DISCRETIZATION OF THE 3D MONGE−AMPERE OPERATOR, BETWEEN WIDE STENCILS AND POWER DIAGRAMS

Jean-Marie Mirebeau[1]

Abstract. We introduce a monotone (degenerate elliptic) discretization of the Monge−Ampere operator, on domains discretized on cartesian grids. The scheme is consistent provided the solution hessian condition number is uniformly bounded. Our approach enjoys the simplicity of the Wide Stencil method [B.D. Froese and A.M. Oberman, *J. Comput. Phys.* **230** (2011) 818–834.], but significantly improves its accuracy using ideas from discretizations of optimal transport based on power diagrams [F. Aurenhammer, F. Hoffmann and B. Aronov, *Algorithmica* (1998)]. We establish the global convergence of a damped Newton solver for the discrete system of equations. Numerical experiments, in three dimensions, illustrate the scheme efficiency.

Mathematics Subject Classification. 65N06, 65N12, 35B50, 35J60, 49L25.

1. Introduction

We introduce a discretization of the Monge−Ampere operator, on three dimensional cartesian grids, which is simultaneously monotone and consistent. Existing consistent schemes, based *e.g.* on Finite Elements [4,14] or Finite Differences [12], are not monotone, and thus require the PDE solution to be sufficiently smooth, and the numerical solver to be well initialized. Existing monotone schemes, based on Wide Stencil discretizations [8,15], suffer from a consistency error depending on the discretization stencil angular resolution. Filtered schemes [9] combine a monotone and a consistent scheme, and attempt to cumulate their robustness and accuracy; improving either of the constituting schemes will benefit to the filtered combination. A monotone and consistent scheme is introduced in [3], but it is limited to two dimensions. Geometric approaches [16] are discussed in the third paragraph.

The proposed numerical scheme belongs to the Wide–Stencil category [15], in the sense that we actually define a family of schemes parameterized by a user chosen stencil. Larger stencils provide consistency for strongly anisotropic problems (*i.e.* for which the solution hessian is almost degenerate), but at the cost of an increased computation time. The choice of stencil is left to the user; let us mention that in the special case of [3] an automatic (solution adaptive, local, anisotropic, and parameter free) stencil construction could be designed. Our numerical experiments show that small stencils, of radius $\sqrt{3}$ or $\sqrt{6}$, see Table 1, already yield convincing

[1] CNRS, University Paris Dauphine, UMR 7534, Laboratory CEREMADE, Paris, France.
jm.mirebeau@gmail.com

TABLE 1. *Left*: Stencil V used with the proposed scheme $\mathcal{D}_V$. In addition to those indicated $(\alpha_1, \alpha_2, \alpha_3)$, the stencil contains all vectors $(\varepsilon_1 \alpha_{\sigma(1)}, \varepsilon_2 \alpha_{\sigma(2)}, \varepsilon_3 \alpha_{\sigma(3)})$ obtained by permuting and changing the sign of their coordinates. Opposite vectors are identified when evaluating $\#(V)$. *Right*: Collection $\mathcal{B}$ of orthogonal triplets used with the Wide–Stencil scheme $\mathcal{D}_{\mathcal{B}}^{\mathrm{WS}}$. Triplets are counted up to reordering their components, and changing their sign: $(e_0, e_1, e_2) \sim (\varepsilon_1 e_{\sigma(1)}, \varepsilon_2 e_{\sigma(2)}, \varepsilon_3 e_{\sigma(3)})$. Vectors with non-coprime coordinates, such as $(2, 2, 0)$, are rejected.

Proposed	Stencil vectors	$\#(V)$
Small	$(1,0,0), (1,1,0), (1,1,1)$	13
Large	Same and $(2,1,0), (2,1,1)$	37

Scheme WS	Orthogonal triplets	$\#(\mathcal{B})$
Small	All within $\{-1,0,1\}^3$	6
Medium	All within $\{-2, \ldots, 2\}^3$	43
Large	All within $\{-3, \ldots, 3\}^3$	82

results. The scheme is dimension independent, but we emphasize its application to three dimensional domains, which is tractable and tested.

Our approach is also inspired by [16] and the discretizations of optimal transport [1, 11, 13] based on global geometric structures, up to two differences. The first modification, a *symmetrization* see Remark 1.9, is required to operate our method with the Dirichlet boundary conditions of the standard Monge−Ampere problem (1.1), instead of the second boundary conditions implicit in optimal transportation. The second modification *localizes* these methods by limiting interactions to close discretization points, see Remark 1.10, which considerably simplifies their numerical implementation. The methods [1, 11, 13, 16] indeed rely on global geometric structures named power diagrams, which generalize Voronoï diagrams. Their construction requires state of the art methods of discrete geometry, which especially in 3D are still an active subject of research. For instance [11] mentions arbitrary precision arithmetic, arithmetic filtering, expansion arithmetics and symbolic perturbation, merely for the consistent evaluation of geometric predicates. Our approach is in contrast local and requires none of these subtleties. We show that this simplification preserves consistency in the setting of viscosity solutions, see Section 2.3, but at the following price: the solution hessian condition number must be uniformly bounded, and the weak Alexandroff solutions cannot be recovered.

We fix throughout this paper an open, convex and bounded domain $\Omega \subseteq \mathbb{R}^d$, $d \geq 2$ ($d = 3$ in the numerical Sect. 3).Given a positive density $\rho \in C^0(\overline{\Omega}, \mathbb{R}_+^*)$, and some Dirichlet data $\sigma \in C^0(\partial\Omega, \mathbb{R})$, we set the goal of approximating numerically the unique viscosity solution [5, 10] of

$$\begin{cases} \det(\nabla^2 u) = \rho & \text{on } \Omega, \\ u = \sigma & \text{on } \partial\Omega, \\ u \text{ convex}, \end{cases} \tag{1.1}$$

where $\nabla^2 u$ denotes the hessian matrix of u. The PDE domain Ω is discretized on a cartesian grid X. Up to a linear change of coordinates, encoding scaling, rotation and offset, we may assume that

$$X \subseteq \Omega \cap \mathbb{Z}^d.$$

The discussion of the discretization of the boundary $\partial\Omega$ is postponed to Remark 2.4.

Definition 1.1. We denote by $\mathbb{U}$ the collection of maps $u : X \cup \partial\Omega \to \mathbb{R}$.

Definition 1.2. A discrete operator $\mathcal{D}$ associates to each $u \in \mathbb{U}$ a discrete map $\mathcal{D}u : X \to \mathbb{R}$.

