Functions with Bounded Hessian–Schatten Variation: Density, Variational, and Extremality Properties

Luigi Ambrosio, Camillo Brena & Sergio Conti

Communicated by G. Dal Maso

Abstract

In this paper we analyze in detail a few questions related to the theory of functions with bounded p-Hessian–Schatten total variation, which are relevant in connection with the theory of inverse problems and machine learning. We prove an optimal density result, relative to the p-Hessian–Schatten total variation, of continuous piecewise linear (CPWL) functions in any space dimension d, using a construction based on a mesh whose local orientation is adapted to the function to be approximated. We show that not all extremal functions with respect to the p-Hessian–Schatten total variation are CPWL. Finally, we prove the existence of minimizers of certain relevant functionals involving the p-Hessian–Schatten total variation in the critical dimension $d = 2$.

Contents

Introduction

In applied sciences, an inverse problem is the process of reconstructing an unknown signal (in practice, a causal factor) from a set of (possibly noisy) observations. Of notable importance is the subclass of linear inverse problems. A linear inverse problem is specified by three main objects:

- an hypothesis space $\mathcal{S}$ in which we look for the signal $f^* \in \mathcal{S}$,
- a linear forward operator $\nu : \mathcal{S} \to \mathbb{R}^N$ which is a proxy for the process of data collection,
- an array $y \in \mathbb{R}^N$ which contains the observed data, to which $\nu(f^*)$ should be close.

Therefore, the inverse problem consists in (approximately) recovering the unknown signal f^* from the observed data y. Also, the problem can be reformulated in variational terms as

$$f^* \in \underset{f \in \mathcal{S}}{\operatorname{argmin}} \, \lambda \mathcal{E}\left(\nu(f), y\right) + \mathcal{R}(f), \tag{0.1}$$

where

- $\mathcal{E} : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ is a convex loss function used to measure the data discrepancy,
- $\mathcal{R} : \mathcal{S} \to \mathbb{R}$ is the regularization functional, mainly used to enforce known structure and regularity on the reconstructed signal,
- $\lambda > 0$ is a parameter that governs the interplay between fidelity to the data and regularity.

Three key effects of the presence of the regularization functional $\mathcal{R}$ are: the enhancement of the stability of the problem, the alleviation of the ill-posedness of the problem and the possibility to invoke the "representer theorem", which provides a parametric form for solutions of (0.1) and has been recently extended to rather general frameworks, see [8,9,19,20]. Roughly (and under suitable assumptions), the just-mentioned abstract result characterizes the set of solutions of (0.1) as linear combinations of the extreme points of

$$\{f \in \mathcal{S} : \mathcal{R}(f) \leqq 1\} \tag{0.2}$$

(which is the unit ball associated to the regularization functional). This strongly motivates the interest in finding and studying the extreme points of the set in (0.2).

In this paper, we are going to study problems arising from a particular, yet general, choice of the items appearing in the functional in (0.1). In particular,

a) the hypothesis space are the functions $f : \Omega \to \mathbb{R}$ with bounded p-Hessian–Schatten variation (see item b)), for some $\Omega \subseteq \mathbb{R}^d$ open. The space coincides indeed with Demengel's space ([11]) of functions with bounded Hessian, which has been introduced to study models of plastic deformations of solids and has proven useful also in the context of image processing, but the norm we adopt is specific and allows for optimal approximation results by continuous and piecewise affine functions when $p = 1$;

b) the regularizing term is the p-Hessian–Schatten variation $|D_p^2 \cdot |(\Omega)$, that co-incides with the relaxation of the functional (here and after $| \cdot |_p$ denotes the p-Schatten norm),

$$|D_p^2 f|(\Omega) := \int_\Omega |\nabla^2 f|_p \mathrm{d}\mathscr{L}^d \quad \text{for every } f \in C^2(\Omega);$$

This is a variant of the classical second-order total variation ([5]). It has been inspired by [6,14–17] and used in [10,18];

c) in the critical case $d = 2$ we consider as linear forward operator the evaluation functional at certain points $x_1, \ldots, x_N \in \mathbb{R}^2$, with observed data $(y_1, \ldots, y_N) \in \mathbb{R}^N$;

d) still in the critical case, the error term is taken to be an ℓ^q norm, i.e.

$$\mathcal{E}(f) := \|(f(x_i) - y_i)_{i=1,\ldots,N}\|_{\ell^q}.$$

e) the tunable parameter is $\lambda \in (0, \infty]$, where by convention $\lambda = \infty$ imposes a perfect fit with the data.

In view of the discussion above, it is evident that some questions arise as natural.

i) The description of the extremal points of the ball (cf. (0.2))

$$\{f : \Omega \to \mathbb{R} : |D_p^2 f|(\Omega) \leqq 1\} \tag{0.3}$$

modulo additive affine functions (since the Hessian–Schatten seminorm is in-variant under the addition of affine functions, this factorization is necessary). A reasonable description of these extremal points was given in [2], under the assumption that a certain density conjecture holds true. Namely, it has been proved that if CPWL functions are dense in energy in the space of functions with bounded Hessian–Schatten variation, then all extremal points, which ob-viously are on the sphere, are found in the closure of the CPWL extremal points (and this last set is rather manageable, see [2]). Here and below, a CPWL (Con-tinuous and PieceWise Linear) function is a piecewise affine function, affine on certain simplexes. In Section 2 we give a positive answer to the just mentioned conjecture, proved only in the two-dimensional case in [2] with a different, more constructive, strategy. As any CPWL function can be exactly represented by a neural network with rectified linear unit (ReLU) activation functions [4], our result (Theorem 2.4) in particular implies approximability of any function whose Hessian has bounded total variation by means of neural networks with ReLU activation functions, with convergence of the 1-Hessian–Schatten norm.

ii) Again with respect to the extremal points of the set described in (0.3), one may wonder whether all the extremal points are CPWL. By a delicate measure-theoretic analysis, in Section 3 we show that the answer is negative: functions whose graphs are cut cones are extremal, modulo affine functions, and these functions are not CPWL if $d \geqq 2$. In connection with this negative answer, as for compact convex sets exposed points are dense in the class of extreme points, it would be interesting to know whether cut cones are also exposed, namely if there exist linear continuous functionals attaining their minimum, when restricted to the closed unit ball of the Hessian–Schatten seminorm, only at a cut cone.

iii) In the two-dimensional case, one may wonder whether the functional (0.1) admits minimizers, with the choice of error and regularizing term described above. In Section 4 we give a positive answer, for a large set of choices of the parameters λ, p and q.

Now we pass to a more detailed description of the content of the paper. Namely, we examine separately the answers to items i), ii) and iii) above and we sketch their proofs.

Density of CPWL functions

In Section 2 we address the problem of density in energy $|D_1^2 \cdot |(\Omega)$ of CPWL functions in the set of functions with bounded Hessian–Schatten variation. Our main result is Theorem 2.2, stated for C^2 targets, and then it follows the localized version Theorem 2.4 for targets with finite p-Hessian–Schatten variation. The proof of Theorem 2.2 heavily relies on a fine study of triangulations of $\mathbb{R}^d$ and consists morally of three parts.

Part 1 is Section 2.1 and deals with general properties of triangulations (considered as couples of sets, the set of vertices and the set of elements), the most important ones being the Delaunay, non degeneracy and uniformity properties (items (a), (b) and (c) of Definition 2.7). Roughly speaking, the Delaunay property states that given an element of the triangulation, no vertex of the triangulation lies inside the circumsphere of the given element; it entails regularity properties, among them, the fact that angles in the elements are not too small. This leads to the non degeneracy property, crucial to estimate geometric quantities related to an element in terms of the volume of the given element. Finally, uniformity states that the vertices of the triangulation are uniformly distributed, in the sense that the maximum size of a ball which contains no vertex is bounded by a constant times the minimal distance between two vertices. The main results are Lemma 2.9, that allows us to gain a Delaunay triangulation starting from a uniform set of vertices and Lemma 2.13 which studies Delaunay triangulation whose vertices locally coincide with a rotation of a rescaling of the lattice $\mathbb{Z}^d$.

Part 2 is Section 2.2 which aims to construct a "good" triangulation (in the sense of **Part 1**) that locally follows a prescribed orientation. The outcome is Theorem 2.14 and the main difficulty in its proof relies in "gluing" the various sub-triangulations to allow for the variable orientation (see Fig. 3).

Part 3 is the proof of the density result, Section 2.3. We exploit the outcome of **Part 2** to build a triangulation that locally follows the orientation given by the Hessian of w, $\nabla^2 w$, in the sense that is given by an orthonormal basis of eigenvectors for $\nabla^2 w$. Then we take u, the affine interpolation for w with respect to this triangulation, which will be a good approximation. The contribution of the Hessian–Schatten variation of u on regions in which the orientation of the triangulation is constant (and hence adapted to the Hessian of w) is estimated thanks to the good choice of the orientation, whereas the contribution around the boundaries of these regions, i.e. where the gluing took place, comes from the regularity properties of the triangulation and the smallness of these regions.

Extremality of cones

In Section 3, we prove that functions whose graphs are cut cones are extremal with respect to the Hessian–Schatten total variation seminorm. Namely, we prove that functions defined as

$$f^{\mathrm{cone}}(x) := (1 - |x|)_+$$

are extremal modulo affine functions, in the sense that, if, for some $\lambda \in (0, 1)$

$$f^{\mathrm{cone}} = \lambda f_1 + (1 - \lambda) f_2$$

with

$$|\mathrm{D}_p^2 f_1|(\mathbb{R}^d) = |\mathrm{D}_p^2 f_2|(\mathbb{R}^d) = |\mathrm{D}_p^2 f^{\mathrm{cone}}|(\mathbb{R}^d),$$

for some $p \in [1, \infty)$, then f_1 and f_2 are equal to f^{cone}, up to affine functions (Theorem 3.1).

Our strategy is as follows. First, we set f_i^{rad} to be the radial symmetrization of f_i, for $i = 1, 2$. As f^{cone} is radial, a simple computation yields that still

$$f^{\mathrm{cone}} = \lambda f_1^{\mathrm{rad}} + (1 - \lambda) f_2^{\mathrm{rad}}$$

and

$$|\mathrm{D}_p^2 f_1^{\mathrm{rad}}|(\mathbb{R}^d) = |\mathrm{D}_p^2 f_2^{\mathrm{rad}}|(\mathbb{R}^d) = |\mathrm{D}_p^2 f^{\mathrm{cone}}|(\mathbb{R}^d).$$

This implies with not much effort that $f_i^{\mathrm{rad}} = f^{\mathrm{cone}}$, up to affine terms, thanks to the explicit computation of Hessian–Schatten total variation of radial functions (Proposition 1.13).

The bulk of the proof is then to prove that whenever we have f such that $f^{\mathrm{rad}} = f^{\mathrm{cone}}$ and $|\mathrm{D}_p^2 f|(\mathbb{R}^d) = |\mathrm{D}_p^2 f^{\mathrm{cone}}|(\mathbb{R}^d)$, then f equals to f^{cone}, up to affine terms. In other words, in the case $f^{\mathrm{rad}} = f^{\mathrm{cone}}$, we have rigidity of the property that $|\mathrm{D}_p^2 f^{\mathrm{rad}}|(\mathbb{R}^d) \leqq |\mathrm{D}_p^2 f|(\mathbb{R}^d)$ stated in Lemma 1.10.

Case $p = 1$ is dealt in Proposition 3.5. For its proof, a key remark is the fact that, if $\boldsymbol{\Delta}$ denotes the distributional Laplacian, then $\int_{B_1} \boldsymbol{\Delta}(f(U \cdot))$ is independent of $U \in SO(\mathbb{R}^d)$. Hence, by $f^{\mathrm{rad}} = f^{\mathrm{cone}}$, we have that

$$\int_{B_1} \boldsymbol{\Delta} f = \int_{B_1} \boldsymbol{\Delta} f^{\mathrm{cone}} = -|\mathrm{D}_1^2 f^{\mathrm{cone}}|(B_1) = -|\mathrm{D}_1^2 f|(B_1),$$

where the second inequality is obtained by explicit computation (or by concavity of f^{cone} in B_1). This then implies that (at the right hand side there is the total variation of the matrix valued measure $\mathrm{D}\nabla f$ with respect to the 1-Schatten norm)

$$\int_{B_1} \mathrm{d}\, tr(\mathrm{D}\nabla f) = -\int_{B_1} \mathrm{d}|\mathrm{D}\nabla f|_1,$$

so that $tr(\mathrm{D}\nabla f) = -|\mathrm{D}\nabla f|_1$ almost everywhere, which implies that the eigenvalues of $\mathrm{D}\nabla f$ are all negative, almost everywhere (Lemma 3.3), by rigidity in the

inequality $|\text{Tr}(A)| \leqq |A|_1$. Then, by Lemma 3.2, it follows that f has a continuous concave representative in B_1. Finally we exploit concavity to obtain the pointwise bound $f \geqq f^{\text{cone}}$ in B_1, which, combined with the integral equality $f^{\text{rad}} = f^{\text{cone}}$, implies the claim.

Case $p \in (1, \infty)$ is dealt in Proposition 3.6, where we reduce ourselves to the case $p = 1$, namely we show that the information $|\text{D}_p^2 f|(\mathbb{R}^d) = |\text{D}_p^2 f^{\text{cone}}|(\mathbb{R}^d)$, coupled with $f^{\text{rad}} = f^{\text{cone}}$, self improves to $|\text{D}_1^2 f|(\mathbb{R}^d) = |\text{D}_1^2 f^{\text{cone}}|(\mathbb{R}^d)$, whence we can use what proved in the **Case** $p = 1$. This reduction is done treating separately the absolutely continuous and singular part of $|\text{D}_p^2 f|$. The former is treated exploiting the strict convexity of the p-Schatten norm together with the scaling property of the map $p \mapsto |\text{D}_p^2 f^{\text{cone}}|$, whereas the latter is treated by Alberti's rank 1 Theorem ([1]), in conjunction with the fact that the p-Schatten norm of rank 1 matrices is independent of p.

Solutions to the minimization problem

In Section 4 we restrict ourselves to the two dimensional Euclidean space. Indeed, we want to exploit the continuity of functions with bounded Hessian–Schatten variation in dimension 2 ([2], see Proposition 1.11) to have a meaningful evaluation functional and define, for $\Omega \subseteq \mathbb{R}^2$ open (cf. (0.1)), $\mathcal{F}_\lambda : L^1_{\text{loc}}(\Omega) \to [0, \infty]$ by

$$\mathcal{F}_\lambda(f) = |\text{D}_1^2 f|(\Omega) + \lambda \|(f(x_i) - y_i)_{i=1,\dots,N}\|_{\ell^1}, \tag{0.4}$$

where $x_1, \dots, x_N \in \Omega$ are distinct points and $y_1, \dots, y_N \in \mathbb{R}$. Also, we are adopting the convention that $\infty \cdot 0 = 0$, hence, if $\lambda = \infty$, we have $\mathcal{F}_\infty : L^1_{\text{loc}}(\Omega) \to [0, \infty]$,

$$\mathcal{F}_\infty(f) = \begin{cases} |\text{D}_1^2 f|(\Omega) & \text{if } f(x_i) = y_i \quad \text{for} \quad i = 1, \dots, N, \\ \infty & \text{otherwise.} \end{cases}$$

Notice that $\mathcal{F}_\lambda$ is the sum of the regularizing term $|\text{D}_1^2 f|$ and the weighted (by λ) error term $\lambda \|(f(x_i) - y_i)_{i=1,\dots,N}\|_{\ell^1}$ and that $\mathcal{F}_\lambda$ can be seen as a relaxed version of $\mathcal{F}_\infty$.

In Section 4, we will consider slightly more general functionals, see (4.1), but for the sake of clarity we reduce ourselves to a particular case in this introduction. Our aim is to prove existence of minimizers of $\mathcal{F}_\lambda$ (Theorem 4.2). Notice that in higher ($\geqq 3$) dimension, $\mathcal{F}_\lambda$ is not well defined (by the lack of continuity), and, even if we try to define it imposing continuity on its domain, minimizers do not exist in general, as the infimum of $\mathcal{F}_\lambda$ is always zero. To see this last claim, simply exploit the scaling property of the Hessian–Schatten total variation (or use Proposition 1.13) for functions of the kind $x \mapsto \sum_{i=1}^{N} y_i (1 - |x - x_i|/r)_+$ as $r \searrow 0$.

