty(\Omega)$ be the convex solution of the Monge–Ampère equation (1.1) and satisfy $\lambda I \leq D^2 u \leq \Lambda I$ in Ω. Let $D^2 u$ be piecewise smooth in the sense that $D^2 u \in W^s_q(\Omega \backslash \omega)$ with $s > 2 + d/q$, $s \leq 3$, $d < q \leq \infty$, and let ω have box dimension $n < d$. If $f \geq \lambda_F > 0$ in Ω and u_h is the solution of (3.3) with right-hand side $\{f_i\}_{i=1}^n$ defined in (3.2) over a mesh $\mathcal{T}_h$ with nodes $\mathcal{N}_h$ and translation invariant basis $\{\phi_i\}_{i=1}^n$, then*

$$\| u - \Gamma(u_h) \|_{L^\infty(\Omega_h)} \leq C h^{s-2} |u|_{W^s_q(\Omega \backslash \omega)} + C h^{\frac{d-n}{d}} |u|_{W^2_\infty(\Omega)},$$

where the constant $C = C(d, \Omega, \omega, \lambda, \Lambda, \lambda_F)$.

Proof We argue as in Theorems 6.1 and 6.2. Therefore, we first observe that $u \in W^2_\infty(\Omega)$ guarantees

$$|(N_h u - u_h)(x_i)| \leq C |u|_{W^2_\infty(\Omega)} h$$

for all $x_i \in \mathcal{N}_h^0$ such that dist $(x_i, \partial\Omega_h) < Rh$. We split the nodes $x_i \in \mathcal{N}_h^0$ such that dist $(x_i, \partial\Omega_h) \geq Rh$ into two sets, those that are at distance Rh to ω, denoted by $\mathcal{N}_h^0(\omega)$, and the complement $\mathcal{N}_h^0(\Omega \backslash \omega)$.

In order to apply Proposition 4.3 (continuous dependence), we start with the estimate

$$|\partial N_h u(x_i)|^{1/d} - |\partial u_h(x_i)|^{1/d} \leq (f_i + \epsilon_i)^{1/d} - f_i^{1/d} \leq d^{-1} \lambda_F^{1/d - 1} \delta_i$$

and $\delta_i = \epsilon_i \big(\int_\Omega \phi_i \big)^{1/d-1}$, in conjunction with the expression of ϵ_i already derived in the second estimate of Proposition 5.4 (interior consistency), to arrive at

$$\delta_i = C h^{s-2-\frac{d}{q}} |u|_{W^s_q(B_i)} \Big(\int_\Omega \phi_i \Big)^{\frac{1}{d}} \qquad \forall\, x_i \in \mathcal{N}_h^0(\Omega \backslash \omega).$$

We next resort to the crude bound $(f_i + \epsilon_i)^{1/d} - f_i^{1/d} \le \epsilon_i^{1/d}$, together with the expression of ϵ_i derived in Lemma 5.5 (boundary consistency), to write

$$\epsilon_i = C|u|_{W^2_\infty(B_i)} \Big(\int_\Omega \phi_i \Big)^{\frac{1}{d}} \qquad \forall\, x_i \in \mathcal{N}_h^0(\omega).$$

We point out that Lemma 5.5 is valid for any $x_i \in \mathcal{N}_h^0$. Arguing as in Theorem 6.2 (rate of convergence for W^s_q solutions), we thus obtain

$$\sum_{x_i \in \mathcal{N}_h^0(\Omega\setminus\omega)} \delta_i^d \le C h^{(s-2)d}\, |u|^d_{W^s_q(\Omega\setminus\omega)},$$

as well as

$$\sum_{x_i \in \mathcal{N}_h^0(\omega)} \epsilon_i^d \le C|u|^d_{W^2_\infty(\Omega)} \sum_{x_i \in \mathcal{N}_h^0(\omega)} |\omega_i| \le C h^{d-n}\, |u|^d_{W^2_\infty(\Omega)}$$

because $\sum_{x_i \in \mathcal{N}_h^0(\omega)} |\omega_i| \le Cm(h)h^d \le h^{d-n}$ with n being the box dimension of ω. Applying now Proposition 4.3 (continuous dependence) to $N_h(u) - u_h$ yields

$$\sup \big(N_h(u) - u_h \big)^- \le C h^{s-2}\, |u|_{W^s_q(\Omega\setminus\omega)} + C h^{\frac{d-n}{d}}\, |u|_{W^2_\infty(\Omega)}.$$