Given some discretization $\mathcal{D}$ of the Monge−Ampere operator, the counterpart of (1.1) takes the form: find $u \in \mathbb{U}$ such that

$$\begin{cases} \mathcal{D}u = \rho & \text{on } X, \\ u = \sigma & \text{on } \partial\Omega. \end{cases} \tag{1.2}$$

The constraint "u convex" is not spelled explicitly in (1.2), contrary to (1.1), but some discrete counterpart of it often follows from the identity $\mathcal{D}u = \rho$ [3]. Before discretizing the Monge−Ampere operator, we need to introduce a more basic operator Δ_e, $e \in \mathbb{Z}^d$, aimed at approximating the second order difference $\langle e, (\nabla^2 u(x))e\rangle$. If $x \in X$, $x + e \in X$ and $x - e \in X$, then we set classically

$$\Delta_e u(x) := u(x + e) - 2u(x) + u(x - e). \tag{1.3}$$

In general, one may not have $x + e \in X$, for instance if $x + e$ lies outside Ω. Hence we introduce

$$h_x^e := \min\{h > 0; \ x + he \in X \cup \partial\Omega\}. \tag{1.4}$$

Let $h^+ := h_x^e$ and $h^- := h_x^{-e}$. Only one linear combination of $u(x)$, $u(x + h^+ e)$ and $u(x - h^- e)$ is consistent with $\langle e, (\nabla^2 u(x))e\rangle$, namely

$$\Delta_e u(x) := \frac{2}{h^+ + h^-}\left(\frac{u(x + h^+ e) - u(x)}{h^+} + \frac{u(x - h^- e) - u(x)}{h^-}\right). \tag{1.5}$$

We next illustrate the notions of consistency and monotonicity (degenerate ellipticity) using two now classical discretizations of the Monge−Ampere operator. The Finite Differences (FD) [12] approximation of the Monge−Ampere operator is defined as the determinant of a naive but consistent approximation of the hessian matrix of u by finite differences: denoting by $(e_i)_{1 \leq i \leq d}$ the canonical basis of $\mathbb{R}^d$

$$\mathcal{D}^{\text{FD}}u(x) := \det(\delta_{ij})_{1 \leq i,j \leq 3}, \quad \text{with } \delta_{ij} = \begin{cases} \Delta_{e_i} u(x) & \text{if } i = j, \\ \frac{1}{4}\left(\Delta_{e_i + e_j} u(x) - \Delta_{e_i - e_j} u(x)\right) & \text{if } i \neq j. \end{cases} \tag{1.6}$$

The Wide Stencil (WS) [8] approximation of the Monge−Ampere operator is defined as follows: denoting by $\mathcal{B} \subseteq (\mathbb{Z}^d)^d$ a finite collection of d-plets of pairwise orthogonal vectors

$$\mathcal{D}_{\mathcal{B}}^{\text{WS}}u(x) := \min_{B \in \mathcal{B}} \prod_{e \in B} \frac{\max\{\Delta_e u(x), 0\}}{\|e\|^2}. \tag{1.7}$$

Definition 1.3. An operator $\mathcal{D}$ is said Degenerate Elliptic of second order (DE2), with stencil V, iff for all $x \in X$, $\mathcal{D}u(x)$ is a non-decreasing function of the second order differences $(\Delta_e u(x))_{e \in V}$.

By construction, scheme WS is DE2, with stencil $\{e \in \mathbb{Z}^d; \exists B \in \mathcal{B}, e \in B\}$. Scheme FD in contrast is *not* DE2. Degenerate Ellipticity comes with strong guarantees including a maximum principle for the discrete solutions of (1.2), and guaranteed convergence of Euler iterative solvers for this discrete system [15]. It allows to recover non-smooth viscosity solutions of (1.1), see Section 2.3, and plays a crucial role in the proof Section 2.2 of the global convergence of a damped Newton solver for (1.2).

Let S_d denote the set of symmetric $d \times d$ matrices, and let $S_d^+ \subseteq S_d$ be the subset of positive definite matrices. In order to analyze the consistency of these schemes, we introduce for each $M \in S_d^+$ the norm $\|\cdot\|_M$ and the map $u_M \in \mathbb{U}$ defined by

$$\|x\|_M := \sqrt{\langle x, Mx\rangle}, \qquad\qquad u_M(x) := \frac{1}{2}\|x\|_M^2. \tag{1.8}$$

Definition 1.4. The consistency set of an operator $\mathcal{D}$ is the collection of all $M \in S_d^+$ such that $\mathcal{D}u_M = \det(M)$, identically on X.

The consistency set of scheme FD is the whole S_d^+; in fact, the identity $\mathcal{D}u_M(x) = \det(M)$ also holds for non-positive matrices, although they are irrelevant for our application. The consistency set of scheme WS is in contrast of *empty interior*; precisely, it consists of matrices $M \in S_d^+$ for which some $B \in \mathcal{B}$ is an eigenbasis [15]. Stated otherwise, the consistency of scheme WS is an asymptotic property, obtained by growing the stencil size to infinity.

A combination $\mathcal{D}u(x) = w(u,x)\mathcal{D}_{\mathcal{B}}^{\mathrm{WS}}u(x) + (1 - w(u,x))\mathcal{D}^{\mathrm{FD}}u(x)$ of schemes FD and WS is considered in [9]. The weight $w(u,x) \in [0,1]$ depends on the local behavior of u close to x, as well as on the discretization scale and the angular resolution of the stencil $\mathcal{B}$. Strictly speaking, the weighted scheme loses both the degenerate ellipticity of $\mathcal{D}_{\mathcal{B}}^{\mathrm{WS}}$, and the consistency of $\mathcal{D}^{\mathrm{FD}}$. Nevertheless [9] establishes convergence in the setting of viscosity solutions, and illustrates numerically (in two space dimensions) that the weight $w(u,x)$ favors the consistent scheme $\mathcal{D}^{\mathrm{FD}}$ except close to the most singular features of the solution. Second order accuracy may in principle be achieved even on a degenerate equation, contrary to the method proposed below. The filtered scheme construction proposed in [9] is quite flexible, and could be improved by replacing the monotone scheme WS with the more accurate, and still monotone, scheme proposed in this paper.

We propose an alternative solution to the apparent conflict between between degenerate ellipticity and consistency. The coordinates of a vector $e = (a_1, \ldots, a_d) \in \mathbb{Z}^d$ are said co-prime iff $\gcd(a_1, \ldots, a_d) = 1$.

Definition 1.5. We limit our attentions to stencils $V \subseteq \mathbb{Z}^d \backslash \{0\}$ which are finite, symmetric w.r.t the origin, span $\mathbb{R}^d$, and which elements have co-prime coordinates. The proposed operator is

$$\mathcal{D}_V u(x) := \mathrm{Leb}\{g \in \mathbb{R}^d; \forall e \in V, 2\langle g, e\rangle \leq \Delta_e u(x)\}, \tag{1.9}$$

where Leb denotes d-dimensional Lebesgue measure.

The Degenerate Ellipticity of $\mathcal{D}_V$ is clear: if any second order difference $\Delta_e u(x)$ increases, then the convex polytope appearing in (1.9) increases for inclusion, hence also in volume. Refining slightly this argument we obtain the derivative of $\mathcal{D}_V u(x)$ with respect to $\Delta_e u(x)$: namely the $(d-1)$-dimensional measure of the facets of (1.9) defined by the equality constraint $2|\langle g, e\rangle| = \Delta_e u(x)$, divided by $\|e\|$ (as a result, $\mathcal{D}_V u(x)$ is continuously differentiable in u, in contrast with (1.7)). Computing polytopes defined by linear inequalities like (1.9) is, by convex duality, equivalent to computing the convex envelope of a set of points [17]. Numerous computer libraries are available for that purpose, such as TetGen® or CGAL® (the author made his own routine which takes advantage of the symmetry of the polytope (1.9)).

The expression (1.9) seems more complex to evaluate than (1.6) or (1.7). Yet in our numerical experiments (Sect. 3), the cost of evaluating the operator $\mathcal{D}_V$ was dominated by the cost of solving linear systems (when solving (1.2) with a damped Newton solver), for which the MUMPS® library was used. The applicability of Newton's method to the Monge−Ampere problem is also investigated in [12], for the continuous problem, and in [8, 14], for some discretizations. Its super-linear rate of convergence, in the neighborhood of the problem solution, makes it appealing for applications. In contrast, gradient descent [1] or first order Euler [15] schemes converge slower but benefit from global convergence guarantees with monotone discretizations (iterates converge to the unique solution of the discrete problem, independently of initialization). For the discretization of interest, we prove Section 2.2 that a damped Newton method benefits from *both* a super-exponential rate of convergence close to the discrete problem solution, and a global convergence guarantee.