We sketch now the proof of the existence of minimizers of $\mathcal{F}_\lambda$. There are two key steps. We denote $\lambda_c := 4\pi$, the "critical" value for λ.

Step 1. First we prove existence of minimizers of $\mathcal{F}_\lambda$, for $\lambda \in [0, \lambda_c]$. This is done via the direct method of calculus of variations, after we prove relative compactness of

minimizing sequences and semicontinuity of this functional. Compactness, proved in Proposition 4.9, is mostly due to the estimates of [2], see Proposition 1.11. Semicontinuity is then proved in Lemma 4.8 and here the choice of $\lambda \in [0, \lambda_c]$ plays a role. The key idea is that, given a point x_i and a converging sequence $f_k \to f$, either $|D_1^2 f_k|$ concentrates at x_i or it does not. In the former case (Lemma 4.7), as a part of $|D_1^2 f_k|$ concentrates at x_i (and $|D_1^2 f|(x_i) = 0$, being points of codimension 2), we experience a drop in the regularizing term of the functional, and this drop is enough to offset the lack of convergence of the evaluation term $f_k(x_i)$ in the error term. In the latter case (Lemma 4.7 again), we have instead convergence of $k \mapsto f_k(x_i)$.

Step 2. We prove the existence of minimizers of $\mathcal{F}^\lambda$, for $\lambda \in [\lambda_c, \infty]$. By **Step 1**, we can take a minimizer f of $\mathcal{F}_{\lambda_c}$. Then we modify f to obtain $\tilde{f}$ satisfying

$$|D_1^2 \tilde{f}|(\Omega) \leqq |D_1^2 f|(\Omega) + \lambda_c \|(f(x_i) - y_i)_i\|_{\ell^1}$$
$$\text{and} \quad \tilde{f}(x_i) = y_i \text{ for } i = 1, \ldots, N.$$

Such modifications is obtained adding to f a suitable linear combination of "cut-cones", namely functions $x \mapsto y_i(1 - |x - x_i|/\bar{r})_+$ for $\bar{r}$ small enough. As $\tilde{f}$ has a perfect fit with the data, for any λ,

$$\mathcal{F}_\lambda(\tilde{f}) = \mathcal{F}_{\lambda_c}(\tilde{f}) \leqq \mathcal{F}_{\lambda_c}(f),$$

where the inequality is due to the construction of $\tilde{f}$. Now, as $\mathcal{F}_\lambda \geqq \mathcal{F}_{\lambda_c}$ (here the choice $\lambda \in [\lambda_c, \infty]$ plays a role) and as f is a minimizer of $\mathcal{F}_{\lambda_c}$, we see that $\tilde{f}$ is a minimizer of $\mathcal{F}_\lambda$.

Therefore, putting together what seen in **Step 1** and in **Step 2** we have that for every $\lambda \in [0, \infty]$ there exists a minimizer of $\mathcal{F}_\lambda$.

1. Preliminaries

In this short section we first recall basic facts about Hessian–Schatten semi-norms and then in Section 1.3 we add an explicit formula to compute Hessian–Schatten variations of radial functions.

1.1. Schatten norms

We recall basic facts about Schatten norms, see [2] and the references therein.

Definition 1.1. (Schatten norm) Let $p \in [1, \infty]$. If $M \in \mathbb{R}^{d \times d}$ and $s_1(M), \ldots, s_d(M) \geqq 0$ denote the singular values of M (counted with their multiplicity), we define the Schatten p-norm of M by

$$|M|_p := \|(s_1(M), \ldots, s_d(M))\|_{\ell^p}.$$

We recall that the scalar product between M, $N \in \mathbb{R}^{d \times d}$ is defined by

$$M \cdot N := \operatorname{tr}(M^t N) = \sum_{i,\,j=1,\ldots,d} M_{i,j} N_{i,j}$$

and induces the Hilbert–Schmidt norm. Next, we enumerate several properties of the Schatten norms that shall be used throughout the paper

Proposition 1.2. *The family of Schatten norms satisfies the following properties.*

i) *If $M \in \mathbb{R}^{d \times d}$ is symmetric, then its singular values $s_1(M), \ldots, s_d(M)$ are equal to $|\lambda_1(M)|, \ldots, |\lambda_d(M)|$, where $\lambda_1(M), \ldots, \lambda_d(M)$ denote the eigenvalues of M (counted with their multiplicity). Hence $|M|_p = \|(\lambda_1(M), \ldots, \lambda_d(M))\|_{\ell^p}$.*

ii) *If $M \in \mathbb{R}^{d \times d}$ and $N \in O(\mathbb{R}^d)$, then $|MN|_p = |NM|_p = |M|_p$.*

iii) *If M, $N \in \mathbb{R}^{d \times d}$, then $|MN|_p \leqq |M|_p |N|_p$.*

iv) *If $M \in \mathbb{R}^{d \times d}$, then $|M|_p = \sup_N M \cdot N$, where the supremum is taken among all $N \in \mathbb{R}^{d \times d}$ with $|N|_{p'} \leqq 1$, for p' the conjugate exponent of p, defined by $1/p + 1/p' = 1$.*

v) *If M has rank 1, then $|M|_p$ coincides with the Hilbert–Schmidt norm of M for every $p \in [1, \infty]$.*

vi) *If $p \in (1, \infty)$, then the Schatten p-norm is strictly convex.*

vii) *If $M \in \mathbb{R}^{d \times d}$, then $|M|_p \leqq C |M|_q$, where $C = C(d, p, q)$ depends only on d, p and q.*

Definition 1.3. (L^r-Schatten norm) Let p, $r \in [1, \infty]$ and let $M \in C_{\mathrm{c}}(\mathbb{R}^d)^{d \times d}$. We define the Schatten (p, r)-norm of M by

$$\|M\|_{p,r} := \||M|_p\|_{L^r(\mathbb{R}^d)}.$$

1.1.1. Poincaré inequalities   We recall basic facts about Poincaré inequalities.

Definition 1.4. Let $A \subseteq \mathbb{R}^d$ be a domain. We say that A supports Poincaré inequalities if for every $q \in [1, d)$ there exists a constant $C = C(A, q)$ depending on A and q such that

$$\left(\fint_A \left| f - \fint_A f \right|^{q^*} \mathrm{d}\mathscr{L}^d \right)^{1/q^*} \leqq C \left(\fint_A |\nabla f|^q \mathrm{d}\mathscr{L}^d \right)^{1/q}$$
$$\text{for every } f \in W^{1,q}(A),$$

where $1/q^* = 1/q - 1/d$.

1.2. Hessian–Schatten total variation

For this section fix $\Omega \subseteq \mathbb{R}^d$ open and $p \in [1, \infty]$. We let p' denote the conjugate exponent of p. Now we recall the definition of Hessian–Schatten total variation and some basic properties, see [2] and the references therein.

Definition 1.5. (Hessian–Schatten variation) Let $f \in L^1_{\text{loc}}(\Omega)$. For every $A \subseteq \Omega$ open we define

$$|D^2_p f|(A) := \sup_F \int_A \sum_{i,\,j=1,\ldots,d} f\, \partial_i \partial_j F_{i,j}\, d\mathscr{L}^d, \tag{1.1}$$

where the supremum runs among all $F \in C^\infty_c(A)^{d \times d}$ with $\|F\|_{p',\infty} \leqq 1$. We say that f has bounded p-Hessian–Schatten variation in Ω if $|D^2_p f|(\Omega) < \infty$.

Remark 1.6. If f has bounded p-Hessian–Schatten variation in Ω, then the set function defined in (1.1) is the restriction to open sets of a finite Borel measure, that we still call $|D^2_p f|$. This can be proved with a classical argument, building upon [12] (see also [3, Theorem 1.53]).

By its very definition, the p-Hessian–Schatten variation is lower semicontinuous with respect to convergence in distributions. ∎

For any couple $p,\, q \in [1, \infty]$, f has bounded p-Hessian–Schatten variation if and only if f has bounded q-Hessian–Schatten variation and, moreover,

$$C^{-1}|D^2_p f| \leqq |D^2_q f| \leqq C|D^2_p f|$$

for some constant $C = C(d,\, p,\, q)$ depending only on d, p and q. This is due to equivalence of matrix norms.

The next proposition connects Definition 1.5 with Demengel's space of functions with bounded Hessian [11], namely Sobolev functions whose partial derivatives are functions of bounded variation. We shall use D to denote the distributional derivative, to keep the distinction with ∇ notation (used also for gradients of Sobolev functions).

Proposition 1.7. [2, Proposition 9] *Let* $f \in L^1_{\text{loc}}(\Omega)$. *Then the following are equivalent:*

- *f has bounded Hessian–Schatten variation in Ω,*
- $f \in W^{1,1}_{\text{loc}}(\Omega)$ *and* $\nabla f \in \text{BV}_{\text{loc}}(\Omega; \mathbb{R}^d)$ *with* $|D\nabla f|(\Omega) < \infty$.

If this is the case, then, as measures,

$$|D^2_p f| = \left| \frac{dD\nabla f}{d|D\nabla f|} \right|_p |D\nabla f|.$$

In particular, there exists a constant $C = C(d, p)$ depending only on d and p such that

$$C^{-1}|D\nabla f| \leqq |D^2_p f| \leqq C|D\nabla f|$$

as measures.

Proposition 1.8. [2, Proposition 11] *Let $f \in L^1_{\mathrm{loc}}(\Omega)$. Then, for every $A \subseteq \Omega$ open, it holds*

$$|\mathrm{D}^2_p f|(A) = \inf \left\{ \liminf_k \int_A |\nabla^2 f_k|_p \mathrm{d}\mathscr{L}^d \right\}$$

where the infimum is taken among all sequences $(f_k) \subseteq C^\infty(A)$ such that $f_k \to f$ in $L^1_{\mathrm{loc}}(A)$. If moreover $f \in L^1(A)$, the convergence in $L^1_{\mathrm{loc}}(A)$ above can be replaced by convergence in $L^1(A)$.

In the statement of the next lemma and in the sequel we denote by $B_\varepsilon(A)$ the open ε-neighbourhood of $A \subseteq \mathbb{R}^d$.

Lemma 1.9. [2, Lemma 12] *Let $f \in L^1_{\mathrm{loc}}(\Omega)$ with bounded Hessian–Schatten variation in Ω. Let also $A \subseteq \mathbb{R}^d$ open and $\varepsilon > 0$ with $B_\varepsilon(A) \subseteq \Omega$. Then, if $\rho \in C_c(\mathbb{R}^d)$ is a convolution kernel with $\mathrm{supp}\,\rho \subseteq B_\varepsilon(0)$, it holds*

$$|\mathrm{D}^2_p(\rho * f)|(A) \leqq |\mathrm{D}^2_p f|(B_\varepsilon(A)).$$

In the same spirit of Lemma 1.9, we have the following lemma.

Lemma 1.10. *Let $f \in L^1_{\mathrm{loc}}(\Omega)$ with bounded Hessian–Schatten variation in Ω. Assume that $A \subseteq \Omega$ is open and invariant under the action of $SO(\mathbb{R}^d)$. For any $U \in SO(\mathbb{R}^d)$ the function $f_U := f(U \cdot)$ satisfies $|\mathrm{D}^2_p f_U|(A) = |\mathrm{D}^2_p f|(A)$. In particular, setting*

$$f^{\mathrm{rad}} := \fint_{SO(\mathbb{R}^d)} f_U \mathrm{d}\mu_d(U),$$

where μ_d is the Haar measure on $SO(\mathbb{R}^d)$, by convexity one has

$$|\mathrm{D}^2_p f^{\mathrm{rad}}|(A) \leqq |\mathrm{D}^2_p f|(A).$$

Proof. The proof is very similar to the one of Lemma 1.9 above i.e. [2, Lemma 12], but we sketch it anyway for the reader's convenience and for future reference.

We take any $F \in C^\infty_c(A)^{n \times n}$ with $\|F\|_{p',\infty} \leqq 1$ and we set $G := U F(U^t \cdot)U^t$. A straightforward computation shows that

$$\sum_{i,j} \partial_i \partial_j G_{i,j}(x) = \sum_{i,j} (\partial_i \partial_j F_{i,j})(U^t x)$$

and that $G \in C^\infty_c(A)^{n \times n}$ with $\|G\|_{p',\infty} \leqq 1$. Then we compute, by a change of variables,

$$\int_A \sum_{i,j} f_U \partial_i \partial_j F_{i,j} \mathrm{d}\mathscr{L}^d = \int_A f(x) \sum_{i,j} (\partial_i \partial_j F_{i,j})(U^t x) \mathrm{d}\mathscr{L}^d(x)$$

$$= \int_A f(x) \sum_{i,j} (\partial_i \partial_j G_{i,j})(x) \mathrm{d}\mathscr{L}^d(x).$$

In particular,

$$\left| \int_A \sum_{i,j} f_U \, \partial_i \partial_j F_{i,j} \mathrm{d}\mathscr{L}^d(x) \right| \leqq |\mathrm{D}_p^2 f|(A).$$

Now, by Fubini's Theorem,

$$\int_A \sum_{i,j} f^{\mathrm{rad}} \partial_i \partial_j F_{i,j} \mathrm{d}\mathscr{L}^d = \int_{SO(\mathbb{R}^d)} \int_A f_U \sum_{i,j} \partial_i \partial_j F_{i,j} \mathrm{d}\mathscr{L}^d \mathrm{d}\mu_d(U)$$

$$\leqq \int_{SO(\mathbb{R}^d)} |\mathrm{D}_p^2 f|(A) \mathrm{d}\mu_d(U) = |\mathrm{D}_p^2 f|(A),$$

whence the claim as F was arbitrary.   □

Proposition 1.11. Sobolev embedding, [2, Proposition 13] *Let $f \in L^1_{\mathrm{loc}}(\Omega)$ with bounded Hessian–Schatten variation in Ω. Then*

$$\begin{aligned}
&f \in L^{d/(d-2)}_{\mathrm{loc}}(\Omega) \cap W^{1,d/(d-1)}_{\mathrm{loc}}(\Omega) && \textit{if } d \geqq 3, \\
&f \in L^\infty_{\mathrm{loc}}(\Omega) \cap W^{1,2}_{\mathrm{loc}}(\Omega) && \textit{if } d = 2, \\
&f \in L^\infty_{\mathrm{loc}}(\Omega) \cap W^{1,\infty}_{\mathrm{loc}}(\Omega) && \textit{if } d = 1
\end{aligned}$$

and, if $d = 2$, f has a continuous representative.

More explicitly, for every $A \subseteq \Omega$ bounded domain that supports Poincaré inequalities and $r \in [1, \infty)$, there exist $C = C(A, r)$ and an affine map $g = g(A, f)$ such that, setting $\tilde{f} := f - g$, it holds that

$$\begin{aligned}
&\|\tilde{f}\|_{L^{d/(d-2)}(A)} + \|\nabla \tilde{f}\|_{L^{d/(d-1)}(A)} \leqq C|\mathrm{D}^2 f|(A) && \textit{if } d \geqq 3, \\
&\|\tilde{f}\|_{L^r(A)} + \|\nabla \tilde{f}\|_{L^2(A)} \leqq C|\mathrm{D}^2 f|(A) && \textit{if } d = 2, \\
&\|\tilde{f}\|_{L^\infty(A)} + \|\nabla \tilde{f}\|_{L^\infty(A)} \leqq C|\mathrm{D}^2 f|(A) && \textit{if } d = 1.
\end{aligned}$$

Lemma 1.12. Rigidity, [2, Lemma 15] *Let $f, g \in L^1_{\mathrm{loc}}(\Omega)$ with bounded Hessian–Schatten variation in Ω and assume that*

$$|\mathrm{D}_p^2(f + g)|(\Omega) = |\mathrm{D}_p^2 f|(\Omega) + |\mathrm{D}_p^2 g|(\Omega).$$

Then

$$|\mathrm{D}_p^2(f + g)| = |\mathrm{D}_p^2 f| + |\mathrm{D}_p^2 g|$$

as measures on Ω.