This is the asserted estimate. □

We conclude with a simple application of Theorem 6.3 (convergence rates for piecewise smooth Hessians) to the example (3.1). Since $d = 2$, $s = 3$, $q = \infty$, and $n = \dim(\partial B_1) = 1$, we deduce

$$\| u - \Gamma(u_h) \|_{L^\infty(\Omega_h)} \le C(u) h^{1/2}.$$

We point out that the singular set $\omega = \partial B_1$ need not be matched by either the mesh $\mathcal{T}_h$ associated with the translation invariant nodal set $\mathcal{N}_h^0$ or the mesh induced by the convex envelope $\Gamma(u_h)$ of u_h.

References

1. Aguilera, N.E., Morin, P.: On convex functions and the finite element method. SIAM J. Numer. Anal. **47**(4), 3139–3157 (2009)
2. Awanou, G.: Standard finite elements for the numerical resolution of the elliptic Monge–Ampère equations: classical solutions. IMA J. Numer. Anal. **35**(3), 1150–1166 (2015)
3. Barles, G., Souganidis, P.E.: Convergence of approximation schemes for fully nonlinear second order equations. Asymptot. Anal. **4**, 271–283 (1991)
4. Benamou, J.D., Collino, F., Mirebeau, J.M.: Monotone and consistent discretization of the Monge–Ampère operator. Math. Comp. **85**(302), 2743–2775 (2016)
5. Benamou, J.D., Froese, D.B., Oberman, A.M.: Two numerical methods for the elliptic Monge–Ampère equation. M2AN. Math. Model. Numer. Anal. **44**, 737–758 (2010)

6. Böhmer, K.: On finite element methods for fully nonlinear elliptic equations of second order. SIAM J. Numer. Anal. **46**(3), 1212–1249 (2008)
7. Brenner, S.C., Gudi, T., Neilan, M., Sung, L.-Y.: C^0 penalty methods for the fully nonlinear Monge–Ampère equation. Math. Comp. **80**(276), 1979–1995 (2011)
8. Brenner, S.C., Neilan, M.: Finite element approximations of the three dimensional Monge–Ampère equation. ESAIM Math. Model. Numer. Anal. **46**(5), 979–1001 (2012)
9. Brenner, S.C., Scott, L.R.: The Mathematical Theory of Finite Element Methods, Volume 15 of Texts in Applied Mathematics, 3rd edn. Springer, New York (2008)
10. Caffarelli, L., Cabré, X.: Fully Nonlinear Elliptic Equations, volume 43 of American Mathematical Society Colloquium Publications. American Mathematical Society, Providence (1995)
11. Caffarelli, L.A.: Interior $W^{2,p}$ estimates for solutions of the Monge–Ampère equation. Ann. of Math. (2) **131**(1), 135–150 (1990)
12. Caffarelli, L.A.: A localization property of viscosity solutions to the Monge–Ampère equation and their strict convexity. Ann. of Math. (2) **131**(1), 129–134 (1990)
13. Carlier, G., Lachand-Robert, T., Maury, B.: A numerical approach to variational problems subject to convexity constraint. Numer. Math. **88**(2), 299–318 (2001)
14. Chen, L., Holst, M.: Efficient mesh optimization schemes based on optimal Delaunay triangulations. Comput. Methods Appl. Mech. Eng. **200**(9–12), 967–984 (2011)
15. Chen, L., Xu, J.: Optimal Delaunay triangulations. J. Comput. Math. **22**(2), 299–308 (2004). Special issue dedicated to the 70th birthday of Professor Zhong-Ci Shi
16. Choné, P., Le Meur, H.J.: Non-convergence result for conformal approximation of variational problems subject to a convexity constraint. Numer. Funct. Anal. Optim. **22**(5–6), 529–547 (2001)
17. Crandall, M.G., Ishii, H., Lions, P.L.: User's guide to viscosity solutions of second order partial differential equations. Bull. Am. Math. Soc. (N.S.) **27**(1), 1–67 (1992)
18. Dean, E.J., Glowinski, R.: 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**(9), 779–784 (2003)
19. Dean, E.J., Glowinski, R.: 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**(12), 887–892 (2004)
20. Dean, E.J., Glowinski, R.: An augmented Lagrangian approach to the numerical solution of the Dirichlet problem for the elliptic Monge–Ampère equation in two dimensions. Electron. Trans. Numer. Anal. **22**, 71–96 (2006). (electronic)
21. Dean, E.J., Glowinski, R.: Numerical methods for fully nonlinear elliptic equations of the Monge–Ampère type. Comput. Methods Appl. Mech. Eng. **195**(13–16), 1344–1386 (2006)
22. Edelsbrunner, H.: Triangulations and meshes in computational geometry. In Acta numerica, 2000, volume 9 of Acta Numer., pp. 133–213. Cambridge Univ. Press, Cambridge (2000)
23. Feng, X., Kao, C.-Y., Lewis, T.: Convergent finite difference methods for one-dimensional fully nonlinear second order partial differential equations. J. Comput. Appl. Math. **254**, 81–98 (2013)
24. Feng, X., Neilan, M.: Mixed finite element methods for the fully nonlinear Monge–Ampère equation based on the vanishing moment method. SIAM J. Numer. Anal. **47**(2), 1226–1250 (2009)
25. Feng, X., Neilan, M.: Vanishing moment method and moment solutions for fully nonlinear second order partial differential equations. J. Sci. Comput. **38**(1), 74–98 (2009)
26. Feng, X., Neilan, M.: Analysis of Galerkin methods for the fully nonlinear Monge–Ampère equation. J. Sci. Comput. **47**(3), 303–327 (2011)
27. Froese, B.D., Oberman, A.M.: Convergent finite difference solvers for viscosity solutions of the elliptic Monge–Ampère equation in dimensions two and higher. SIAM J. Numer. Anal. **49**, 1692–1714 (2011)
28. Froese, B.D., Oberman, A.M.: Fast finite difference solvers for singular solutions of the elliptic Monge–Ampère equation. J. Comput. Phys. **230**(3), 818–834 (2011)
29. Gardner, R.J.: The Brunn–Minkowski inequality. Bull. Am. Math. Soc. (N.S.) **39**(3), 355–405 (2002)
30. Gutiérrez, C.E.: The Monge–Ampère Equation. Progress in Nonlinear Differential Equations and their Applications, vol. 44. Birkhäuser Boston, Boston (2001)
31. Joe, B.: Construction of three-dimensional Delaunay triangulations using local transformations. Comput. Aided Geom. Des. **8**(2), 123–142 (1991)
32. Kuo, H.J., Trudinger, N.S.: Linear elliptic difference inequalites with random coefficients. Math. Comp. **55**, 37–53 (1990)