The following results, established Section 2.1, show that the consistency set of the proposed scheme is of non-empty interior, in contrast with scheme WS. By Corollary 1.8, a finite stencil is sufficient to guarantee consistency for all matrices $M \in S_d^+$ with condition number below a given bound.

Definition 1.6. For each matrix $M \in S_d^+$, we introduce the Voronoi cell and facet

$$\mathrm{Vor}(M) := \{g \in \mathbb{R}^d; \forall e \in \mathbb{Z}^d, \|g\|_M \leq \|g - e\|_M\},$$
$$\mathrm{Vor}(M; e) := \{g \in \mathrm{Vor}(M); \|g\|_M = \|g - e\|_M\}.$$

An M-Voronoi vector is an element $e \in \mathbb{Z}^d \backslash \{0\}$ such that $\mathrm{Vor}(M; e) \neq \emptyset$; it is said *strict* iff the facet $\mathrm{Vor}(M; e)$ is $(d-1)$-dimensional.
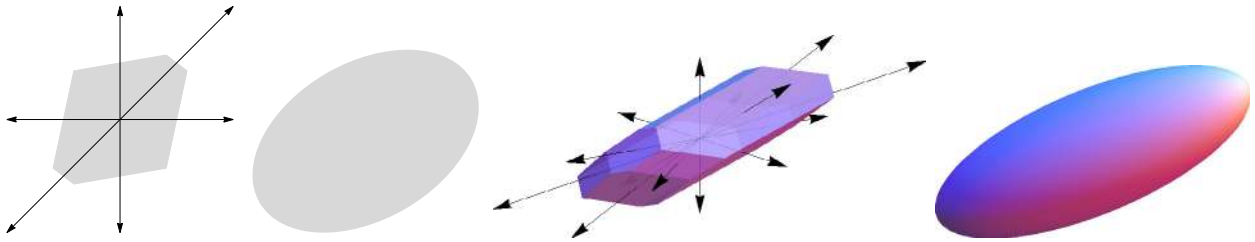
FIGURE 1. Voronoi cell $\text{Vor}(M)$, Voronoi vectors e, and ellipse $\{x \in \mathbb{R}^d; \langle x, Mx \rangle \leq 1\}$. *Left*: 2D, *Right*: 3D. Note that e traverses the facet $\text{Vor}(M; e)$ at the point $e/2$.

Proposition 1.7 (Consistency). *A matrix M is in the consistency set of $\mathcal{D}_V$ iff V contains all strict M-Voronoi vectors.*

Corollary 1.8 (Finite stencils are enough). *Let $\kappa \geq 1$ and let V collect all elements $e \in \mathbb{Z}^d$ with norm $\|e\| \leq \kappa\sqrt{d}$ and co-prime coordinates. Then the consistency set of $\mathcal{D}_V$ contains all $M \in S_d^+$ such that $\kappa \geq \sqrt{\|M\|\|M^{-1}\|}$.*

Outline: The results on consistency are established Section 2.3, and the global convergence of the damped Newton solver in Section 2.2. We also study in Section 2.3 convergence as the grid scale tends to zero, in the setting of viscosity solutions. Numerical experiments Section 3 illustrate the method's efficiency.

Remark 1.9 (Symmetrization). A non-symmetrical variant of our Monge−Ampere operator discretization (1.9) can be defined as follows: for $x \in X$ such that $x + e \in X$ for all $e \in V$

$$\mathcal{D}_V' u(x) := \text{Leb}\{g \in \mathbb{R}^d; \forall e \in V, \langle g, e \rangle \leq u(x+e) - u(x)\}. \tag{1.10}$$

In the case of a quadratic function u_M, $M \in S_d^+$, the polytope appearing in (1.10) is merely a translation of (1.9) by the vector Mx, so that $\mathcal{D}_V' u_M = \mathcal{D}_V u_M$. In the case of a general u, the polytope (1.10) can be non-symmetric, in contrast with (1.9). This asymmetry can actually help extract weak Alexandroff solutions of (1.1) with non-symmetric subgradient sets, such as the two Diracs test case in [2].

The drawback of the expression (1.10) is that it is only correctly defined sufficiently far from the boundary $\partial\Omega$. There is no simple way, to the knowledge of the author, to incorporate in (1.10) the values of u on $\partial\Omega$, and remain consistent with the Monge−Ampere operator $\det(\nabla^2 u)$ in the strong sense of Definition 1.4. Hence our choice of (1.9), defined in terms of the symmetric second order differences $\Delta_e u(x)$, $e \in V$.

Remark 1.10 (Localization). Consider the point dependent, largest possible stencil $V(x) = \{e \in \mathbb{Z}^d; x + e \in X\}$. Then $\mathcal{D}_{V(x)}' u(x)$, see (1.10), is the measure of the power cell associated to x in the power diagram based discretization [1, 11, 13, 16] of optimal transport. Since the stencils $(V(x))_{x \in X}$ are typically huge, it is not reasonable to evaluate $\mathcal{D}_{V(x)}' u(x)$ independently for each x. Instead a global power diagram needs to be constructed, using complex geometric procedures which validity requires extremely careful evaluations of geometric predicates [11].

2. PROOFS OF THE MAIN RESULTS

We establish in Section 2.1 the numerical scheme consistency, show in Section 2.2 the global convergence of a damped Newton solver for the discrete system, and prove in Section 2.3 a convergence result in the setting of viscosity solutions as the discretization grid scale tends to zero.

2.1. Consistency

Our consistency analysis relies on elementary properties of Voronoi cells of lattices, see Definition 1.6, and [6] for more on this topic. Our first step is to bound their diameter. Let $\kappa(M)^2$ denote the condition number of a matrix $M \in S_d^+$, *i.e.* $\kappa(M) := \sqrt{\|M\|\|M^{-1}\|}$.

Lemma 2.1. *Let $M \in S_d^+$. Any point $g \in \text{Vor}(M)$ satisfies $\|g\| \leq \frac{1}{2}\kappa(M)\sqrt{d}$. Any M-Voronoi vector e satisfies $\|e\| \leq \kappa(M)\sqrt{d}$, and has co-prime coordinates.*

Proof. Bound on the norm. Let $g \in \text{Vor}(M)$, and let e_g be obtained by rounding the coordinates of g to the nearest integer, so that $\|g - e_g\| \leq \frac{1}{2}\sqrt{d}$. One has $\lambda^-\|g\| \leq \|g\|_M \leq \|g - e_g\|_M \leq \lambda^+\|g - e_g\|$, where λ^-, λ^+ are respectively the smallest and largest eigenvalues of $M^{\frac{1}{2}}$. Thus $\|g\| \leq (\lambda^+/\lambda^-)\frac{1}{2}\sqrt{d}$ as announced. If e is an M-Voronoi vector, then there exists $g \in \text{Vor}(M)$ such that $\|g\| = \|e - g\|$, hence $\|e\| \leq 2\|g\| \leq \kappa(M)\sqrt{d}$.