1.3. Hessian–Schatten variation of radial functions

The following result is new and aims at computing the Hessian–Schatten variation of radial functions. This will be needed in Sections 3 and 4. Notice also that, as expected, the contribution involving the singular part of $|Dg'|$ in (1.2) below does not depend on p.

In the proof we shall use the auxiliary function $F : (0, R) \times \mathbb{R}^2 \to [0, \infty)$

$$F(s, (v_1, v_2)) := d\omega_d \|(sv_2, v_1, \ldots, v_1)\|_{\ell^p} s^{d-2},$$

where v_1 is repeated $d-1$ times and $\omega_d := \mathscr{L}^d(B_1)$ (d will be the dimension of the Euclidean ambient space). Notice that F is continuous, convex and 1-homogeneous with respect to the (v_1, v_2) variable. Therefore, for intervals $(r_1, r_2) \subseteq (0, R)$, the functional

$$\Phi_{(r_1, r_2)}(\mu) := \int_{(r_1, r_2)} F\left(s, \frac{d\mu}{d|\mu|}\right) d|\mu| = \int_{(r_1, r_2)} F\left(s, \frac{d\mu}{d\lambda}\right) d\lambda \qquad \text{whenever } |\mu| \ll \lambda,$$

defined on $\mathbb{R}^2$-valued measures μ makes sense and is convex. Furthermore, Reshetnyak lower semicontinuity Theorem (e.g. [3, Theorem 2.38]) grants its lower semicontinuity with respect to weak convergence in duality with $C_c((r_1, r_2))$.

Proposition 1.13. *Let $d \geq 2$ and let $g \in L^1_{loc}((0, R)) \to \mathbb{R}$ be such that $\int_0^r s^{d-1} |g(s)| ds < \infty$ for every $r \in (0, R)$. Define $f(\cdot) := g(|\cdot|) \in L^1_{loc}(B_R(0))$.*

Assume that f has bounded Hessian–Schatten total variation in $B_R(0)$. Then $g \in W^{1,1}_{loc}((0, R))$ and $g' \in BV_{loc}((0, R))$. Write the decomposition $Dg' = D^s g' + g'' \mathscr{L}^1$, where $D^s g' \perp \mathscr{L}^1$. Then, for every $r \in (0, R]$ and $p \in [1, \infty]$, one has

$$|D^2_p f|(B_r(0)) = d\omega_d \left(\int_{(0, r)} s^{d-1} d|D^s g'|(s) + \int_0^r \|(sg''(s), g'(s), \ldots, g'(s))\|_{\ell^p} s^{d-2} ds \right). \tag{1.2}$$

Conversely, assume that $g \in W^{1,1}_{loc}((0, R))$ and $g' \in BV_{loc}((0, R))$, and, with the same notation above, that

$$\int_{(0, R)} s^{d-1} d|D^s g'|(s) + \int_0^R \|(sg''(s), g'(s), \ldots, g'(s))\|_{\ell^p} s^{d-2} ds < \infty.$$

Then f has bounded Hessian–Schatten total variation in $B_R(0)$ and the Hessian–Schatten variation of f is computed as above.

Proof. Let $r \in (0, R)$. Let ρ_k be radial Friedrich mollifiers for $\mathbb{R}^d$ and define $f_k := \rho_k * f$. As f_k is still radial, we write $f_k(\cdot) = g_k(|\cdot|)$, where $g_k \in C^\infty((0, r))$. As $f_k \to f \in L^1(B_r(0))$, $g_k \to g$ in $L^1_{loc}((0, r))$. Now we compute, on $B_r(0)$,

$$\nabla^2 f_k(x) = g_k''(|x|) \frac{x \otimes x}{|x|^2} + g_k'(|x|) \frac{|x|^2 \mathrm{Id} - x \otimes x}{|x|^3}.$$

Notice that the eigenvalues of the matrix appearing at the right hand side of the equation above are $g_k''(|x|)$ with multiplicity 1 and $g_k'(|x|)/|x|$ with multiplicity

$d - 1$, the eigenvectors being x and a basis of $x^\perp$. Therefore, by Proposition 1.7, on $B_r(0)$ one has

$$|D_p^2 f_k| = |x|^{-1} \big\| \big(|x| g_k''(|x|), g_k'(|x|), \ldots, g_k'(|x|)\big) \big\|_{\ell^p} \mathscr{L}^d \geqq g_k''(|x|) \mathscr{L}^d. \quad (1.3)$$

As $|D_p^2 f_k|(B_r(0))$ is uniformly bounded by Lemma 1.9, we obtain the claimed membership for g, letting eventually $r \nearrow R$.

For the purpose of proving the inequality $\geqq$ in (1.2). It is enough to compute $|D_p^2 f|(A_{r_1, r_2})$, where we define the open annulus

$$A_{r_1, r_2} := B_{r_2}(0) \setminus \bar{B}_{r_1}(0)$$

for $[r_1, r_2] \subseteq (0, R)$. Also, there is no loss of generality in assuming that r_1 and r_2 are such that $|Dg'|(\{r_1\}) = |Dg'|(\{r_2\}) = 0$, as well as $|D\nabla f|(\partial A_{r_1, r_2}) = 0$, hence we will tacitly assume this condition in what follows.

From (1.3), with the notation $\mu_g := (g' \mathscr{L}^1, Dg')$, we get

$$|D_p^2 f_k|(A_{r_1, r_2}) = \int_{A_{r_1, r_2}} |D_p^2 f_k|(x) \mathrm{d}\mathscr{L}^d(x) = \Phi_{(r_1, r_2)}(\mu_{g_k}).$$

Now notice that Lemma 1.9 and our choice of radii grant $|D_p^2 f|(A_{r_1, r_2}) = \lim_k |D_p^2 f_k|(A_{r_1, r_2})$, so that the lower semicontinuity of Φ together with the weak* convergence of μ_{g_k} to μ_g grants

$$|D_p^2 f|(A_{r_1, r_2}) \geqq \Phi_{(r_1, r_2)}(\mu_g)$$
$$= \mathrm{d}\omega_d \left(\int_{(r_1, r_2)} s^{d-1} \mathrm{d}|D^s g'|(s) + \int_{r_1}^{r_2} \|(sg''(s), g'(s), \ldots, g'(s))\|_{\ell^p} s^{d-2} \mathrm{d}s \right).$$

Letting $r_1 \to 0$ and $r_2 \to r$ provides the inequality $\geqq$ in (1.2).

Now we prove the converse implication and inequality. This time we denote by (ρ_k) a sequence of Friedrich mollifiers on $\mathbb{R}$ and we call $g_k := \rho_k * g$, then $f_k(\cdot) := g_k(|\cdot|)$. Notice that, with our choice of the radii, $|\mu_{g_k}|((r_1, r_2))$ converges to $|\mu_g|((r_1, r_2))$ as $k \to \infty$, therefore invoking Reshetnyak continuity Theorem (e.g. [3, Theorem 2.39]) we get

$$|D_p^2 f|(A_{r_1, r_2}) \leqq \liminf_k |D_p^2 f_k|(A_{r_1, r_2}) = \liminf_k \Phi_{(r_1, r_2)}(\mu_{g_k})$$

$$= \Phi_{(r_1, r_2)}(\mu_g) \leqq \Phi_{(0, R)}(\mu_g)$$

$$= \mathrm{d}\omega_d \left(\int_{(0, R)} s^{d-1} \mathrm{d}|D^s g'|(s) + \int_0^R \|(sg''(s), g'(s), \ldots, g'(s))\|_{\ell^p} s^{d-2} \mathrm{d}s \right).$$

Letting $r_1 \to 0$ and $r_2 \to R$ gives that f has bounded Hessian–Schatten total variation in $B_R(0) \setminus \{0\}$. To conclude, obtaining also the converse inequality in (1.2), we need just to apply the classical Lemma 1.14 below to f and to the partial derivatives of f, taking into account the mutual absolute continuity of $|D_p^2 f|$ and $|D\nabla f|$ (Proposition 1.7). $\quad \square$

Lemma 1.14. *Let $B_R(0) \subseteq \mathbb{R}^d$, $d \geqq 2$ and let $h \in W^{1,1}(B_R(0) \setminus \{0\})$ (resp. $h \in \mathrm{BV}(B_R(0) \setminus \{0\})$). Then $h \in W^{1,1}(B_R(0))$ (resp. $h \in \mathrm{BV}(B_R(0))$ and $|Dh|(\{0\}) = 0$).*

Proof. By a truncation argument, we can assume with no loss of generality that h is bounded. Then, the approximation of h by the functions $h_k = h(1 - \psi_k) \in W^{1,1}(B_R(0))$ (resp. $\mathrm{BV}(B_R(0))$), where $\psi_k \in C^1_c(B_{1/k}(0))$ satisfy $|\nabla \psi_k| \leqq 2k$, $0 \leqq \psi_k \leqq 1$ and $\psi_k = 1$ in a neighbourhood of 0, together with Leibniz rule, provides the result. □

2. Density of CPWL Functions

We recall the definition of continuous piecewise linear (CPWL) functions. In view of this definition we state that a simplex in $\mathbb{R}^d$ is the convex hull of $d + 1$ points (called vertices of the simplex) that do not lie on an hyperplane, and a face of a simplex is the convex hull of a subset of its vertices.

Definition 2.1. Let $\Omega \subseteq \mathbb{R}^d$ open and let $f \in C(\Omega)$. We say that f is CPWL (or $f \in \mathrm{CPWL}(\Omega)$) if there exists a decomposition of $\mathbb{R}^d$ in d-dimensional simplexes $\{P_k\}_{k \in \mathbb{N}}$, such that

 (i) $P_k \cap P_h$ is either empty or a common face of P_k and P_h, for every $h \neq k$;
(ii) for every k, the restriction of f to $P_k \cap \Omega$ is affine;
(iii) the decomposition is locally finite, in the sense that for every ball B, only finitely many P_k intersect B.

The main theorem of this section is the following density result:

Theorem 2.2. *For any $w \in C^2(\mathbb{R}^d)$ there exists a sequence $(u_j) \subseteq \mathrm{CPWL}(\mathbb{R}^d)$ with $u_j \to w$ in the $L^\infty_{\mathrm{loc}}(\mathbb{R}^d)$ topology and such that for any bounded open set $\Omega \subseteq \mathbb{R}^d$ with $\mathscr{L}^d(\partial\Omega) = 0$,*

$$\lim_{j \to \infty} |\mathrm{D}_1^2 u_j|(\Omega) = |\mathrm{D}_1^2 w|(\Omega).$$

Recall that, as explained in [2, Remark 22], because of lower semicontinuity the exponent $p = 1$ is the only meaningful exponent in a density result as above, namely this sharp approximation by CPWL functions is not possible for the energy $|\mathrm{D}_p^2 f|$ when $p > 1$.

We defer the proof of Theorem 2.2 to Section 2.3, after having studied properties of "good" triangulations in Sections 2.1 and 2.2. Namely, we aim to construct triangulations of $\mathbb{R}^d$ which locally follow a prescribed orientation. The general scheme is illustrated in Fig. 2. In each of the large squares it coincides with a rotation of a triangulation of $\varepsilon\mathbb{Z}^d$; the difficulty resides in the interpolation region between different squares. In Section 2.1 we discuss standard material on general properties of triangulations. In Section 2.2 we present the specific construction, the key result is Theorem 2.14. This is then used to prove density in Theorem 2.2.

First, we start with a brief discussion around the result of Theorem 2.2. We recall the following extension result [2, Lemma 17] its last claim is immediate, once one takes into account also Proposition 1.11:

Lemma 2.3. *Let $\Omega := (0,1)^d \subseteq \mathbb{R}^d$ and let $f \in L^1_{\text{loc}}(\Omega)$ with bounded Hessian–Schatten variation in Ω. Then there exist an open neighbourhood $\tilde{\Omega}$ of $\bar{\Omega}$ and $\tilde{f} \in L^1_{\text{loc}}(\tilde{\Omega})$ with bounded Hessian–Schatten variation in $\tilde{\Omega}$ such that*

$$|D^2_1 \tilde{f}|(\partial\Omega) = 0 \tag{2.1}$$

and

$$\tilde{f} = f \qquad a.e. \text{ on } \Omega.$$

In particular, $f \in L^1(\Omega)$.

The next result gives a positive answer to [2, Conjecture 1], partially proved in the two-dimensional case in [2, Theorem 21]. The proof is based on Theorem 2.2 and a diagonal argument.

Theorem 2.4. *Let $\Omega := (0,1)^d \subseteq \mathbb{R}^d$. Then* CPWL *functions are dense with respect to the energy $|D^2_1 \cdot |(\Omega)$ in the space*

$$\{f \in L^1_{\text{loc}}(\Omega) : f \text{ has bounded Hessian–Schatten variation in } \Omega\}$$

with respect to the $L^1(\Omega)$ topology. Namely, for any $f \in L^1_{\text{loc}}(\Omega)$ with bounded Hessian–Schatten variation in Ω, there exists $\{f_k\}_k \subseteq \text{CPWL}(\Omega)$ with $f_k \to f$ in $L^1(\Omega)$ and $|D^2_1 f_k|(\Omega) \to |D^2_1 f|(\Omega)$.

Proof. Take f as in the statement, and let $\tilde{f}$ be given by Lemma 2.3. By using smooth cut-off functions, there is no loss of generality in assuming that $\tilde{f}$ is compactly supported in $\tilde{\Omega}$, hence, in particular, $\tilde{f} \in L^1(\mathbb{R}^d)$. Also, we see that we can assume that $\mathscr{L}^d(\partial\tilde{\Omega}) = 0$.

Now we take $(\tilde{f}_k) \subseteq C^\infty_c(\mathbb{R}^d)$ be mollifications of $\tilde{f}$ by means of compactly supported mollifiers, notice that $\tilde{f}_k \to \tilde{f}$ in $L^1(\mathbb{R}^d)$ and $|D^2_1 f_k|(\tilde{\Omega}) = |D^2_1 f_k|(\mathbb{R}^d) \to |D^2_1 \tilde{f}|(\mathbb{R}^d) = |D^2_1 \tilde{f}|(\tilde{\Omega})$, thanks to Proposition 1.9 and lower semicontinuity. Now, for any k, take $(\tilde{f}_{k,h}) \subseteq \text{CPWL}(\mathbb{R}^d)$ be given by Theorem 2.2 for $\tilde{f}_k$. With a diagonal argument, we obtain $(g_\ell) \subseteq \text{CPWL}(\mathbb{R}^d)$ with $g_\ell \to \tilde{f}$ in $L^1(\tilde{\Omega})$ and such that $|D^2_1 g_\ell|(\tilde{\Omega}) \to |D^2_1 \tilde{f}|(\tilde{\Omega})$. By lower semicontinuity, the fact that $|D^2_1 g_\ell|(\tilde{\Omega}) \to |D^2_1 \tilde{f}|(\tilde{\Omega})$ and (2.1), it easily follows that

$$|D^2_1 g_\ell|(\Omega) \to |D^2_1 \tilde{f}|(\Omega) = |D^2_1 f|(\Omega).$$

Clearly, $g_\ell \to f$ in $L^1(\Omega)$, so that the proof is concluded. □

Remark 2.5. Let $\Omega := (0,1)^d$. As a consequence of Theorem 2.4, the description of the extremal points of the unit ball with respect to the $|D^2_1 \cdot |(\Omega)$ seminorm obtained in [2, Theorem 25] remains in place in arbitrary dimension. In a slightly imprecise way, the result states that CPWL extremal points are dense in 1-Hessian–Schatten energy in the set of extremal points with respect to the $L^1(\Omega)$ topology. Notice that the description of CPWL extremal points is made explicit in [2, Proposition 23]. ∎
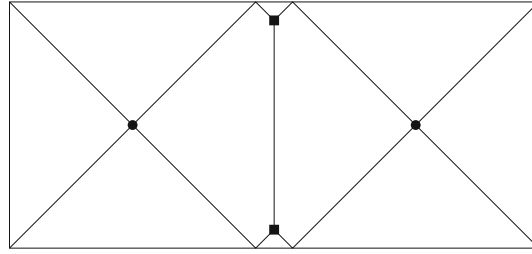
Fig. 1. Sketch of the function G_h used in proving Remark 2.6. The function equals 1 on the two points marked by black dots, $-h$ on the two points marked by black squares, vanishes outside the large rectangle, and is affine in each of the ten polygons in the figure

Remark 2.6. The set of extremal points is not closed with respect to the convergence considered here. For example, with $d = 2$, one can easily check that the function $g(x) := \max\{1 - \|x\|_{\ell^\infty}, 0\}$ is extremal, but the function $G_0(x) := g(x + e_1) + g(x - e_1)$ is not. Indeed, $G_0 = \frac{1}{2}(2\, g(\cdot + e_1) + 2\, g(\cdot - e_1))$, with $|D_p^2 G_0|(\mathbb{R}^2) = |D_p^2 2\, g(\cdot + e_1)|(\mathbb{R}^2) = |D_p^2 2\, g(\cdot - e_1)|(\mathbb{R}^2)$. For $h \in (0, 1/4)$ we then define $G_h : \mathbb{R}^2 \to \mathbb{R}$ by

$$G_h(x) := \max\{1 - \|x - (1 + h)e_1\|_{\ell^\infty}, 1 - \|x + (1 + h)e_1\|_{\ell^\infty}, -\mathrm{dist}_{\ell^\infty}(x, \partial R_h)\}$$

if $x \in R_h := [-2 - h, 2 + h] \times [-1, 1]$, and $G_h(x) = 0$ if $x \in \mathbb{R}^2 \backslash R_h$ (see Fig. 1). Then each G_h is CPWL, is extremal, and $G_h \to G_0$ uniformly with $|D_p^2 G_h|(\mathbb{R}^2) \to |D_p^2 G_0|(\mathbb{R}^2)$ for any $p \in [1, \infty]$, but G_0 is not extremal.