33. Kuo, H.J., Trudinger, N.S.: Discrete methods for fully nonlinear elliptic equations. SIAM J. Numer. Anal. **29**, 123–135 (1992)
34. Kuo, H.-J., Trudinger, N.S.: Positive difference operators on general meshes. Duke Math. J. **83**, 415–433 (1996)
35. Kuo, H.-J., Trudinger, N.S.: A note on the discrete Aleksandrov–Bakelman maximum principle. In: Proceedings of 1999 International Conference on Nonlinear Analysis (Taipei), vol. 4, pp. 55–64 (2000)
36. Mérigot, Q., Oudet, É.: Handling convexity-like constraints in variational problems. SIAM J. Numer. Anal. **52**(5), 2466–2487 (2014)
37. Mirebeau, J.M.: Discretization of the 3D Monge–Ampere operator, between wide stencils and power diagrams. ESAIM Math. Model. Numer. Anal. **49**(5), 1511–1523 (2015)
38. Neilan, M.: A nonconforming Morley finite element method for the fully nonlinear Monge–Ampère equation. Numer. Math. **115**(3), 371–394 (2010)
39. Neilan, M.: Quadratic finite element approximations of the Monge–Ampère equation. J. Sci. Comput. **54**(1), 200–226 (2013)
40. Nochetto, R.H., Zhang, W.: Discrete ABP estimate and convergence rates for linear elliptic equations in non-divergence form. Found. Comput. Math. **18**(3), 537–593 (2018)
41. Oberman, A.M.: 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**, 221–238 (2008)
42. Oberman, A.M.: A numerical method for variational problems with convexity constraints. SIAM J. Sci. Comput. **35**(1), A378–A396 (2013)
43. Oliker, V.I., Prussner, L.D.: 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**, 271–293 (1988)
44. Sorensen, D.C., Glowinski, R.: A quadratically constrained minimization problem arising from PDE of Monge–Ampère type. Numer. Algor. **53**(1), 53–66 (2010)