Coordinates are co-prime. Consider a vector which coordinates are not co-prime: ke, with $k \geq 2$ and $e \in \mathbb{Z}^d \setminus \{0\}$. Then for any $g \in \mathbb{R}^d$ one has $\|ke - g\|_M^2 + (k-1)\|g\|_M^2 = k\|e - g\|_M^2 + (k^2 - k)\|e\|_M^2$. Hence $\|e - g\|_M < \max\{\|ke - g\|_M, \|g\|_M\}$ and therefore ke cannot be an M-Voronoi vector. $\qquad\square$

Corollary 2.2. *Let $M \in S_d^+$, and let E be the collection of strict M-Voronoi vectors. For any set $V \subseteq \mathbb{Z}^d$ one has $\text{Vor}(M) \subseteq \{g \in \mathbb{R}^d; \forall e \in V, 2\langle g, Me\rangle \leq \|e\|_M^2\}$, with equality iff $E \subseteq V$.*

Proof. For any $g, e \in \mathbb{R}^d$, we have the equivalence $\|g\|_M \leq \|g - e\|_M \Leftrightarrow 2\langle g, Me\rangle \leq \|e\|_M^2$, obtained by squaring both sides of the first inequality. Hence $\text{Vor}(M)$ is a convex polytope, defined by a family of linear inequalities indexed by $e \in \mathbb{Z}^d$. By Lemma 2.1 one can eliminate all inequalities but a finite number. Among these inequalities, only those corresponding to strict M-Voronoi vectors are active, in the sense that they define a facet of $\text{Vor}(M)$. The result follows. $\qquad\square$

We next identify the volume of a Voronoi cell, and use it to establish our scheme consistency.

Lemma 2.3. *For any $M \in S_d^+$ one has $\text{Leb}(\text{Vor}(M)) = 1$.*

Proof. The set $\text{Vor}(M)$ collects elements $g \in \mathbb{R}^d$ which are closer to the origin than to any other point $e \in \mathbb{Z}^d$. As a result the translates of $\text{Vor}(M)$, by all offsets $e \in \mathbb{Z}^d$, tile the space $\mathbb{R}^d$ up to a negligible set. Thus $\text{Leb}(\text{Vor}(M))$ is the co-volume of the lattice $\mathbb{Z}^d$, namely 1 as announced. $\qquad\square$

Proof of Proposition 1.7 (Consistency). Let $M \in S_d^+$, let $x \in X$, and let u_M be the quadratic map (1.8). By construction $\Delta_e u_M(x) = \|e\|_M^2$, for any $e \in \mathbb{Z}^d$. Hence $\mathcal{D}_V u_M(x)$ is the volume of:

$$\{g \in \mathbb{R}^d; \forall e \in V, 2\langle g, e\rangle \leq \|e\|_M^2\} = M\{g \in \mathbb{R}^d; \forall e \in V, 2\langle g, Me\rangle \leq \|e\|_M^2\} \supseteq M\text{Vor}(M).$$

where we used Corollary 2.2 and abusively denoted by $ME := \{Me; e \in E\}$ the image of a set E by the linear map M. Therefore $\mathcal{D}_V u_M(x) \geq \det(M)\text{Leb}(\text{Vor}(M)) = \det(M)$ by Lemma 2.3. Equality holds iff the above inclusion of polytopes is an equality, equivalently iff V contains all strict M-Voronoi vectors by Corollary 2.2. $\quad\square$

Proof of Corollary 1.8 (Finite stencils are enough). Combine Proposition 1.7 and Lemma 2.1. $\qquad\square$

2.2. Global convergence of a damped Newton solver

We establish the convergence of a damped Newton solver for the discrete system (1.2). Before doing so, we need to clarify the implementation of boundary conditions. The stencil V is fixed in the following, and obeys the properties of Definition 1.5.

Remark 2.4 (Boundary discretization). Maps $u \in \mathbb{U}$ are in principle, see Definition 1.1, defined both on the discrete sampling X of Ω, and on the whole uncountable boundary $\partial\Omega$. Fortunately, only finitely many values of u on $\partial\Omega$ play an active role in the system of equations (1.2): those on

$$\partial X := \{x + h_x^e e; \ x \in X, \ e \in V\} \cap \partial\Omega.$$

This discretization, with $\#(X \cup \partial X)$ unkowns, is used in our experiments, and in the proof of the damped Newton solver global convergence below.

Alternatively one may (i) not introduce any unknown for the boundary values, but use the given Dirichlet data σ as in [3], or (ii) introduce some unknowns on an arbitrary boundary sampling $\partial X'$, extended to $\partial\Omega \backslash \partial X'$ by some interpolation procedure. Unfortunately (i) leads to initialization difficulties for the iterative solver, since one needs to find a strictly convex seed u_0 with prescribed boundary values on $\partial\Omega$, and (ii) may limit the accuracy of the scheme if first order interpolation is used, or violate its degenerate ellipticity in the case of high order interpolation.

Our first step is to establish, in Corollary 2.6, that the Jacobian matrix associated to the discrete Monge–Ampere system (1.2) is invertible. This follows from the invertibility of diagonally dominant matrices, recalled in the next lemma, and from the degenerate ellipticity of the proposed operator $\mathcal{D}_V$.

Lemma 2.5. *Let A be an $n \times n$ matrix such that for each $1 \le i \le n$*

$$|A_{ii}| \ge \sum_{j \neq i} |A_{ij}|. \tag{2.1}$$

Assume that for each index $1 \le i_0 \le n$ there exists a chain $i_0, \ldots, i_k$ such that i_k satisfies strictly *inequality (2.1), and $A_{i_r, i_{r+1}} \neq 0$ for all $0 \le r < k$. Then A is invertible.*

Proof. For each index $1 \le i \le n$, let $k(i)$ denote the length of the smallest chain of indices as above. Let $x \in \mathbb{R}^d$ be such that $Ax = 0$. Among the indices $1 \le i \le n$ such that the vector component $|x_i|$ is maximal, choose one which minimizes $k(i)$. From $(Ax)_i = 0$ we obtain

$$|A_{ii}||x_i| \le \sum_{j \neq i} |A_{ij}||x_j|, \quad \text{thus} \quad \left(|A_{ii}| - \sum_{j \neq i} |A_{ij}|\right)|x_i| + \sum_{j \neq i}|A_{ij}|(|x_i| - |x_j|) = U + V \le 0.$$

Both terms U and V are non-negative by construction, hence $U = V = 0$. If $k(i) = 0$, then from $U = 0$ we obtain $|x_i| = 0$, hence $x = 0$. If $k(i) \neq 0$, then from $V = 0$ we obtain $|x_j| = |x_i|$ for all indices $1 \le j \le n$ such that $A_{ij} \neq 0$. One of these indices satisfies $k(j) = k(i) - 1$, which contradicts our choice of i. This concludes the proof. $\square$

Let $\mathbb{U}_0$ be the collection of maps $u : X \cup \partial X \to \mathbb{R}$ such that:

$$\forall x \in X, \ \mathcal{D}_V u(x) > 0. \qquad (\text{Equivalently: } \forall x \in X, \ \forall e \in V, \ \Delta_e u(x) > 0.) \tag{2.2}$$

Let $f : \mathbb{U}_0 \to \mathbb{R}^{X \cup \partial X}$ be defined by

$$f(u)(x) := \begin{cases} \ln \mathcal{D}_V u(x) & \text{if } x \in X, \\ u(x) & \text{if } x \in \partial X. \end{cases} \tag{2.3}$$

Corollary 2.6. *For each $u \in \mathbb{U}_0$, the Jacobian matrix $df(u)$ is invertible.*

Proof. Let $u \in \mathbb{U}_0$ and let $A := df(u)$, which formally is a matrix associating a coefficient to each index pair (x, y), $x, y \in X \cup \partial X$. The line of A corresponding to each $x \in \partial X$ has a single non-zero coefficient, namely 1 at index (x, x), hence this line satisfies (2.1) strictly. By degenerate ellipticity of $\mathcal{D}_V$, the line corresponding to any $x \in X$ satisfies (2.1). We prove in the following the chain property of Lemma 2.5, which implies the announced invertibility.