Let us briefly comment on the proof of extremality of G_h (the same argument implies extremality of g). If $G_h = \lambda f + (1 - \lambda) f'$, with $\lambda \in (0, 1)$ and $|D_p^2 f|(\mathbb{R}^2) = |D_p^2 f'|(\mathbb{R}^2) = |D_p^2 G_h|(\mathbb{R}^2)$, then by Lemma 1.12 the support of $|D_p^2 f|$ is contained in the support of $|D_p^2 G_h|$, so that f (after choosing the continuous representative) is affine in each of the sets on which G_h is affine. Adding an irrelevant affine function, we can reduce to the case that $f = 0$ outside R_h. Using the fact that if two affine functions coincide on three non-collinear points then they coincide everywhere, one obtains $f = aG_h$, where $a := f((1 + h)e_1) \in \mathbb{R}$ (see Fig. 1); by equality of the norms $a = \pm 1$. Similarly, $f' = \pm G_h$, so that by $G_h = \lambda f + (1 - \lambda) f'$ we obtain $G_h = f = f'$. ∎

2.1. General properties of triangulations

We define a triangulation of $\mathbb{R}^d$ as a pair of two sets, the first one, V, containing the vertices (nodes), the second one, E, containing the elements, which are non-degenerate compact simplexes with pairwise disjoint interior. Each simplex is the convex hull of its $d + 1$ vertices. One further requires a compatibility condition that ensures that neighbouring elements share a complete face (and not a strict subset of a face). We remark that there is a large literature which studies this in the more general framework of simplicial complexes. For the present application the metric and regularity properties are crucial, we present in this section the few properties which are relevant here in a self-contained way.
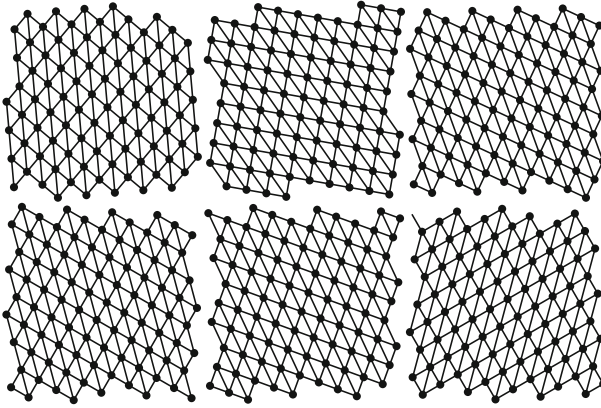
Fig. 2. Sketch of the desired triangulation without the interpolation region. Aim of this section is to find a suitable interpolation between the squares

Definition 2.7. A triangulation of $\mathbb{R}^d$ is a pair (V, E), with $V \subseteq \mathbb{R}^d$ and $E \subseteq \mathcal{P}(\mathbb{R}^d)$ such that

  i) for every $e \in E$, e has non empty interior and there is $v_e \subseteq \mathbb{R}^d$ with $\#v_e = d+1$ and $e = \operatorname{conv}(v_e)$;

  ii) $V = \bigcup_{e \in E} v_e$;

  iii) for any $e, e' \in E$ one has $e \cap e' = \operatorname{conv}(v_e \cap v_{e'})$;

  iv) $\bigcup_{e \in E} e = \mathbb{R}^d$.

We introduce four regularity properties:

  (a) The triangulation has the Delaunay property if for each $e \in E$, the unique open ball B with $v_e \subseteq \partial B$ obeys $B \cap V = \emptyset$.

  (b) The triangulation is c_*-non degenerate, for some $c_* > 0$, if $(\operatorname{diam} e)^d \leqq c_* \mathcal{L}^d(e)$ for all $e \in E$.

  (c) The set $V \subseteq \mathbb{R}^d$ is $(\bar{c}, \varepsilon)$-uniform, for some $\bar{c}, \varepsilon > 0$, if $|x - y| \geqq \varepsilon/\bar{c}$ for all $x \in V$, $y \in V$ with $x \neq y$ and $B_{\bar{c}\varepsilon}(q) \cap V \neq \emptyset$ for all $q \in \mathbb{R}^d$.

  (d) The triangulation is locally finite if, for every ball B, only finitely many elements of E intersect B.

Condition iii) states that two distinct elements of E are either disjoint or share a face of dimension between 0 and $d - 1$; in particular distinct elements have disjoint interior. Notice that $\operatorname{conv}(\emptyset) = \emptyset$.

The Delaunay property (a) states that the circumscribed sphere to each simplex does not contain any other vertex, and implies $\partial e \cap V = v_e$ for all $e \in E$. It can be interpreted as a statement that the vertices have been matched to form simplexes in an "optimal" way.

The non-degeneracy property (b) states that simplexes are uniformly non-degenerate, so that the affine bijection that maps e onto the standard simplex has a uniformly bounded condition number. It implies that there is $C = C(c_*, d)$ such

that for any $e \in E$, any $x \in v_e$, any $F \in \mathbb{R}^d$ one has

$$|F| \leqq C(c_*, d) \sum_{y \in v_e \setminus \{x\}} \frac{|F \cdot (y - x)|}{|y - x|}. \tag{2.2}$$

The uniformity property (c) of a set V of vertices ensures (for Delaunay triangulations) that all sides of all elements have length comparable to ε. Also, property (c) immediately implies property (d), as it forces V to be a locally finite set.

Remark 2.8. Let (V, E) be a triangulation that has the Delaunay property (property (a)) and is $(\bar{c}, \varepsilon)$-uniform (property (c)). Then $\mathrm{diam}(e) \leqq 2\bar{c}\varepsilon$, for any $e \in E$.
∎

Proof. Take $e \in E$ and let $q \in \mathbb{R}^d$ and $r \in (0, \infty)$ such that $v_e \subseteq \partial B_r(q)$. By the Delaunay property, $V \cap B_r(q) = \emptyset$, so that, by $(\bar{c}, \varepsilon)$-uniformity, $\bar{c}\varepsilon > r \geqq \mathrm{diam}(e)/2$. □

We next show how given the set of vertices V one can abstractly obtain a good triangulation. The construction is standard up to a perturbation argument. As we could not find a reference with the complete result, we prove it.

Lemma 2.9. *Let $V \subseteq \mathbb{R}^d$ be uniform in the sense of property (c) of Definition 2.7. Then there is $E \subseteq \mathcal{P}(\mathbb{R}^d)$ such that (V, E) is a triangulation of $\mathbb{R}^d$ with the Delaunay property (a)*

Proof. We define $f : \mathbb{R}^d \to [0, \infty]$ by

$$f(x) := \begin{cases} |x|^2 & \text{if } x \in V, \\ \infty & \text{otherwise.} \end{cases}$$

Let g be the convex envelope of f, which is CPWL (see Lemma 2.10 below). Moreover, notice that

$$g(x) = |x|^2 = f(x) \quad \text{for every } x \in V.$$

Let $q \in \mathbb{R}^d$, $\mu \in \mathbb{R}$ be such that

$$A := \{x : g(x) = \mu + 2x \cdot q\} \tag{2.3}$$

has nonempty interior. Notice that A is compact, convex and coincides with the closure of its interior, and $g(x) > \mu + 2x \cdot q$ for every $x \in \mathbb{R}^d \setminus A$. Also, we set

$$w := \{x \in V : \mu + 2x \cdot q = |x|^2\} = A \cap V, \tag{2.4}$$

then,

$$\mu + 2x \cdot q < |x|^2 \quad \text{for all } x \in V \setminus w.$$

Now we show that $\mathrm{ext}(A) \subseteq V$ so that $\mathrm{ext}(A) \subseteq w$ and hence $A = \mathrm{conv}(w)$ with $\#w \geqq d + 1$ (as A has nonempty interior). Take indeed $p \in \mathrm{ext}(A)$ and assume $p \notin V$. Then, take a minimal set of points $\{p_1, \ldots, p_k\} \subseteq V$ such

that $(p, g(p)) \in \text{conv}\left((p_1, f(p_1)), \ldots, (p_k, f(p_k))\right)$ (this is possible by (2.7) of Lemma 2.10 below). As $p \in \text{ext}\,(A)$, up to reordering, we can assume that $p_1 \notin A$, hence by $g(p_1) > \mu + 2p_1 \cdot q$ we have that $g(p) > \mu + 2p \cdot q$, a contradiction.

The above equations can be rewritten as

$$|x - q|^2 = \mu + |q|^2 \qquad \text{for all } x \in w$$

and

$$|x - q|^2 > \mu + |q|^2 \qquad \text{for all } x \in V \setminus w.$$

We set $r := \sqrt{\mu + |q|^2}$, so that these conditions are $w \subseteq \partial B_r(q)$ and $V \cap B_r(q) = \emptyset$, so that the set w has the Delaunay property.

Notice then that for every $x \in V$, there is at least one set A as in (2.3) with nonempty interior and with $x \in A \cap V$ (this set was called w): this follows from the fact that g is CPWL.

Any decomposition of those elements A in (2.3) with nonempty interior into non degenerate simplexes with vertices in w leads to a pair (V, E) with all 4 claimed properties of triangulations, except for iii) of Definition 2.7. In the rest of the proof we show by a perturbation argument that a decomposition exists such that property iii), which relates neighbouring pieces in which g is affine, also holds.

We first remark that property iii) is automatically true if g is non degenerate, in the sense that each A is a simplex, which is the same as $\#w = d + 1$ (we are going to add a few details about this in the sequel of the proof). In turn, this is true if for every choice of $X := \{x_1, \ldots, x_{d+2}\} \subseteq V$ the $d + 2$ points $\{(x, g(x))\}_{x \in X} \in \mathbb{R}^{d+1}$ do not lie in a d-dimensional hyperplane, so that (2.4) cannot hold for all $x \in X$.

We fix an enumeration $\varphi : V \to \mathbb{N} \setminus \{0, 1\}$ and recall that V is $(\bar{c}, \varepsilon)$-uniform. For any $\rho \in (0, \varepsilon \wedge 1]$ we consider $f_\rho : \mathbb{R}^d \to [0, \infty]$ defined by

$$f_\rho(x) := \begin{cases} |x|^2 + \rho^{\varphi(x)} & \text{if } x \in V, \\ \infty & \text{otherwise.} \end{cases}$$

For a given set $X := \{x_1, \ldots, x_{d+2}\} \subseteq V$ consider the $d + 2$ equations

$$\mu + 2x_i \cdot q = |x_i|^2 + \rho^{\varphi(x_i)} \qquad \text{for } i = 1, \ldots, d + 2 \tag{2.5}$$

in the $d + 1$ unknowns (μ, q). The affine map $T : \mathbb{R}^{d+1} \to \mathbb{R}^{d+2}$ defined by $T_i(\mu, q) := \mu + 2x_i \cdot q - |x_i|^2$ has an image which is at most $d + 1$ dimensional, hence contained in a set of the form $\{\Xi \in \mathbb{R}^{d+2} : \Xi \cdot v = a\}$ for some $v \in S^{d+1}$, $a \in \mathbb{R}$ (which depend on X). If the system (2.5) has a solution, then

$$\sum_{i=1}^{d+2} v_i \rho^{\varphi(x_i)} = a.$$

As $|v| = 1$ and the exponents are all distinct, this is a nontrivial polynomial equation in ρ, and has at most finitely many solutions. As there are countably many possible choices of the set $X \subseteq V$, for all but countably many values of ρ no such system

has a solution. Therefore we can choose $\rho_j \searrow 0$ such that (2.5) has no solution for any choice of X with $X = \{x_1, \ldots, x_{d+2}\} \subseteq V$.

Fix now an index j and let g_{ρ_j} be the convex envelope of f_{ρ_j}. Notice that if ρ_j is sufficiently small (that we are going to assume from here on), then, as V is discrete and $|x|^2$ is strictly convex,

$$g_{\rho_j}(x) = |x|^2 + \rho_j^{\varphi(x)} = f_{\rho_j}(x) \qquad \text{for every } x \in V.$$

Our choice of ρ_j implies that for every j, for every choice of $X := \{x_1, \ldots, x_{d+2}\} \subseteq V$ the $d + 2$ points $\{(x, g_{\rho_j}(x))\}_{x \in X} \in \mathbb{R}^{d+1}$ do not lie in a d-dimensional hyperplane. Now pick μ, q such that

$$A := \{x : g_{\rho_j}(x) = \mu + 2x \cdot q\}$$

has nonempty interior (the function g_{ρ_j} is CPWL, by Lemma 2.10 below). By non-degeneracy, arguing as above, $A = \text{conv}(w)$, with $\#w = d+1$ and $\text{Int}(A) \cap V = \emptyset$. We define E_j as the family of those sets.

Let us justify why (V, E_j) is a triangulation of $\mathbb{R}^d$. It is enough to show that property iii) holds. Take then $e_1, e_2 \in E_j$ (with vertices w_1, w_2), so that there exist two affine functions L_1, L_2 such that $g_{\rho_j} = L_i$ on e_i and $g_{\rho_j} > L_i$ on $\mathbb{R}^d \setminus e_i$, for $i = 1, 2$. Assume that $\xi \in e_1 \cap e_2$, so that $L_1(\xi) = g_{\rho_j}(\xi) = L_2(\xi)$. Take a minimal set $\{\zeta_1, \ldots, \zeta_k\} \subseteq w_2$ with $\xi \in \text{conv}(\{\zeta_1, \ldots, \zeta_k\})$. As for every $a = 1, \ldots, k$, $L_2(\zeta_a) = g_{\rho_j}(\zeta_a) \geqq L_1(\zeta_a)$, it follows that for every $a = 1, \ldots, k$, $g_{\rho_j}(\zeta_a) = L_1(\zeta_a)$ hence $\{\zeta_1, \ldots, \zeta_k\} \subseteq w_1 \cap w_2$, so that we have verified property iii).