Let $x \in X$, and let K_x be the polytope appearing in (1.9), which by construction is convex, compact and symmetric w.r.t the origin. For each $e \in V$ let F_e denote the facet of K_x defined by the equality constraint $2\langle g, e \rangle = \Delta_e u(x)$. Let $V_x := \{e \in V; \Lambda_e > 0\}$ where Λ_e denotes the $(d-1)$-dimensional measure of F_e. Since the polytope is symmetric, one has $\Lambda_e = \Lambda_{-e}$ for all $e \in V$, hence V_x is symmetric. Since the polytope is compact, and since the exterior normal to F_e is $e/\|e\|$, the set V_x spans $\mathbb{R}^d$. The coefficient of the Jacobian matrix A at index $(x, x + h_x^e e)$ is $4\Lambda_e/(h_x^e(h_x^e + h_x^{-e})\|e\|\mathrm{Leb}(K_x))$, using (1.5) and the geometric argument that the polytope volume variation is at first order given by the facets areas $\Lambda_e + \Lambda_{-e}$ times their normal displacement. This coefficient is hence positive if $e \in V_x$.

Among the points $x + h_x^e e$, $e \in V_x$ one at least is strictly closer to $\partial \Omega$ than x. By induction we can thus build a finite chain $x = x_0, \ldots, x_{k-1} \in X$ such that $x_k \in \partial X$ and the coefficient of A at index (x_r, x_{r+1}) is non-zero for each $0 \leq r < k$. The announced result then follows from Lemma 2.5. □

Our second step is to show that (2.3) is a proper map: the preimage of any compact set is a compact set, a property which is tightly linked with the well-posedness of the PDE (1.1). Let

$$\partial_x U := \{g \in \mathbb{R}^d; \forall y \in \overline{\Omega}, \langle g, y - x \rangle \leq U(y) - U(x)\} \tag{2.4}$$

denote the subgradient of a convex map $U \in C^0(\overline{\Omega}, \mathbb{R})$ at a point $x \in \Omega$. We connect in Lemma 2.7 the proposed operator $\mathcal{D}_V u(x)$ with the Lebesgue measure of $\partial_x U$, where U is the lower convex envelope of u. The properness of (2.3) then follows in Corollary 2.9 from the maximum principle of Alexandroff–Bakelman–Pucci.

Lemma 2.7. *Let $u \in \mathbb{U}_0$, and let $U : \mathrm{Hull}(X \cup \partial X) \to \mathbb{R}$ be the maximal convex map bounded above by u. Then $\mathrm{Leb}(\partial_x U) \leq h_*^d \mathcal{D}_V u(x)$ for all $x \in X$, with $h_* := \max\{h_x^e; x \in X, e \in V\}$.*

Proof. By construction $U \leq u$, and for any $x \in X$ such that $U(x) < u(x)$ one has $\mathrm{Leb}(\partial_x U) = 0$, so that the announced inequality holds. Assume that $U(x) = u(x)$, and let $g^+, g^- \in \partial_x U$. Let $e \in V$ and let $h^+ := h_x^e$, $h^- := h_x^{-e}$ as in (1.5). One has $\langle g^+, h^+ e \rangle \leq u(x + h^+ e) - u(x)$, and $\langle g^-, -h^- e \rangle \leq u(x - h^- e) - u(x)$ by (2.4), thus

$$\langle g^+ - g^-, e \rangle \leq \frac{u(x + h^+ e) - u(x)}{h^+} + \frac{u(x - h^- e) - u(x)}{h^-} = \frac{h^+ + h^-}{2}\Delta_e u(x) \leq h_* \Delta_e u(x).$$

Therefore $\frac{1}{2}(\partial_x U - \partial_x U) \subseteq h_* K_x$, where the left hand side is a Minkowski sum of sets, and K_x is the polytope appearing in (1.9). The announced estimate then follows from this inclusion and Brunn−Minkowski's inequality. □

Theorem 2.8 (Alexandroff–Bakelman–Pucci's maximum principle [10]). *Let $u \in C^0(\overline{\Omega})$, and let U be the maximal convex map bounded above by u. Then*

$$\min_{\partial \Omega} u - \min_{\overline{\Omega}} u \leq (\mathrm{Leb}(G)/\omega_d)^{\frac{1}{d}} \mathrm{diam}(\Omega), \qquad \text{with } G := \bigcup_{\substack{x \in \Omega \\ u(x) = U(x)}} \partial_x U, \tag{2.5}$$

where $\mathrm{diam}(\Omega) := \max\{\|x - y\|; x, y \in \Omega\}$, and ω_d is the volume of the d-dimensional unit ball.

Corollary 2.9. *The map $f : \mathbb{U}_0 \to \mathbb{R}^{X \cup \partial X}$ is proper.*

Proof. Since f is continuous, it suffices to bound by above and below the values of an arbitrary $u \in \mathbb{U}_0$ in terms of those of $f(u)$. We have the obvious estimate $f(u)(x) = u(x)$ for values at $x \in \partial X$.

Lower bound on X. Let Ω' be the interior of $\mathrm{Hull}(X \cup \partial X)$, and let $u' \in C^0(\overline{\Omega}', \mathbb{R})$ be defined by $u'(z) = \min_{x \in X \cup \partial X} u(x) + \lambda \|x - z\|$, where $\lambda > 0$. Let $U \in C^0(\overline{\Omega}', \mathbb{R})$ be the maximal convex function bounded above by u'. If the constant λ is sufficiently large, then U also is the maximal convex function bounded above by u. In addition for all $x \in \Omega'$, one has $U(x) = u'(x) \Leftrightarrow x \in X \cup \partial X$. We obtain applying Theorem 2.8 the desired lower bound on $\min\{u(x); x \in X\}$, since

$$\min_{\overline{\Omega}'} u' = \min_{X \cup \partial X} u, \qquad \min_{\partial \Omega'} u' = \min_{\partial X} u, \qquad \mathrm{Leb}(G) = \sum_{x \in X} \mathrm{Leb}(\partial_x U) \leq h_*^d \sum_{x \in X} \mathcal{D}_V u(x) = h_*^d \sum_{x \in X} e^{f(u)(x)}.$$

Upper bound on X. Let $x \in X \cup \partial X$ be such that $u(x)$ is maximal, and let us assume for contradiction that $x \notin \partial X$. Since $\Delta_e u(x) > 0$, for any $e \in V$, one has either $u(x + h^+e) > u(x)$, or $u(x - h^-e) > u(x)$, with the notations of (1.5). This contradicts the maximality of $u(x)$, and concludes the proof. $\qquad\square$

Numerous variants of the Newton method exist [7]; an elementary one is presented below for completeness, which guarantees global convergence for the system (1.2) of interest. This damped Newton algorithm is an iterative equation solver, which recursion rule (2.6) involves an adaptively chosen "damping" parameter δ. In practice δ is typically small in the first iterations, and equal to 1 in the last ones, which coincide with the classical Newton method and enjoy quadratic convergence. Convergence is guaranteed for maps which, like (1.5), are shown to be proper and at each point a local diffeomorphism. Note that these assumptions, plus the connectedness of the source domain and the simple connectedness of the target domain (here both satisfied), imply by Hadamard−Levy's theorem that f is a global diffeomorphism.