Now take $e \in E_j$ and consider the associated set of vertices w. The conditions

$$\mu + 2x \cdot q = |x|^2 + \rho_j^{\varphi(x)} \geqq |x|^2 \qquad \text{for all } x \in w$$

and

$$\mu + 2x \cdot q \leqq |x|^2 + \rho_j^{\varphi(x)} \leqq |x|^2 + \rho_j^2 \qquad \text{for all } x \in V$$

lead to

$$|x - q|^2 \leqq \mu + |q|^2 \qquad \text{for all } x \in w$$

and

$$\rho_j^2 + |x - q|^2 \geqq \mu + |q|^2 \qquad \text{for all } x \in V.$$

Therefore $w \subseteq \overline{B}_r(q)$, and either $r \leqq \rho_j$ or $V \cap B_{r-\rho_j}(q) = \emptyset$, where $r := \sqrt{\mu + |q|^2}$. By uniformity of the grid, necessarily $r - \rho_j < \bar{c}\varepsilon$, which gives $\text{diam}(e) \leqq 2r < 2\bar{c}\varepsilon + 2\rho_j \leqq 2(\bar{c} + 1)\varepsilon$.

For any $x \in V$, the possible choices of e with $x \in v_e$ are restricted by $\text{diam}(e) < 2(\bar{c} + 1)\varepsilon$, which implies $v_e \subseteq V \cap B_{2(\bar{c}+1)\varepsilon}(x)$. As the grid is uniform, the latter set is finite, with a bound depending only on $\bar{c}$. Therefore for any $x \in V$ we can choose a subsequence of ρ_j such that the set

$$\{e \in E_j : x \in v_e\}$$

is, after finitely many steps, constant. As there are countably many $x \in V$, we can choose a common diagonal subsequence. Along this sequence, for any bounded set K the set $\{e \in E_j : e \subseteq K\}$ is, after finitely many steps, constant. Property iii) holds for E_j, and therefore for those sets. Therefore we obtain a common set E with all desired properties. We remark that indeed the Delaunay property follows from the construction of E and the discussion of the first part of the proof: indeed, if $e \in E$, it is easy to see that there exists an affine function coinciding with g on e. $\quad\square$

We next present the result on the regularity of convex envelopes used above.

Lemma 2.10. *Let $V \subseteq \mathbb{R}^d$ be a uniform set of vertices, in the sense of item (c) of Definition 2.7. Let $f : V \to [0, \infty)$ be superlinear, in the sense that*

$$\lim_{x \in V, \, |x| \to \infty} \frac{f(x)}{|x|} = \infty. \tag{2.6}$$

Let $g : \mathbb{R}^d \to [0, \infty)$ be the convex envelope of f (f is extended by ∞ to $\mathbb{R}^d \setminus V$). Then g is CPWL. Moreover,

$$\{(x, g(x)) : x \in \mathbb{R}^d\} \subseteq \operatorname{conv}\left(\{(x, f(x)) : x \in V\}\right) \tag{2.7}$$

(notice that we are not taking the closure of the convex hull at the right hand side).

Remark 2.11. It is easy to verify the following:

(i) That V is uniform implies that g is real-valued.
(ii) That the assumption of superlinearity is necessary. Indeed, consider $d = 2$, $V = \mathbb{Z}^2$, $f(x) = |x|$. Obviously $g(x) \geq |x|$. For any $x \in \mathbb{Q}^2$ there is $n \in \mathbb{N} \setminus \{0\}$ such that $xn \in \mathbb{Z}^2$, which implies $g(x) \leq (1 - \frac{1}{n})f(0) + \frac{1}{n}f(xn) = |x|$, so that $g(x) = |x|$ on $\mathbb{Q}^2$. As g is a real-valued convex function, it is continuous. We conclude $g(x) = |x|$ on $\mathbb{R}^2$, which is not CPWL.

Proof of Lemma 2.10. For $r \in (0, \infty)$, we write

$$f_r(x) := \begin{cases} f(x) & \text{if } x \in V \cap B_r, \\ \infty & \text{otherwise,} \end{cases}$$

and let $g_r \geqq g$ be the convex envelope of f_r. Since V is uniform, any set $V \cap B_r$ is finite, and therefore g_r is CPWL on $\operatorname{conv}(V \cap B_r)$, and infinity outside. If $r \geqq \bar{c}\varepsilon$, with $\bar{c}, \varepsilon > 0$ the constants from item (c) of Definition 2.7, the set $V \cap B_r$ is nonempty.

We shall show below that for any $r > 0$ there is $R > 0$ such that $g = g_R$ on $B_{r/4}$. This implies that g is CPWL on $B_{r/4}$ for any r, and therefore the assertion. The choice of R (which depends on f and r) is done in (2.9) below.

For $r \geqq \bar{c}\varepsilon$ we define $\alpha_r := \max f(V \cap [-r, r]^d)$. We first prove that if $R/\sqrt{d} > r \geqq 4\bar{c}\varepsilon$ then

$$g_R(x) \leqq \alpha_r \text{ for all } x \in B_{r/2}. \tag{2.8}$$

To see this, let $q_1, \ldots, q_{2^d}$ denote the vertices of the cube $[-1, 1]^d$. By uniformity of V, for each i we can pick $p_i \in V \cap B_{\bar{c}\varepsilon}((r - \bar{c}\varepsilon)q_i)$. One checks that $B_{r/2} \subseteq (r - 2\bar{c}\varepsilon)[-1, 1]^2 \subseteq \operatorname{conv}(\{p_1, \ldots, p_{2^d}\})$. As $p_i \in V \cap [-r, r]^d \subseteq V \cap B_R$, we have $g_R(p_i) \leqq f(p_i) \leqq \alpha_r$ for all i, and therefore $g_R \leqq \alpha_r$ on $B_{r/2}$, which proves (2.8).

We next show that, if R is chosen sufficiently large, then $g_R = g$ on $B_{r/4}$. By convexity, (2.8), and $g_R \geqq 0$ we obtain $\operatorname{Lip}(g_R; B_{r/4}) \leqq 4\alpha_r/r$. As g_R is CPWL in $B_{r/4}$, for any $y \in B_{r/4}$ there is an affine function $a : \mathbb{R}^d \to \mathbb{R}$ such that $y \in T_a := \{g_R = a\} \cap B_{r/4}$ and T_a has nonempty interior. The Lipschitz bound on g_R then carries over to a, and we obtain $|\nabla a| \leqq 4\alpha_r/r$. By convexity of g_R, we have $a \leqq g_R$, so that $a \leqq f$ on $V \cap B_R$. In order to obtain the same inequality outside B_R, we consider any x with $|x| \geqq R \geqq r$. Then, recalling $y \in T_a \subseteq B_{r/4}$,

$$a(x) \leqq a(y) + |\nabla a|\,|x - y| \leqq \alpha_r + \frac{4\alpha_r}{r}\left(|x| + \frac{r}{4}\right) \leqq \frac{6\alpha_r}{r}|x|.$$

Finally, by (2.6) we can choose $R > \sqrt{d}r$ such that

$$f(x) \geqq \frac{6\alpha_r}{r}|x| \qquad \text{for all } x \in V \setminus B_R. \tag{2.9}$$

Therefore $a \leqq f$ everywhere, which implies $a \leqq g \leqq g_R$, and in turn $g = g_R$ on T_a and therefore on $B_{r/4}$.

We prove now (2.7). Take $x \in \mathbb{R}^d$, so that, by what proved above, $g(x) = g_R(x)$ for some $R > 0$. Now notice that the epigraph of g_R coincides with the convex hull of the epigraph of f_R (here we are using that the convex hull of the epigraph of f_R is closed), so that the conclusion is easily achieved. $\qquad\square$

We next investigate in more detail Delaunay triangulations such that V locally coincides with $\mathbb{Z}^d$ (possibly up to translations and rotations). We show in Lemma 2.13 below that the elements necessarily are the "natural" ones. Before we recall some basic properties of $\mathbb{Z}^d$, where, as usual, for $F \in \mathbb{R}^{d \times d}$, $A \subseteq \mathbb{R}^d$, $p \in \mathbb{R}^d$, we set $p + FA := \{p + Fa : a \in A\}$.

Remark 2.12. The following hold.

i) Let $R \in SO(\mathbb{R}^d)$ and let $\varepsilon \in (0, \infty)$. Then $\operatorname{dist}(x, \varepsilon R\mathbb{Z}^d) \leqq \varepsilon\sqrt{d}/2$ for any $x \in \mathbb{R}^d$.

ii) If $v \subseteq \mathbb{Z}^d$, $\#v = d + 1$, then either v is contained in a $(d - 1)$-dimensional affine subspace, or

$$\mathcal{L}^d(\operatorname{conv} v) \geqq \frac{1}{d!}.$$

iii) If $w \subseteq \mathbb{Z}^d$, $\#w = d$, then either w is contained in a $(d - 2)$-dimensional affine subspace, or

$$\mathcal{H}^{d-1}(\operatorname{conv} w) \geqq \frac{1}{(d - 1)!}. \tag{2.10}$$

Proof. To prove the first item, we can change coordinates to assume that $R = \text{Id}$, and then, by scaling, we see that we can assume $\varepsilon = 1$. For each $i = 1, \ldots, d$ we select $z_i \in \mathbb{Z}$ with $|x_i - z_i| \leqq \frac{1}{2}$, so that $z \in \mathbb{Z}^d$ and

$$|x - z| = \left(\sum_{i=1}^{d} (x_i - z_i)^2 \right)^{1/2} \leqq \sqrt{d}/2.$$

For the second one, by translation we can assume $0 \in v$. The volume of the simplex conv v is given by $1/d!$ times the absolute value of the determinant of the matrix whose columns are the vectors of $v \setminus \{0\}$. As each component of each vector is integer, the determinant is an integer. Hence it is either 0, or at least 1.

The proof of the third item is similar. Again, assume $0 \in w$. At least one e_i is not contained in the linear space generated by w. We apply the first assertion to $v := w \cup \{e_i\}$, and obtain that the volume of $T := \text{conv } v$ is either zero or at least $1/d!$. Since the volume of T is also given by $1/d$ times the area of conv w times the distance of e_i to the space generated by w, which is at most 1 since $0 \in w$, we obtain (2.10). □

Lemma 2.13. *Let (V, E) be a triangulation of $\mathbb{R}^d$ with the Delaunay property and let $B_r(q)$ be a ball such that $V \cap B_r(q) = \varepsilon R \mathbb{Z}^d \cap B_r(q)$, for some $\varepsilon > 0$ and $R \in SO(\mathbb{R}^d)$. If $e \in E$ is such that $e \cap B_{r - \sqrt{d}\varepsilon}(q) \neq \emptyset$, then there is a unique $y \in \varepsilon R(\mathbb{Z} + \frac{1}{2})^d$ such that $v_e \subseteq y + \varepsilon R\{-\frac{1}{2}, \frac{1}{2}\}^d$, characterized by $v_e \subseteq \partial B_{\sqrt{d}/2}(y)$.*

We remark that the assumption $e \cap B_{r - \sqrt{d}\varepsilon}(q) \neq \emptyset$ implies $r > \sqrt{d}\varepsilon$.

Proof. By scaling and a change of coordinates it suffices to consider the case $\varepsilon = 1$, $R = \text{Id}$. Let e be as in the statement, and let $B_\rho(y)$ be such that $v_e \subseteq \partial B_\rho(y)$. By the Delaunay property, using also the assumption in force here,

$$B_\rho(y) \cap \mathbb{Z}^d \cap B_r(q) \subseteq B_\rho(y) \cap V = \emptyset; \qquad (2.11)$$

by $e \cap B_{r - \sqrt{d}}(q) \neq \emptyset$ and $e \subseteq \overline{B}_\rho(y)$ we have

$$|q - y| < r - \sqrt{d} + \rho \quad (\text{and } r > \sqrt{d}). \qquad (2.12)$$

We want to show now that $\rho = \sqrt{d}/2$.

First, we assume (by contradiction) that $\rho > \sqrt{d}/2$. We show that this possibility cannot occur. We define $\rho' := \min\{\rho, r, (r + \rho - |q - y|)/2\}$. Condition (2.12) implies $\rho' > \sqrt{d}/2$ and the definition of ρ' gives

$$|q - y| \leqq -2\rho' + r + \rho = (r - \rho') + (\rho - \rho'),$$

so that there exists $y' \in \overline{B}_{r - \rho'}(q) \cap \overline{B}_{\rho - \rho'}(y)$ (we adopt the convention that $\overline{B}_0(x) = \{x\}$). The point y' obeys then $B_{\rho'}(y') \subseteq B_r(q) \cap B_\rho(y)$ and therefore, recalling (2.11), $B_{\rho'}(y') \cap \mathbb{Z}^d = \emptyset$, which contradicts $\rho' > \sqrt{d}/2$ (Remark 2.12(i)).

Hence $\rho \leqq \sqrt{d}/2$, so that, using also (2.12), $\overline{B}_\rho(y) \subseteq B_r(q)$, and therefore, recalling (2.11), $B_\rho(y) \cap \mathbb{Z}^d = \emptyset$ and $v_e \subseteq \mathbb{Z}^d$. We define $z \in \mathbb{Z}^d$ by choosing for each i a component $z_i \in \mathbb{Z}$ which minimizes $|z_i - y_i|$, notice that $|z_i - y_i| \leqq 1/2$. As $B_\rho(y) \cap \mathbb{Z}^d = \emptyset$, we have $|z - y| \geqq \rho$. By minimality of z_i, for any $x \in$
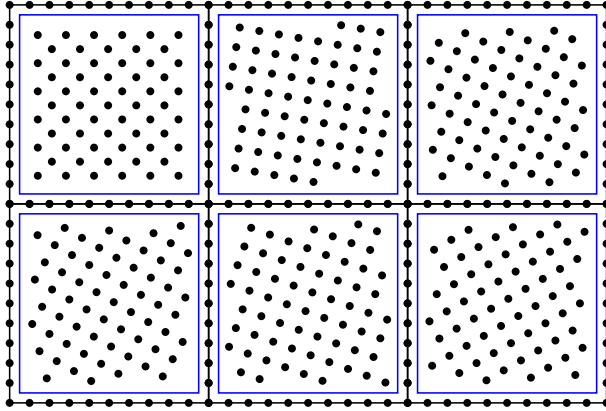
Fig. 3. Sketch of the set of vertices V built in Theorem 2.14. The blue squares indicate the irregular regions where V_{mid} is used

$v_e \subseteq \mathbb{Z}^d$ and any i we have $|x_i - y_i| \geqq |z_i - y_i|$, which by $x \in \partial B_\rho(y)$ implies $\rho = |x - y| \geqq |z - y| \geqq \rho$. Therefore, equality holds throughout and

$$\rho = |x - y| = |z - y| \text{ and } |x_i - y_i| = |z_i - y_i| \qquad \text{for every } i \in \{1, \dots, d\} \text{ and } x \in v_e.$$

Assume that there exists i with $|z_i - y_i| < \frac{1}{2}$, so that $|z_i - x_i| < 1$ for all $x \in v_e$. As $x_i, z_i \in \mathbb{Z}$, this implies $x_i = z_i$ for all $x \in v_e$, hence v_e is contained in a $(d-1)$-dimensional subspace of $\mathbb{R}^d$. As e is non degenerate (i.e. has non empty interior), this is impossible, hence $|z_i - y_i| = \frac{1}{2}$ for all i. We conclude that $\rho = \sqrt{d}/2$ and then $v_e \subseteq y + \{-\frac{1}{2}, \frac{1}{2}\}^d$, which also implies the membership of y to $(\mathbb{Z} + 1/2)^d$ by $v_e \subseteq \mathbb{Z}^d$. □

2.2. Construction of the triangulation

We write $Q_\ell(x) := x + (-\ell/2, \ell/2)^d$ and $Q_\ell := Q_\ell(0)$. Notice the factor $1/2$, i.e. ℓ is the length of the edge of the open cube $Q_\ell(x)$.