Proposition 2.10 (Global convergence of the damped Newton algorithm)**.** *Let $N > 0$, let $\mathbb{U}_0 \subseteq \mathbb{R}^N$ be an open set, and let $f \in C^1(\mathbb{U}_0, \mathbb{R}^N)$. Assume that f is proper and that the Jacobian matrix $F(x) := df(x)$ is invertible for each $x \in \mathbb{U}_0$. Consider $y \in \mathbb{R}^N$, and for each $x \in \mathbb{U}_0$, $\delta \in [0, 1]$, the Newton update*

$$\mathrm{Newton}(x, \delta) := x + \delta F(x)^{-1}(y - f(x)). \tag{2.6}$$

Let $x_0 \in \mathbb{U}_0$, and for each $n \geq 0$ let $x_{n+1} := \mathrm{Newton}(x_n, \delta_n)$, where $\delta_n = 2^{-k_n}$ and k_n is the smallest non-negative integer such that $\|y - f(x_{n+1})\| \leq (1 - \delta_n/2)\|y - f(x_n)\|$. Then $f(x_n) \to y$ as $n \to \infty$.

Proof. Introduce the set $K := \{x \in \mathbb{U}_0; \|y - f(x)\| \leq \|y - f(x_0)\|\}$, which is compact since f is proper. For any $x \in \mathbb{U}_0$ one has the Taylor expansion: for small δ

$$\begin{aligned} f(\mathrm{Newton}(x, \delta)) &= f(x) + \delta F(x)F(x)^{-1}(y - f(x)) + o(\delta\|F(x)^{-1}(y - f(x))\|) \\ &= y - (1 - \delta)(y - f(x)) + o(\delta\|y - f(x)\|). \end{aligned} \tag{2.7}$$

Hence $\|y - f(\mathrm{Newton}(x, \delta))\| = (1 - \delta + o(\delta))\|y - f(x)\|$ using (2.7), which is smaller than $(1 - \delta/2)\|y - f(x)\|$ for an open range of $\delta \in]0, \lambda(x)[$. By compactness, this property holds for all $x \in K$ with an uniform open range $]0, \lambda[$.

Thus δ_0 is well defined, bounded below by $\lambda/2$, and by construction $\|y - f(x_1)\| \leq (1 - \lambda/2)\|y - f(x_0)\|$ so that $x_1 \in K$. The result follows from an immediate induction argument. $\qquad\square$

2.3. Convergence in the setting of viscosity solutions

In this section, we let the discretization grid scale tend to zero, and study the convergence of the minimizers to the discrete problems (1.2). For all integers $n \geq 1$, let $X_n := \Omega \cap n^{-1}\mathbb{Z}^d$. Let $\mathbb{U}_n$ be the collection of

(semi-)discrete maps $u : X_n \cup \partial\Omega \to \mathbb{R}$. The stencil V is fixed and obeys the assumptions of Definition 1.5. For any $u \in \mathbb{U}_n$, $x \in X_n$, and $M \in S_d$ let

$$\mathcal{D}_n(u) := \mathrm{Leb}\{g \in \mathbb{R}^d; \forall e \in V, \ 2\langle g, e\rangle \leq \Delta_{e/n}u(x)\},$$
$$D(M) := \mathrm{Leb}\{g \in \mathbb{R}^d; \forall e \in V, \ 2\langle g, e\rangle \leq \langle e, Me\rangle\}. \tag{2.8}$$

We denoted by $\Delta_{e/n}u(x) := n^2(u(x + e/n) - 2u(x) + u(x - e/n))$ the standard approximation of $\langle e, (\nabla^2 u(x))e\rangle$ on the grid X_n (This expression assumes that $x \pm e/n \in X_n$, and should be modified as in (1.5) otherwise). Given a density $\rho \in C^0(\overline{\Omega}, \mathbb{R}^*_+)$, and some Dirichlet data $\sigma \in C^0(\partial\Omega, \mathbb{R})$, we study the problems

$$\begin{cases} u \in \mathbb{U}_n, \\ \mathcal{D}_n(u) = \rho & \text{on } \Omega, \\ u = \sigma & \text{on } \partial\Omega. \end{cases} \qquad \begin{cases} u : \overline{\Omega} \to \mathbb{R}, \\ D(\nabla^2 u) = \rho & \text{on } \Omega, \\ u = \sigma & \text{on } \partial\Omega. \end{cases} \tag{2.9}$$

By Section 2.2, the discretized problem (2.9, left) has a unique solution $u_n \in \mathbb{U}_n$. We show in Corollary 2.15 that (2.9, right) also admits a unique solution u_∞. Importantly, if $\nabla^2 u_\infty$ exists at each $x \in \Omega$ and belongs to the consistency set of $\mathcal{D}_V$, then u_∞ is also the unique solution to the Monge−Ampere equation (1.1). The study of (2.9, right) relies on the concept of viscosity solutions [5].

Definition 2.11. A sub-solution (resp. super-solution) of (2.9), right is an upper-semi-continuous (resp. lower-semi-continuous) map $u : \overline{\Omega} \to \mathbb{R}$ such that (I) $u \leq \sigma$ (resp. $u \geq \sigma$) on $\partial\Omega$ and (II) for any $x \in \Omega$ and any $\varphi \in C^2(\Omega)$ such that $u - \varphi$ has a local maximum (resp. minimum) at x, one has $D(\nabla^2\varphi(x)) \geq \rho(x)$ (resp. $\leq \rho(x)$).

A solution to (2.9, right) is a map $u : \overline{\Omega} \to \mathbb{R}$ which is both a super-solution and a sub-solution.

Replacing "local maximum" with "strict global maximum" (resp. minimum) in Definition 2.11 yields an equivalent notion of sub- and super-solutions [5]. Super- and sub-solutions of (2.9, right) obey a comparison principle, proved Proposition 2.13. Given symmetric matrices $M, M' \in S_d$, we write $M \preceq M'$ iff the difference $M' - M$ is positive semi-definite.

Lemma 2.12. *For any $M, M' \in S_d$, one has: $M \preceq M' \Rightarrow D(M) \leq D(M')$. More quantitatively, if $M, H \in S_d$ are such that $D(M) > 0$ and $D(H) > 0$ then*

$$D(M + H)^{\frac{1}{d}} \geq D(M)^{\frac{1}{d}} + D(H)^{\frac{1}{d}}.$$

Proof. Let $K(M) \subseteq \mathbb{R}^d$ be the polytope appearing in (2.8), so that $D(M) = \mathrm{Leb}K(M)$. Then

$$M \preceq M' \Rightarrow \forall e \in V, \|e\|_M^2 \preceq \|e\|_{M'}^2 \Rightarrow K(M) \subseteq K(M') \Rightarrow D(M) \leq D(M').$$

Second point: since $\langle e, (M + H)e\rangle = \langle e, Me\rangle + \langle e, He\rangle$, the set $K(M + H)$ contains the Minkowski sum $K(M) + K(H)$. The announced result follows from Brunn−Minkowski's inequality. □

Proposition 2.13. *If u_* is a sub-solution of (2.9, right), and u^* a super-solution, then $u_* \leq u^*$.*

Proof. The result does not immediately follow from [5] because the operator $-D(\nabla^2 u)$ is only degenerate elliptic. The following operator is in contrast strictly elliptic when $\varepsilon > 0$:

$$F_\varepsilon(u) := -D(\nabla^2 u)^{\frac{1}{d}} + \varepsilon u. \tag{2.10}$$

Let $M := \sup_{\overline{\Omega}} u_* - u^*$, which is finite by upper-semi-continuity. For contradiction, we assume $M > 0$. Let also $r := \max\{\|x\|; x \in \overline{\Omega}\}$ and

$$u_\varepsilon(x) := u_*(x) + \frac{\varepsilon M}{2}(\|x\|^2 - r^2). \tag{2.11}$$

One has $u_\varepsilon \leq u_* \leq u^*$ on $\partial\omega$, and by Lemma 2.12

$$F_\varepsilon(u^*) - F_\varepsilon(u_\varepsilon) \geq \varepsilon M - \varepsilon(u_* - u^*) \geq 0. \tag{2.12}$$

Applying the standard comparison principle [5] to the strictly elliptic F_ε we obtain $u^* \geq u_\varepsilon$, hence $u^* - u_* \geq -\varepsilon M r^2/2$. Letting $\varepsilon \to 0$ yields $u^* \geq u_*$ as announced. $\qquad\square$

Weal solutions can be extracted from sequences of discrete solutions: for each $n \geq 1$, extend u_n by bilinear interpolation (or any other local interpolation procedure) on $\Omega_n := \{x \in \Omega; \, d(x, \partial\Omega) \geq n^{-1}\sqrt{d}\}$, and define $u_*, u^* : \overline{\Omega} \to \mathbb{R}$ by

$$u_*(x) := \limsup_{n\to\infty} u_n(x), \qquad\qquad u^*(x) := \liminf_{n\to\infty} u_n(x). \tag{2.13}$$

Lemma 2.14. *u_* and u^* are respectively a sub-solution and a super-solution of* (2.9), *right.*

Proof. Inspection of the proof of Corollary 2.9 yields the quantitive estimate

$$\min_{\partial\Omega} \sigma - \left(\frac{1}{n^d} \sum_{x \in X_n} \frac{\rho(x)}{\omega_d}\right)^{\frac{1}{d}} \operatorname{diam}(\Omega) \leq u_n \leq \max_{\partial\Omega} \sigma. \tag{2.14}$$

Note that $n^{-d} \sum_{x \in X_n} \rho(x) \to \int_\Omega \rho$ as $n \to \infty$. (Also, the multiplicative term h_* appearing in Cor. 2.9 equals 1 with our choice of discrete domain $X_n = \Omega \cap n^{-1}\mathbb{Z}^d$, since $]x, x + e/n] \cap (X_n \cup \partial\Omega)$ is non-empty for all $x \in X_n$, $e \in \mathbb{Z}^d$.)

The maps $u_n \in \mathbb{U}_n$ are therefore bounded independently of n, hence u_* and u^* are well defined. From this point, the announced result follows by a standard argument: let $x \in \Omega$ and $\varphi \in C^2(\Omega)$ be such that $u^* - \varphi$ attains a strict global maximum at $x \in \Omega$. For each $n \geq 1$, let x_n be the element of X_n which maximizes $u_n - \varphi$. By monotonicity $\mathcal{D}_n\varphi(x_n) \geq \mathcal{D}_n u_n(x_n) = \rho(x_n)$. Observing that $x_n \to x$ as $n \to \infty$, we obtain $D(\nabla^2\varphi(x)) \geq \rho(x)$ in the limit. This establishes that u^* is a super-solution, and likewise u_* is a sub-solution. $\qquad\square$

Corollary 2.15. *One has $u_* = u^*$, and this map is the unique solution u_∞ of* (2.9), *right. As a result, u_n converges pointwise to u_∞ as $n \to \infty$.*

Proof. By construction $u_* \leq u^*$, but by the comparison principle $u_* \geq u^*$ (see Prop. 2.13). Therefore (2.9), *right*, admits the solution $u_\infty := u_* = u^*$ which, again by the comparison principle, is its unique solution. Finally, for any $x \in \Omega$ on has as $n \to \infty$: $\limsup u_n(x) = u^*(x) = u_*(x) = \liminf u_n(x)$, hence $u_\infty(x) = \lim u_n(x)$. $\qquad\square$

3. Numerical experiments

We implemented the three[2] Monge Ampere discretizations described in the introduction: the Finite Differences (FD) scheme $\mathcal{D}^{\mathrm{FD}}$, the Wide Stencil (WS) scheme $\mathcal{D}^{\mathrm{WS}}_{\mathcal{B}}$, and the proposed scheme $\mathcal{D}_V$. All presented experiments are three dimensional. The latter two schemes require choosing a collection $\mathcal{B}$ of triplets of orthogonal vectors, or a stencil V, see Table 1. We recall that scheme FD is consistent but not monotone (or degenerate elliptic). Scheme WS is monotone and thus benefits from the associated convergence guarantees, but suffers from significant consistency errors, see Figure 2. The proposed scheme is simultaneously monotone and consistent, provided the PDE solution hessian condition number is bounded, and the scheme stencil V is sufficiently large (see Prop. 1.7 and Cor. 1.8). A consistency error arises when these conditions are not satisfied, see Figure 2.

We limit our attention to synthetic test cases, posed on the unit cube $\Omega :=]0, 1[^3$. A known convex function $U : \overline{\Omega} \to \mathbb{R}$ is numerically recovered from its hessian determinant $\rho := \det(\nabla^2 U)$, and its boundary values $\sigma := U_{|\partial\Omega}$. Three test cases are considered.

---

[2]The filtered method [9], which accuracy may be competitive, was not implemented due to lack of time.
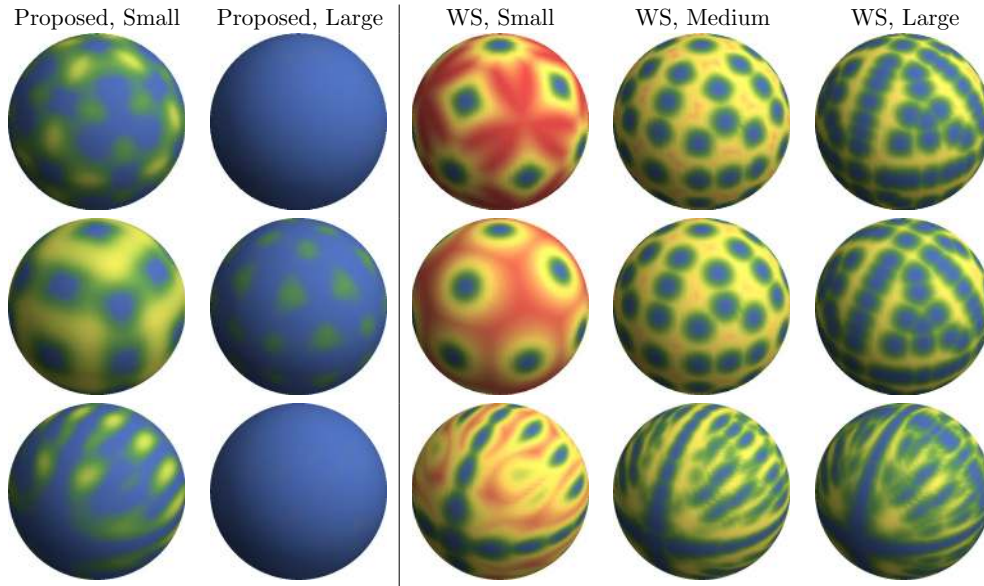
FIGURE 2. Relative consistency error $(\mathcal{D}u_M - \det(M))/\mathcal{D}u_M$ (blue $= 0$, red $= 1$) for a quadratic (1.8) map u_M associated to a matrix $M = M(v)$, with v in the unit sphere (hence the spherical plot). First line: $M(v)$ has eigenvalues $6^2, 1, 1$, the former with eigenvector v. Second line likewise with eigenvalues $6^{-2}, 1, 1$. Third line $M(v) = R(v)DR(v)$ where $R(v)$ is the rotation of axis v and angle π, and D is diagonal of entries $6, 1, 1/6$. (Color online).
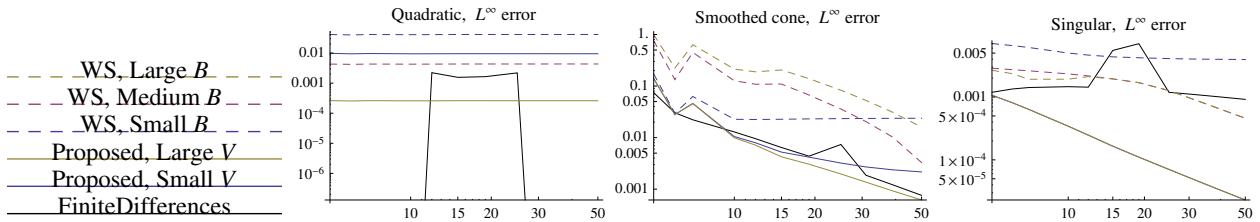