Aim of this section is to prove the following (see Figs. 2 and 3 for illustrations):

Theorem 2.14. *For any $d \geqq 2$ there is $C_G = C_G(d)$ with the following property.*

Let $0 < \varepsilon < \delta$ with $\delta \geqq C_G \varepsilon$, and let $R : \delta \mathbb{Z}^d \to SO(\mathbb{R}^d)$. Then there is a triangulation (V, E) of $\mathbb{R}^d$, in the sense of Definition 2.7, with the following properties:

  i) *Regularity: The triangulation has the Delaunay property (property (a)), is C_G-non degenerate (property (b)), and is (C_G, ε)-uniform (property (c)).*
  ii) *Orientation: for each $z \in \delta \mathbb{Z}^d$ one has $V \cap Q_{\delta - C_G \varepsilon}(z) = \varepsilon R(z) \mathbb{Z}^d \cap Q_{\delta - C_G \varepsilon}(z)$.*

We start by proving that in a single cube we can construct a set of vertices V which coincides with $\varepsilon \mathbb{Z}^d$ on the boundary, with a rotation of the same lattice inside, and which is uniform and non-degenerate, in a sense made precise in the statement below. This will then be used to prove Theorem 2.14.
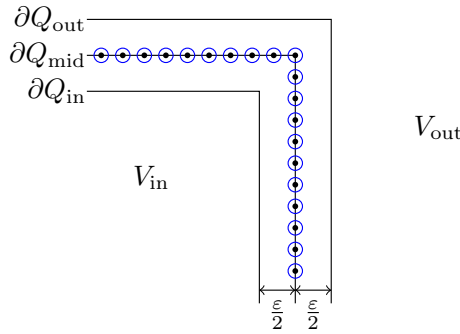
Fig. 4. Sketch of the boundary region as considered in Lemma 2.15

Lemma 2.15. *Let $z \in \mathbb{R}^d$, $\varepsilon > 0$, $R \in SO(\mathbb{R}^d)$, $M \in \mathbb{N}$ with $M \geqq 6 + 2d$. Then there is $V \subseteq \mathbb{R}^d$ with the following properties:*

i) *Orientation: $V \setminus Q_{M\varepsilon}(z) = \varepsilon \mathbb{Z}^d \setminus Q_{M\varepsilon}(z)$ and $V \cap Q_{(M-2)\varepsilon}(z) = R\varepsilon\mathbb{Z}^d \cap Q_{(M-2)\varepsilon}(z)$;*

ii) *$(2d, \varepsilon)$-uniformity: for any $q \in \mathbb{R}^d$ we have $B_{2d\varepsilon}(q) \cap V \neq \emptyset$; for any $x \neq y \in V$ we have $|x - y| \geqq \varepsilon/(2d)$;*

iii) *Non-degeneracy: There is $C' = C'(d)$ such that if $v \subseteq V$, $\#v = d + 1$, v is not contained in a $(d-1)$-dimensional affine subspace, and there is a ball $B_r(y)$ with $v \subseteq \partial B_r(y)$, $B_r(y) \cap V = \emptyset$, then $\mathscr{L}^d(\operatorname{conv} v) \geqq \varepsilon^d / C'$.*

Proof. We divide the proof in several steps.

Step 1: general setting. To simplify notation we denote by $Q_{\mathrm{out}} := Q_{M\varepsilon}(z)$ the outer cube, by $Q_{\mathrm{in}} := Q_{(M-2)\varepsilon}(z)$ the inner cube, and by $Q_{\mathrm{mid}} := Q_{(M-1)\varepsilon}(z)$ the intermediate one (see Fig. 4). We set $V_{\mathrm{out}} := \varepsilon\mathbb{Z}^d \setminus Q_{\mathrm{out}}$; $V_{\mathrm{in}} := R\varepsilon\mathbb{Z}^d \cap \overline{Q}_{\mathrm{in}}$, and shall construct below a finite set $V_{\mathrm{mid}} \subseteq Q_{(M-\frac{1}{2})\varepsilon}(z) \setminus Q_{(M-\frac{3}{2})\varepsilon}(z)$ such that

$$V := V_{\mathrm{in}} \cup V_{\mathrm{out}} \cup V_{\mathrm{mid}}$$

has the desired properties. The property i) is true for any choice of V_{mid}. Next we deal with ii), and leave the more delicate treatment of iii) at the end.

We show that for any $q \in \mathbb{R}^d$ one has $B_{2d\varepsilon}(q) \cap (V_{\mathrm{in}} \cup V_{\mathrm{out}}) \neq \emptyset$. Consider first the case $q \in Q_{\mathrm{mid}}$. Let q' be the point of $\overline{Q}_{(M-2-\sqrt{d})\varepsilon}(z)$ closest to q. This implies

$$|q - q'| \leqq \frac{1}{2}\sqrt{d}(1 + \sqrt{d})\varepsilon \tag{2.13}$$

and $B_{\sqrt{d}\varepsilon/2}(q') \subseteq Q_{\mathrm{in}}$. By Remark 2.12, we can take $p \in R\varepsilon\mathbb{Z}^d \cap \overline{B}_{\sqrt{d}\varepsilon/2}(q') \subseteq V_{\mathrm{in}}$. Since by (2.13)

$$2d\varepsilon > |q - q'| + \sqrt{d}\varepsilon/2$$

we have $p \in \overline{B}_{\sqrt{d}\varepsilon/2}(q') \subseteq B_{2d\varepsilon}(q)$, and the first assertion in ii) is proved in this case. In the case $q \notin Q_{\mathrm{mid}}$ we argue similarly, projecting onto $\mathbb{R}^d \setminus Q_{(M+\sqrt{d})\varepsilon}(z)$,

with $\mathbb{R}^d \setminus Q_{\text{out}}$ instead of $\overline{Q}_{\text{in}}$. Therefore the first assertion in ii) is true for any choice of V_{mid}.

It remains to choose V_{mid} so that the property $|x - y| \geq \varepsilon/(2d)$ for all $x \neq y \in V$ (i.e. the second assertion in ii)) is preserved, and iii) holds. In order to understand the strategy (cf. iii)), consider a set v and a ball $B_r(y)$ such that

$$v \subseteq V \text{ with } \#v = d + 1, v \subseteq \partial B_r(y), V \cap B_r(y) = \emptyset. \qquad (2.14)$$

The construction strategy of V_{mid} then will ensure that:

(a) sets v as in (2.14) cannot contain elements of both V_{in} and V_{out};
(b) for any choice of v as in (2.14), with additionally $v \subseteq V_{\text{in}} \cup V_{\text{mid}}$ or $v \subseteq V_{\text{out}} \cup V_{\text{mid}}$, it holds that v is either contained in a $(d - 1)$-dimensional affine subspace or obeys $\mathscr{L}^d(\text{conv } v) \geq \varepsilon^d / C'$.

Step 2: construction of U_ε. We show here that there is a finite set $U_\varepsilon \subseteq \partial Q_{\text{mid}}$ such that if the set V_{mid} is constructed picking exactly one point of each $B_{\varepsilon/(4d)}(u)$, for $u \in U_\varepsilon$, then (a) and the second assertion in ii) hold. The specific choice of the points of $B_{\varepsilon/(4d)}(u)$, for $u \in U_\varepsilon$, will be done in Step 3 to ensure (b) of (and hence iii), by (a)).

We let $U_\varepsilon := \partial Q_{\text{mid}} \cap (\frac{1}{d}\varepsilon\mathbb{Z}^d + p)$, where $p := z - \frac{M-1}{2}\varepsilon \sum_i e_i$ is a vertex of Q_{mid}. The shift p is chosen so that the set is nonempty; we recall that Q_{mid} is a cube of side length $(M - 1)\varepsilon \in \varepsilon\mathbb{Z}$, but the centre z is a generic point in $\mathbb{R}^d$.

Assume now that V_{mid} is chosen so that it contains exactly one point of each $B_{\varepsilon/(4d)}(u)$, for $u \in U_\varepsilon$. We claim that then V satisfies also the second assertion in ii). Let indeed $x, y \in V$, $x \neq y$. If both are in V_{in}, or both in V_{out}, then $|x - y| \geq \varepsilon$. If both are in V_{mid}, then there are $u_x \neq u_y \in U_\varepsilon$ with $|u_x - x| + |u_y - y| \leq \varepsilon/(2d)$. As $u_x - u_y \in \frac{1}{d}\varepsilon\mathbb{Z}^d \setminus \{0\}$, we obtain

$$|x - y| \geq |u_x - u_y| - |u_x - x| - |u_y - y| \geq \varepsilon/(2d).$$

In the other cases, we use

$$\text{dist}(V_{\text{out}}, V_{\text{mid}}) \geq \text{dist}(\partial Q_{\text{out}}, \partial Q_{\text{mid}}) - \varepsilon/(4d) = \varepsilon/2 - \varepsilon/(4d) \geq \varepsilon/4$$

and similarly $\text{dist}(V_{\text{in}}, V_{\text{mid}}) \geq \varepsilon/4$ to conclude. This proves the second assertion in ii).

We finally check that (a) holds. Let $v \subseteq V$ be as in (2.14). Assume by contradiction that v contains elements of both V_{in} and V_{out}, then the sphere $\partial B_r(y)$ intersects both ∂Q_{out} and ∂Q_{in}. We show that there exists $x' \in \partial Q_{\text{mid}}$ such that $B_{\varepsilon/2}(x') \subseteq B_r(y)$. Assume first $y \in Q_{\text{mid}}$. Let $y' \in \partial B_r(y) \cap \partial Q_{\text{out}}$, and choose $x' \in [y, y'] \cap \partial Q_{\text{mid}}$. Then $|x' - y'| \geq \varepsilon/2$, so that

$$|x' - y| = |y - y'| - |x' - y'| \leq r - \varepsilon/2$$

and $B_{\varepsilon/2}(x') \subseteq B_r(y)$. If instead $y \notin Q_{\text{mid}}$, we select $y' \in \partial B_r(y) \cap \partial Q_{\text{in}}$, and proceed analogously. Let x be the point in U_ε closest to x'. As every component x_i is the element of $\frac{1}{d}\varepsilon\mathbb{Z} + p_i$ closest to x_i', we have $|x - x'| \leq \sqrt{d}\varepsilon/(2d) = \varepsilon/(2\sqrt{d})$. As $\frac{1}{2} > \frac{1}{4d} + \frac{1}{2\sqrt{d}}$, we obtain $B_{\varepsilon/(4d)}(x) \subseteq B_{\varepsilon/2}(x') \subseteq B_r(y)$. As $x \in U_\varepsilon$, there

is a point of V_{mid} in $B_{\varepsilon/(4d)}(x)$, which contradicts the condition $V \cap B_r(y) = \emptyset$ stated in (2.14). Therefore this cannot happen, and hence (a) holds.

Step 3: choice of the elements of V_{mid}. We write $\{u_1, \dots, u_J\} := U_\varepsilon$ and iteratively for every j pick a point $z_j \in B_{\varepsilon/(4d)}(u_j)$ which ensures (b). We collect in $V^j_{\text{mid}} := \{z_1, \dots, z_j\}$ the points chosen in the first j steps, and at the end we will use $V_{\text{mid}} := V^J_{\text{mid}}$. Fix

$$\ell := 1 + 2d, \tag{2.15}$$

the reason for this specific choice will be clear later.

An admissible set of vertices at stage j is a set v with $\#v = d + 1$ such that there is $q \in \partial Q_{\text{mid}}$ with $v \subseteq B_{\ell\varepsilon}(q)$, $\mathscr{L}^d(\text{conv } v) > 0$, and either $v \subseteq V^j_{\text{mid}} \cup V_{\text{in}}$ or $v \subseteq V^j_{\text{mid}} \cup V_{\text{out}}$.

An admissible face at stage j is a set w with $\#w = d$ such that there is $q \in \partial Q_{\text{mid}}$ with $w \subseteq B_{\ell\varepsilon}(q)$, $\mathcal{H}^{d-1}(\text{conv } w) > 0$, and either $w \subseteq V^j_{\text{mid}} \cup V_{\text{in}}$ or $w \subseteq V^j_{\text{mid}} \cup V_{\text{out}}$. We denote by $N_w := \#(w \cap V^j_{\text{mid}})$ the number of items of w in V^j_{mid}, clearly $N_w \leqq d$.

We intend to show that there are α, β, γ, $C_F > 0$ (depending only on d) such that we can choose $z_j \in B_{\varepsilon/(4d)}(u_j)$ iteratively with the following two properties:

(i) If v is an admissible set of vertices at stage j, then

$$\mathscr{L}^d(\text{conv } v) \geqq \beta\varepsilon^d. \tag{2.16}$$

(ii) If w is an admissible face at stage j, then

$$\mathcal{H}^{d-1}(\text{conv } w) \geqq \frac{\alpha^{N_w}}{C_F}\varepsilon^{d-1}. \tag{2.17}$$

The key to the choice of z_j, which eventually leads to (2.16) at stage j building upon (2.17) at stage $j-1$, is the following geometric observation. If v is an admissible set of vertices at stage j, and it contains the point z_j, then $w := v\backslash\{z_j\}$ is an admissible face at stage $j-1$ and for any $q \in w$ we have

$$\mathscr{L}^d(\text{conv } v) = \frac{1}{d}|(z_j - q) \cdot v_w|\mathcal{H}^{d-1}(\text{conv } w) \tag{2.18}$$

where v_w is a unit normal to the affine space generated by w. The factor $\mathcal{H}^{d-1}(\text{conv } w)$ will be estimated via (2.17) at stage $j-1$, the choice of z_j needs to ensure that the first factor is not too small, for any possible choice of w.

Now we start choosing $z_1, \dots, z_J$. As stated before, we proceed by iteration. Assume that we have already chosen $z_1, \dots, z_{j-1}$, we want to choose z_j (if $j = 1$ we use $V^0_{\text{mid}} = \emptyset$). Let w be an admissible face at stage $j-1$ such that $w \subseteq B_{(2\ell+1/(4d))\varepsilon}(u_j)$. If no such face exists, choose $z_j := u_j$. Since no two points in V are at distance smaller than $\varepsilon/(2d)$ (by ii)), the number of possible choices of w is bounded by a number K which depends only on d. Let $w_1, \dots, w_K$ be these possible choices. We choose z_j such that

$$|(z_j - p_k) \cdot v_{w_k}| \geqq \gamma\varepsilon \tag{2.19}$$

for all $k = 1, \ldots, K$ and an arbitrary choice of $p_k \in w_k$ (the condition does not depend on the choice of p_k, as v_{w_k} is orthogonal to $p_k - p_k'$ for any $p_k, p_k' \in w_k$). We show now why we can choose such z_j. We observe that

$$\mathcal{L}^d\left(\{z \in B_{\varepsilon/(4d)}(u_j) : |(z - p_k) \cdot v_{w_k}| < \gamma\varepsilon\}\right) \leqq 2\gamma\varepsilon \left(\frac{\varepsilon}{2d}\right)^{d-1} = \gamma 2^{2-d} d^{1-d} \varepsilon^d$$

and thus the total volume of these sets is controlled by $K\gamma 2^{2-d} d^{1-d} \varepsilon^d$. Then we choose γ such that this expression equals $\frac{1}{2}\mathcal{L}^d(B_{\varepsilon/(4d)}(u_j))$ and hence we have a suitable z_j. Continuing in this way, we have thus constructed V_{mid}^J.

It remains to show by induction that the points we constructed have the properties (2.16) and (2.17). Assume first $j = 0$, and recall $V_{\text{mid}}^0 = \emptyset$, so that $N_w = 0$. By Remark 2.12, (2.16) and (2.17) hold provided $C_F \geqq (d - 1)!$ and $\beta \leqq 1/d!$. Assume now that (2.16) and (2.17) hold at stage $j - 1$, we are going to prove that they hold also at stage j.

Let v be an admissible set of vertices at stage j. If $z_j \notin v$, then v was already admissible at stage $j - 1$, hence (2.16) holds. Then we assume that $z_j \in v$, so that $w := v \setminus \{z_j\}$ is an admissible face at stage $j - 1$ and $v \subseteq B_{\ell\varepsilon}(q) \subseteq B_{2\ell\varepsilon}(z_j) \subseteq B_{(2\ell+1/(4d))\varepsilon}(u_j)$, where $q \in \partial Q_{\text{mid}}$ is given by the admissibility of v. In particular, $w \subseteq B_{(2\ell+1/(4d))\varepsilon}(u_j)$, so that (2.19) holds for w in place of w_k. By (2.17) at stage $j - 1$, (2.18), (2.19) and $N_w \leqq d$ we have, provided $\alpha \leqq 1$,

$$\mathcal{L}^d(\text{conv } v) = \frac{1}{d}|(z_j - p) \cdot v_w|\mathcal{H}^{d-1}(\text{conv } w) \geqq \frac{\gamma\alpha^d}{C_F d}\varepsilon^d$$

for any $p \in w$, so that setting $\beta := \min\{\gamma\alpha^d/(C_F d), 1/d!\}$ we obtain (2.16).