FIGURE 3. Log-Log plot of the L^∞ error, as a function of resolution n, for the three test cases. Discretization set $X \subseteq \Omega$ has n^3 points. Scheme stencils in Figure 1.

- (Quadratic) $U(x) := \frac{1}{2}\langle x, Mx\rangle$, where M was chosen randomly, with $\sqrt{\|M\|\|M^{-1}\|} \approx 8.5$.
- (Smoothed cone) $U(x) := \sqrt{\delta^2 + \|x - x_0\|^2}$, with $\delta := 0.1$ and $x_0 := (1/2, 1/2, 1/2)$.
- (Singular, [9]) $U(x) := -\sqrt{3 - \|x\|^2}$.

A Damped Newton solver is applied to the discrete system (1.2), starting from the trivial seed $u(x) := \|x\|^2$.

Quadratic test case. The chosen matrix M does not belong to the consistency set of schemes $\mathcal{D}_V$ and $\mathcal{D}_{\mathcal{B}}^{\text{WS}}$, with the chosen V and $\mathcal{B}$; hence the numerical error reflects their consistency error, and is resolution independent. On the topic of consistency, it was determined using Proposition 1.7 and some semi-definite programming that the consistency set of the proposed scheme $\mathcal{D}_V$ contains all $M \in S_3^+$ such that $\text{Tr}(M)/(\det M)^{\frac{1}{3}}$ is less than 7.8 with the small stencil V, and less than 11.9 with the large V, see Table 1. The always consistent but non-monotone scheme $\mathcal{D}^{\text{FD}}$ finds the exact solution for a range of resolutions, up to machine precision, but switches to a completely erroneous solution at other resolutions.

Smoothed Cone test case. The test function U is smooth, but the test is not as simple as it looks since (i) ∇U varies quickly close to the center x_0, but (ii) $\nabla^2 U$ is almost degenerate far from x_0. Point (i) favors small stencils, while point (ii) requires a good angular resolution. Scheme WS is most efficient with a small triplets collection $\mathcal{B}$ at low resolutions, but with a medium or large one at high resolutions. Picking a stencil V is easier with the proposed scheme: the larger one here always yields a smaller error, albeit at a higher computational cost. The non-monotone scheme FD passes this test as well. We give some computation times for this test, which are typical of our experience. On a 2.7 Ghz Laptop using a single core, using a $50 \times 50 \times 50$ discretization grid X, computations took (in minutes): 4.7 and 9.7 with the proposed scheme, small and large V respectively; 6.4, 24 and 65 with scheme WS, small medium and large $\mathcal{B}$ respectively; 6.5 with scheme FD.

Singular test case. The function U is not smooth at the point $(1, 1, 1)$, where its gradient is formally $(+\infty, +\infty, +\infty)$. The non-monotone scheme FD completely fails this test: numerical error does not decrease as resolution increases, and initializing the Newton solver with the exact solution U does not even help, as observed in [9]. Scheme WS will recover the solution, if the bases collection $\mathcal{B}$ is large enough, but numerical error decays quite slowly. The proposed scheme works flawlessly: it has the same convergence guarantees as scheme WS, but yields an L^∞ error about 20 times smaller on a 50^3 grid. Curiously, numerical error is identical with the small and the large stencil V.

References

[1] F. Aurenhammer, F. Hoffmann and B. Aronov, Minkowski-type theorems and least-squares clustering. *Algorithmica* **20** (1998) 61–76.

[2] J.D. Benamou and B.D. Froese, A viscosity framework for computing Pogorelov solutions of the Monge−Ampere equation. Preprint (2014) ArXiv:1407.1300.

[3] J.-D. Benamou, F. Collino and J.-M. Mirebeau, Monotone and Consistent discretization of the Monge−Ampere operator Preprint (2014) ArXiv:1409.6694.

[4] S.C. Brenner and M. Neilan, Finite element approximations of the three dimensional Monge-Ampère equation. *ESAIM: M2AN* **46** (2012) 979–1001.

[5] M.G. Crandall, H. Ishii and P.-L. Lions, User's guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc.* **27** (1992) 1–67.

[6] J.H. Conway and N.J.A. Sloane, Low-Dimensional Lattices. VI. Voronoi Reduction of Three-Dimensional Lattices. *Proc. Roy. Soc. A: Math., Phys. Eng. Sci.* **436** (1992) 55–68.

[7] S.C. Eisenstat and H.F. Walker, Globally convergent inexact Newton methods. *SIAM J. Optim.* **4** (1994) 393–422.

[8] 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) 818–834.

[9] B.D. Froese and A.M. Oberman, Convergent filtered schemes for the Monge–Ampère partial differential equation. *SIAM J. Numer. Anal.* **51** (2013) 423–444.

[10] C.E. Gutiérrez, The Monge-Ampère Equation. In *Progr. Nonlin. Differ. Eqs. Appl.* Springer (2001).

[11] B. Lévy, A numerical algorithm for L_2 semi-discrete optimal transport in 3D. To appear in *ESAIM: M2AN* (2015). Doi:10.1051/m2an/2015055.

[12] G. Loeper and F. Rapetti, Numerical solution of the Monge-Ampère equation by a Newton's algorithm. *C. R. Math. Acad. Sci. Paris* (2005).

[13] Q. Merigot, A Multiscale approach to optimal transport. *Computer Graphics Forum* **30** (2011) 1583–1592.

[14] M. Neilan, Quadratic finite element approximations of the Monge-Ampère equation. *J. Sci. Comput.* **54** (2012) 200–226.

[15] A.M. Oberman, Convergent difference schemes for degenerate elliptic and parabolic equations: Hamilton–Jacobi equations and free boundary problems. *SIAM J. Numer. Anal.* **44** (2006) 879–895.

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

[17] F.P. Preparata and M. Shamos, Computational Geometry. Springer (2012).