Let w be an admissible face at stage j. As above, by the inductive assumption it suffices to consider the case $z_j \in w$. Assume $w \subseteq V_{\text{mid}}^j \cup V_{\text{in}}$, the other case is analogous and will not be treated. Being w admissible, $w \subseteq B_{\ell\varepsilon}(q)$, for some $q \in \partial Q_{\text{mid}}$. Let q' be the point of $\partial Q_{(M-4-\sqrt{d})\varepsilon}(z)$ closest to q, so that $|q - q'| \leqq \sqrt{d}(3 + \sqrt{d})\varepsilon/2$, and choose $p_* \in \varepsilon R\mathbb{Z}^d \cap \overline{B}_{\varepsilon\sqrt{d}/2}(q') \subseteq \overline{Q}_{(M-4)\varepsilon}(z)$ (Remark 2.12). By the choice of ℓ made in (2.15), we get

$$|p_* - q| \leqq |p_* - q'| + |q' - q| \leqq (\sqrt{d} + 3\sqrt{d} + d)\varepsilon/2 < (\ell - 1)\varepsilon.$$

Then the $2d$ points $p_* \pm \varepsilon Re_i$ are all in $B_{\ell\varepsilon}(q) \cap V_{\text{in}}$, and at least one of them is not in the affine space generated by $w \setminus \{z_j\}$. Denote it by p, and set

$$\hat{w} := (w \setminus \{z_j\}) \cup \{p\}.$$

Then $\hat{w}$ is an admissible face at stage $j - 1$, with $N_{\hat{w}} = N_w - 1$ and $\mathcal{H}^{d-1}(\text{conv } \hat{w}) \neq 0$, so that (2.17) holds for $\hat{w}$. Further, $\hat{w} \subseteq B_{\ell\varepsilon}(q) \subseteq B_{2\ell\varepsilon}(z_j) \subseteq B_{(2\ell+1/(4d))\varepsilon}(u_j)$ implies that $\hat{w}$ is one of the faces $w_1, \ldots, w_K$ considered for (2.19), so that the choice of z_j implies that (2.19) holds for $\hat{w}$.

We compute the volume of the simplex with vertices in $\hat{w} \cup \{z_j\} = w \cup \{p\}$ in two different ways:

$$|(z_j - p) \cdot v_{\hat{w}}|\mathcal{H}^{d-1}(\text{conv } \hat{w}) = |(z_j - p) \cdot v_w|\mathcal{H}^{d-1}(\text{conv } w).$$

By (2.19) and (2.17) for $\hat{w}$, recalling that $z_j, p \in B_{\ell\varepsilon}(q)$ implies $|z_j - p| \leqq 2\ell\varepsilon$, we obtain

$$\mathcal{H}^{d-1}(\text{conv } w) \geqq \frac{1}{2\ell\varepsilon}|(z_j - p) \cdot v_{\hat{w}}|\mathcal{H}^{d-1}(\text{conv } \hat{w}) \geqq \frac{\gamma}{2\ell}\alpha^{N_{\hat{w}}}\varepsilon^{d-1}/C_F$$

which concludes the proof of (2.17) with $\alpha := \min\{1, \gamma/(2\ell)\}$. $\quad\square$

At this point we conclude the proof of Theorem 2.14.

Proof of Theorem 2.14. Set

$$\ell := 2d \quad \text{and} \quad M := \lfloor \delta/\varepsilon \rfloor - 4\ell,$$

so that $Q_{M\varepsilon} \subseteq Q_\delta$, with

$$\text{dist}(Q_{M\varepsilon}, \partial Q_\delta) \geqq 2\ell\varepsilon. \tag{2.20}$$

We first select a background lattice,

$$V^0 := \varepsilon\mathbb{Z}^d \setminus \bigcup_{z \in \delta\mathbb{Z}^d} Q_{M\varepsilon}(z).$$

For each $z \in \delta\mathbb{Z}^d$, if $C_G \geqq 7 + 2d + 4\ell$ we can use (by $M \geqq C_G - 1 - 4\ell$) Lemma 2.15 to obtain a set V_z such that $V_z \cap Q_{(M-2)\varepsilon}(z) = R(z)\varepsilon\mathbb{Z}^d \cap Q_{(M-2)\varepsilon}(z)$, and $V_z \setminus Q_{M\varepsilon}(z) = \varepsilon\mathbb{Z}^d \setminus Q_{M\varepsilon}(z)$. We then set

$$V := V^0 \cup \bigcup_{z \in \delta\mathbb{Z}^d} (V_z \cap Q_\delta(z)) = V^0 \cup \bigcup_{z \in \delta\mathbb{Z}^d} (V_z \cap Q_{M\varepsilon}(z)).$$

This set obviously has the orientation property stated in ii), provided that $C_G \geqq 4\ell + 3$.

We show that for any $x \neq y \in V$, one has $|x - y| \geqq \varepsilon/\ell$. Indeed, if there is $z \in \delta\mathbb{Z}^d$ with $x, y \in V_z$ then item ii) of Lemma 2.15 implies $|x - y| \geqq \varepsilon/\ell$. If $x, y \in V^0$ then $|x - y| \geqq \varepsilon$. We are left with the case $x \in Q_{M\varepsilon}(z)$ and $y \in Q_{M\varepsilon}(z')$ for some $z \neq z' \in \delta\mathbb{Z}^d$, which implies $|x - y| \geqq 2\text{dist}(Q_{M\varepsilon}, \partial Q_\delta) \geqq 4\ell\varepsilon \geqq \varepsilon/\ell$, by (2.20).

We next similarly show that for any $q \in \mathbb{R}^d$ one has $V \cap B_{\ell\varepsilon}(q) \neq \emptyset$. If there is $z \in \delta\mathbb{Z}^d$ such that $q \in Q_{(M+2\ell)\varepsilon}(z)$ then $B_{\ell\varepsilon}(q) \subseteq Q_\delta(z)$, and the required property follows from item ii) of Lemma 2.15, since $V \supseteq V_z \cap Q_\delta(z)$. If not, then $B_{\ell\varepsilon}(q)$ does not intersect any $Q_{M\varepsilon}(z)$, so that $B_{\ell\varepsilon}(q) \cap V^0 = B_{\ell\varepsilon}(q) \cap \varepsilon\mathbb{Z}^d$, which is nonempty by Remark 2.12.

This proves that the set V is (ℓ, ε)-uniform, in the sense of Property (c) of Definition 2.7. By Lemma 2.9 there is a set E so that (V, E) is a triangulation with the Delaunay property.

It only remains to show that (V, E) is non-degenerate. Let $e \in E$ be a simplex, and let $\partial B_r(q) \supseteq v_e$ be its circumscribed sphere. By the Delaunay property $B_r(q) \cap V = \emptyset$, by the (ℓ, ε)-uniformity proven above this implies $r < \ell\varepsilon$. If there is $z \in \delta\mathbb{Z}^d$ such that $q \in Q_{(M+2\ell)\varepsilon}(z)$ then $v_e \subseteq V_z$, and item iii) of Lemma 2.15 implies $\mathcal{L}^d(e) \geqq \varepsilon^d/C'$. Otherwise $v_e \subseteq V^0 \subseteq \varepsilon\mathbb{Z}^d$, and since $\mathcal{L}^d(e) > 0$ by Remark 2.12 we obtain $\mathcal{L}^d(e) \geqq \varepsilon^d/d!$. This concludes the proof, with $C_G := \max\{7 + 2d + 4\ell, 4\ell + 3, C', d!\}$. $\quad\square$

2.3. Proof of the main result

We now recall how one can use a triangulation to define continuous, piecewise affine approximations.

Lemma 2.16. *Let (V, E) be a triangulation of $\mathbb{R}^d$. For any $w : V \to \mathbb{R}$ there is a unique $u \in C^0(\mathbb{R}^d)$ which coincides with w on V and is affine on each $e \in E$.*

If the triangulation is c_*-non degenerate, and if moreover w is obtained as the restriction to V of a $C^2(\mathbb{R}^d)$ function that we still denote w, then the function u obtained above obeys

$$\|\nabla u\|_{L^\infty(e)} \leqq C\|\nabla w\|_{L^\infty(e)} \tag{2.21}$$

and

$$\|\nabla w - \nabla u\|_{L^\infty(e)} \leqq C\mathrm{diam}(e)\|\nabla^2 w\|_{L^\infty(e)} \tag{2.22}$$

for all $e \in E$, with C depending on c_* and d.

Proof. For each $e \in E$ one defines $u_e : e \to \mathbb{R}$ by $u_e = w$ on v_e and as the affine interpolation in the rest of $e = \mathrm{conv}(v_e)$. To prove existence of u we only need to check that $u_e = u_{e'}$ on $e \cap e'$, for any pair $e \neq e' \in E$. Assume $e \cap e' \neq \emptyset$. Then $e \cap e' = \mathrm{conv}(v_e \cap v_{e'})$. As $u_e = u_{e'}$ on $v_e \cap v_{e'}$, and both are affine in conv $(v_e \cap v_{e'})$, they coincide on $e \cap e'$. This concludes the proof of the first assertion.

To prove the two estimates, we focus on an element $e \in E$ and let G be the constant gradient of u on e. For any pair $x, y \in v_e$,

$$G(y - x) = u(y) - u(x) = w(y) - w(x) = \int_0^1 \nabla w(x + t(y - x))(y - x)\mathrm{d}t, \tag{2.23}$$

which implies

$$|G(y - x)| \leqq \|\nabla w\|_{L^\infty(e)}|y - x|.$$

With (2.2) we obtain (2.21).

To prove the last estimate, we pick any $\xi \in e$ and rewrite (2.23) as

$$(G - \nabla w(\xi))(y - x) = \int_0^1 (\nabla w(x + t(y - x)) - \nabla w(\xi))(y - x)\mathrm{d}t.$$

By the mean-value theorem $|\nabla w(\eta) - \nabla w(\xi)| \leqq \mathrm{diam}(e)\|\nabla^2 w\|_{L^\infty(e)}$ for any $\eta \in e$, so that

$$|(G - \nabla w(\xi))(y - x)| \leqq \mathrm{diam}(e)\|\nabla^2 w\|_{L^\infty(e)}|y - x|.$$

With (2.2) we obtain (2.22).   □

We are ready to prove our main result, Theorem 2.2.

Proof of Theorem 2.2. Before entering into the proof of the theorem, we stress that we are going to use the fact that for a piecewise affine function u_j,

$$|D_1^2 u_j| = |D\nabla u_j|. \tag{2.24}$$

This follows from the fact that u_j is piecewise affine, hence the distributional derivative of $D\nabla u_j$ is only of jump type, so that the density of $D\nabla u_j$ with respect to $|D\nabla u_j|$ is a rank 1 matrix, and hence we can use item v) of Proposition 1.2 in conjunction with Proposition 1.7.

Fix two sequences $\delta_j \to 0$, $\varepsilon_j \to 0$, with $\delta_j > 0$, $\varepsilon_j > 0$, and $\varepsilon_j/\delta_j \to 0$. For each j and each $z \in \delta_j \mathbb{Z}^d$ we select a matrix $R_z \in SO(\mathbb{R}^d)$ such that $R_z^t \nabla^2 w(z) R_z$ is diagonal, and let (V_j, E_j) be the grid constructed in Theorem 2.14 with these parameters. We define u_j as the piecewise affine interpolation of w, constructed as in Lemma 2.16. This concludes the construction.

In order to prove convergence and the energy bound, it suffices to work in a large ball B_r, with $\Omega \subseteq B_{r/2}$. For large j, we can assume $C_G \varepsilon_j \leqq \delta_j \leqq r/(2d)$. Here and below C_G is the (fixed) constant from Theorem 2.14, we can assume $C_G > 2\sqrt{d}$. We use C for a generic constant that depends only on d (and C_G) and may vary from line to line. By Lemma 2.16 one immediately obtains a uniform Lipschitz bound on u_j,

$$\|\nabla u_j\|_{L^\infty(B_{2r})} \leqq C\|\nabla w\|_{L^\infty(B_{3r})}.$$

By the uniformity property of the grid, for any $x \in B_r$ and any j there is $y \in V_j$ with $|x - y| \leqq C_G \varepsilon_j$, therefore

$$\|w - u_j\|_{L^\infty(B_r)} \leqq C_G \varepsilon_j (\|\nabla u_j\|_{L^\infty(B_{2r})} + \|\nabla w\|_{L^\infty(B_{2r})}) \to 0.$$

This proves local uniform convergence.

Since $\nabla^2 w$ is continuous, one has that

$$\omega_\rho := \sup\{|\nabla^2 w(x) - \nabla^2 w(y)| : x, y \in B_{2r}, |x - y| \leqq \rho\sqrt{d}\} \tag{2.25}$$

converges to zero as $\rho \to 0$.

The estimate of the energy is done separately in the interior of the cubes, where the grid is regular, and in the boundary regions. We start from the boundary, where the grid is irregular. As ∇w is continuous, equation (2.22) in Lemma 2.16 permits to estimate $|[\nabla u_j]|$, the jump in ∇u_j across the boundary between two neighbouring elements e and e' which intersect B_r, and gives

$$|[\nabla u_j]| \leqq C\varepsilon_j \|\nabla^2 w\|_{L^\infty(B_{2r})} \quad \text{in all } e \text{ with } e \cap B_r \neq \emptyset,$$

here we used also Remark 2.8. Using non-degeneracy and uniformity of the triangulation to control the volume of e, we obtain

$$|D\nabla u_j|(\partial e) \leqq C\mathcal{H}^{d-1}(\partial e) \max |[\nabla u_j]|(\partial e) \leqq C\mathcal{L}^d(e)\|\nabla^2 w\|_{L^\infty(B_{2r})}$$

for all elements $e \in E_j$ with $e \subseteq B_r$. Fix now $z \in \delta_j \mathbb{Z}^d$ such that $Q_{\delta_j}(z) \cap \Omega \neq \emptyset$. Summing the previous condition over all elements $e \in E_j$ with $e \cap \overline{Q}_{\delta_j}(z) \backslash Q_{\delta_j - 4C_G \varepsilon_j}(z) \neq \emptyset$ leads to

$$
\begin{aligned}
|\mathrm{D}\nabla u_j|(\overline{Q}_{\delta_j}(z) \setminus Q_{\delta_j - 4C_G\varepsilon_j}(z)) \\
\leq C \mathscr{L}^d(Q_{\delta_j + 4C_G\varepsilon_j}(z) \setminus Q_{\delta_j - 8C_G\varepsilon_j}(z)) \|\nabla^2 w\|_{L^\infty(B_{2r})} \\
\leq C((\delta_j + 4C_G\varepsilon_j)^d - (\delta_j - 8C_G\varepsilon_j)^d) \|\nabla^2 w\|_{L^\infty(B_{2r})} \\
\leq C \delta_j^{d-1} \varepsilon_j \|\nabla^2 w\|_{L^\infty(B_{2r})},
\end{aligned}
\tag{2.26}
$$

provided j is large enough, since $\varepsilon_j \ll \delta_j$. Here we used that for every $e \in E_j$, $\mathrm{diam}(e) \leq 2C_G \varepsilon_j$, being the triangulation (V_j, E_j) (C_G, ε_j)-uniform and with the Delaunay property.

We next estimate the energy inside $Q_{\delta_j - 3C_G\varepsilon_j}(z)$, for some $z \in \delta_j \mathbb{Z}^d \cap B_r$. Let $H_z := \nabla^2 w(z)$, and recall that R_z was chosen so that $R_z^t H_z R_z = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$ for some $\lambda \in \mathbb{R}^d$, which implies $|H_z|_1 = \sum_{i=1}^d |\lambda_i|$, see items i) and ii) of Proposition 1.2. In the next estimates we write briefly δ and ε for δ_j and ε_j.

For any element $e \in E_j$ with $e \cap Q_{\delta - 2C_G\varepsilon}(z) \neq \emptyset$, we can select $p_e \in e \cap Q_{\delta - 2C_G\varepsilon}(z)$. Then $B_{C_G\varepsilon/2}(p_e) \subseteq Q_{C_G\varepsilon}(p_e) \subseteq Q_{\delta - C_G\varepsilon}(z)$, so that the orientation property of Theorem 2.14 gives $B_{C_G\varepsilon/2}(p_e) \cap V_j = B_{C_G\varepsilon/2}(p_e) \cap \varepsilon R_z \mathbb{Z}^d$. Recalling $C_G > 2\sqrt{d}$, by applying Lemma 2.13 with $q = p_e$, $r = C_G\varepsilon/2$, there exists $y \in \varepsilon R_z(\mathbb{Z} + \frac{1}{2})^d$ such that $v_e \subseteq y + \varepsilon R_z\{-\frac{1}{2}, \frac{1}{2}\}^d$. Let $F_y := \nabla w(y)$. For all $x \in v_e$, Taylor remainder term in integral form and (2.25) yield

$$
w(x) = w(y) + F_y(x - y) + \frac{1}{2} H_z(x - y) \cdot (x - y) + R(x)
$$

(this can be seen as the definition of $R(\cdot)$) with

$$
\begin{aligned}
|R(x)| \leq d\varepsilon^2 |\nabla^2 w(y) - H_z| + \int_0^1 |\nabla^2 w(x + t(y - x)) \\
- \nabla^2 w(y)| \, |y - x|^2 dt \leq C\varepsilon^2 \omega_\delta.
\end{aligned}
\tag{2.27}
$$

As $x - y = \sum_i \varepsilon \gamma_i R_z e_i$, with $\gamma_i \in \{-\frac{1}{2}, \frac{1}{2}\}$, recalling that $R_z^t H_z R_z = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$ we have

$$
H_z(x - y) \cdot (x - y) = \varepsilon^2 \sum_{i,k=1}^d \gamma_i \gamma_k e_i R_z^t H_z R_z e_k = \frac{1}{4} \varepsilon^2 \sum_{i=1}^d \lambda_i
$$

which does not depend on the γ_i, and therefore is the same for all $x \in v_e$. Hence

$$
w(x) = w(y) + F_y(x - y) + \frac{1}{8} \varepsilon^2 \sum_{i=1}^d \lambda_i + R(x) \qquad \text{for all } x \in v_e.
$$

The function u_j is affine on the element e, assume it has the form $u_j(\xi) = a_e + G_e \xi$ for $\xi \in e$. As $u_j = w$ on v_e, for every pair $x, x' \in v_e$ we obtain

$$
G_e(x - x') = u_j(x) - u_j(x') = w(x) - w(x') = F_y(x - x') + R(x) - R(x').
$$

Recalling that e is a non-degenerate simplex by (2.2), (2.27) and what just proved we obtain

$$|G_e - F_y| \leqq C\varepsilon\omega_\delta. \tag{2.28}$$

In summary, if $e \in E_j$ obeys $e \cap Q_{\delta-2C_G\varepsilon}(z) \neq \emptyset$ then there exists $y_e \in \varepsilon R_z(\mathbb{Z}+\frac{1}{2})^d$ with $v_e \subseteq y_e+\varepsilon R_z\{-\frac{1}{2}, \frac{1}{2}\}^d$, and the vector $G_e := \nabla u_{j|_e}$ obeys (2.28).

Consider now some $y \in \varepsilon R_z(\mathbb{Z}+\frac{1}{2})^d$ such that $(y+R_zQ_\varepsilon) \cap Q_{\delta-4C_G\varepsilon}(z) \neq \emptyset$. If e, e' are two elements with $v_e, v_{e'} \subseteq y + R_z\overline{Q}_\varepsilon$, then (by $C_G > \sqrt{d}$) both intersect $Q_{\delta-2C_G\varepsilon}(z)$, so that the above discussion applies and (2.28) gives $|G_e - G_{e'}| \leqq C\varepsilon\omega_\delta$, having used that the above discussion forces $y = y_e$ (since $y, y_e \in \varepsilon R_z(\mathbb{Z}+\frac{1}{2})^d$ and $y \neq y_e$ imply that $(y + R_z\overline{Q}_\varepsilon) \cap (y_e + \varepsilon R_z\{-\frac{1}{2}, \frac{1}{2}\}^d) \supseteq v_e$ has at most dimension $d-1$) and analogously $y = y_{e'}$. In particular, those elements constitute a decomposition of $y + R_zQ_\varepsilon$. Arguing as before, summing over all pairs,

$$|D\nabla u_j|(y + R_zQ_\varepsilon) \leqq C\varepsilon^{d-1} \max |G_e - G_{e'}| \leqq C\varepsilon^d\omega_\delta. \tag{2.29}$$

In order to estimate the contribution from the boundary of these cubes, let $y' = y \pm \varepsilon R_ze_i$ be the centre of one of the neighbouring small cubes. Since $C_G > 2\sqrt{d}$, $y' + R_zQ_\varepsilon \subseteq Q_{\delta-2C_G\varepsilon}(z)$, so that (2.28) holds for any element e'' contained in $y' + R_z\overline{Q}_\varepsilon$ (with e'' in place of e and y' in place of y). As the common boundary has area ε^{d-1},

$$|D\nabla u_j|(\partial(y + R_zQ_\varepsilon)) \leqq C\varepsilon^d\omega_\delta + \sum_{y'\in y+R_z\varepsilon\{\pm e_1,...,\pm e_d\}} \varepsilon^{d-1}|F_y - F_{y'}|.$$

As we did before, we represent $F_{y'} - F_y = \nabla w(y') - \nabla w(y)$ with Taylor's theorem

$$F_{y'} = F_y + H_z(y' - y) + R'(y', y) \quad \text{and} \quad |R'(y', y)| \leqq C\varepsilon\omega_\delta$$

(this can be seen as the definition of $R'(\cdot, \cdot)$) to obtain

$$|D\nabla u_j|(\partial(y + R_zQ_\varepsilon)) \leqq C\varepsilon^d\omega_\delta + \sum_{y'\in y+R_z\varepsilon\{\pm e_1,...,\pm e_d\}} \varepsilon^{d-1}|H_z(y' - y)|$$

$$= C\varepsilon^d\omega_\delta + 2\varepsilon^d|H_z|_1 \leqq C\varepsilon^d\omega_\delta + 2\int_{y+R_zQ_\varepsilon} |\nabla^2 w|_1 \mathrm{d}\mathscr{L}^d, \tag{2.30}$$

where we used that the R_ze_i are eigenvectors of H_z by the choice of R_z, the definition of the Schatten norm and in the final step (2.25). Let

$$A_z := \left\{ y \in \varepsilon R_z\left(\mathbb{Z}+\frac{1}{2}\right)^d : (y + R_zQ_\varepsilon) \cap Q_{\delta-4C_G\varepsilon}(z) \neq \emptyset \right\}.$$

Summing over all $y \in A_z$, taking into account (2.29) and (2.30) and recalling that the boundaries between the cubes appear twice in the sum, gives

$$|D\nabla u_j|(Q_{\delta-4C_G\varepsilon}(z)) \leq C\delta^d\omega_\delta + \int_{Q_\delta(z)} |\nabla^2 w|_1 \mathrm{d}\mathscr{L}^d$$

and combining with (2.26)

$$|\mathrm{D}\nabla u_j|(\overline{Q}_\delta(z)) \leqq C\delta^d \left(\omega_\delta + \frac{\varepsilon}{\delta}\|\nabla^2 w\|_{L^\infty(B_{2r})}\right) + \int_{Q_\delta(z)} |\nabla^2 w|_1 \mathrm{d}\mathscr{L}^d.$$

Summing over all z such that $Q_\delta(z) \cap \Omega \neq \emptyset$, and inserting back the indices j,

$$|\mathrm{D}\nabla u_j|(\Omega) \leqq C|(\Omega)_{\delta_j}| \left(\omega_{\delta_j} + \frac{\varepsilon_j}{\delta_j}\|\nabla^2 w\|_{L^\infty(B_{2r})}\right) + \int_{(\Omega)_{\delta_j}} |\nabla^2 w|_1 \mathrm{d}\mathscr{L}^d$$

where $(\Omega)_\rho := \{x \in \mathbb{R}^d : \mathrm{dist}(x, \Omega) \leqq \rho\sqrt{d}\}$. Taking the limit $j \to \infty$, and recalling that $\delta_j \to 0$, $\omega_{\delta_j} \to 0$ and $\varepsilon_j/\delta_j \to 0$, concludes the proof (recalling (2.24)). $\square$

3. Extremality of Cones

In this section we consider functions of the kind

$$f^{\mathrm{cone}}(x) := (1 - |x|)_+. \tag{3.1}$$

It is clear that our forthcoming discussion will apply also to slightly different functions, e.g. $a(1 - b|x - x_0|)_+$ for $a, b \in \mathbb{R}$ with $b > 0$ and $x_0 \in \mathbb{R}^d$, but this will not make much difference, as one can reduce to the particular case of (3.1) via a change of coordinates and a rescaling. Notice that, by Proposition 1.13, if $d \geqq 2$,

$$|\mathrm{D}_p^2 f^{\mathrm{cone}}|(B_r(0)) = d\omega_d\left((d-1)^{1/p-1}(r \wedge 1)^{d-1} + \chi_{(1,\infty)}(r)\right). \tag{3.2}$$

Our aim is to investigate extremality of such kind of functions with respect to p-Hessian–Schatten seminorms, for $p \in [1, \infty]$. It turns out that these functions are extremal, and now we state our main result in this direction. Its proof is deferred to Section 3.3 and will follow easily from the results of Sections 3.1 and 3.2, taking into account also Section 1.3.

Theorem 3.1. *Let $d \geqq 2$ and let $p \in [1, \infty)$. Let f_1, $f_2 \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$ with bounded Hessian–Schatten variation in $\mathbb{R}^d$ such that*

$$|\mathrm{D}_p^2 f_1|(\mathbb{R}^d) = |\mathrm{D}_p^2 f_2|(\mathbb{R}^d) = |\mathrm{D}_p^2 f^{\mathrm{cone}}|(\mathbb{R}^d)$$

and such that for some $\lambda \in (0, 1)$,

$$f^{\mathrm{cone}} = \lambda f_1 + (1 - \lambda) f_2.$$

Then f_1 and f_2 are equal to f^{cone}, up to affine terms: there exist affine functions $L_1, L_2 : \mathbb{R}^d \to \mathbb{R}$ such that $f_i = f^{\mathrm{cone}} + L_i$ for $i = 1, 2$.

Notice that Theorem 3.1 is stated only for $d \geq 2$. Indeed, for $d = 1$, it is easy to realize that f^{cone} is *not* extremal, according to the meaning described in the statement of the theorem.

To simplify the notation, as in this section we are going to consider only balls centred at the origin, we will omit to write the centre of the ball, i.e. $B_r := B_r(0)$. Before going on, we recall that given $f \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$, we denote by f^{rad} the function given by Lemma 1.10. As an explicit expression, notice that

$$f^{\mathrm{rad}}(x) = \fint_{\partial B_{|x|}} f(\sigma) \, \mathrm{d}\mathscr{H}^{d-1}(\sigma) \quad \text{for } \mathscr{L}^d\text{-a.e. } x. \tag{3.3}$$

Notice also that $f^{\mathrm{rad}}(x) = g(|x|)$ for $g(r)$ given by the right hand side of (3.3) with r in place of $|x|$.

3.1. Convexity

We prove that if a function $f \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$ is such that $f^{\mathrm{rad}} = f^{\mathrm{cone}}$ and such that $|\mathrm{D}^2_p f|(\mathbb{R}^d) = |\mathrm{D}^2_p f^{\mathrm{cone}}|(\mathbb{R}^d)$, then f is the cone. The case $p = 1$ is treated in Proposition 3.5, using the fact that the absolutely continuous part of $\mathrm{D}\nabla f$ has a sign, which makes f concave inside the unit ball. The case $p > 1$ is treated in Proposition 3.6, using strict convexity of the p-Schatten norm to show that the absolutely continuous part of $\mathrm{D}\nabla f$ is a scalar multiple of the absolutely continuous part of $\mathrm{D}\nabla f^{\mathrm{cone}}$, and then scaling to reduce to the $p = 1$ case.

First, we need a couple of lemmas. The first is an extension of a well known criterion to recognize convexity.

Lemma 3.2. *Let $\Omega \subseteq \mathbb{R}^d$ be open and convex and let $f \in L^1_{\mathrm{loc}}(\Omega)$ with bounded Hessian–Schatten variation in Ω. Assume that $\mathrm{D}\nabla f \geq 0$ (as a measure with values in symmetric matrices). Then f has a representative which is continuous and convex.*

Proof. The property of having a continuous representative is clearly local. Since Ω is open and convex, a continuous function $g : \Omega \to \mathbb{R}$ is convex if and only if it is convex in a neighbourhood of any point. Therefore it suffices to prove the assertion in a neighbourhood of any point, so that we can assume $f \in W^{1,1}(\Omega)$ with $\nabla f \in \mathrm{BV}(\Omega; \mathbb{R}^d)$, by Proposition 1.11 and Proposition 1.7.

Let $x \in \Omega$, and pick $r > 0$ such that $Q_{4r}(x) \subseteq \Omega$ (we write here $Q_\ell(y) := y + (-\ell, \ell)^n$). Fix a mollifier $\eta_\varepsilon \in C^\infty_c(B_\varepsilon; [0, \infty))$, with $\varepsilon \leq r$, and define $f_\varepsilon := \eta_\varepsilon * f \in C^\infty(Q_{3r}(x))$. Then an immediate computation yields $\mathrm{D}\nabla f_\varepsilon = \eta_\varepsilon * \mathrm{D}\nabla f \geq 0$ in $Q_{3r}(x)$, therefore f_ε is convex in $Q_{3r}(x)$. Further, $f_\varepsilon \to f$ in $W^{1,1}(Q_{3r})$. It remains to show that f_ε (possibly after passing to a subsequence) converges uniformly in Q_r, which implies the conclusion in Q_r and therefore in a neighbourhood of any point of Ω.

We prove now uniform convergence in Q_r, the argument is classical, see e.g. the proof of [13, Theorem 7.6]. Passing to a subsequence, $f_{\varepsilon_j} \to f$ pointwise almost everywhere. Pick $\bar{x} \in Q_{r/2}(x)$ such that the sequences $f_{\varepsilon_j}(\bar{x})$ and $f_{\varepsilon_j}(y)$, for any vertex y of $Q_{2r}(\bar{x}) \subseteq Q_{3r}(x)$, are bounded (as we can assume them to be

convergent), and let $M = M_{\bar{x},r}$ be the common bound. By convexity, $f_{\varepsilon_j} \leqq M$ on $\bar{Q}_{2r}(\bar{x})$. To prove the uniform lower bound, we observe that for any $w \in Q_{2r}(\bar{x}) \backslash \{\bar{x}\}$ there is $z \in \partial Q_{2r}(\bar{x})$ such that $\bar{x}$ is in the interior of the segment joining w with z. As convexity implies monotonicity of the difference quotients,

$$\frac{f_{\varepsilon_j}(\bar{x}) - f_{\varepsilon_j}(w)}{|\bar{x} - w|} \leqq \frac{f_{\varepsilon_j}(z) - f_{\varepsilon_j}(\bar{x})}{|z - \bar{x}|} \leqq \frac{2M}{2r},$$

where in the last step we used $|z - \bar{x}| \geqq 2r$. Since $f_{\varepsilon_j}(\bar{x}) \geqq -M$ and $|w - \bar{x}| \leqq 2r\sqrt{d}$ we have $f_{\varepsilon_j}(w) \geqq -(1 + 2\sqrt{d})M$. Passing to the smaller cube $Q_r(x)$ and using again monotonicity of the difference quotients we obtain $\text{Lip}(f_{\varepsilon_j}; Q_r(x)) \leqq C'M$ for all j, so that f_{ε_j} converges uniformly in $Q_r(x)$ to a continuous convex function, which coincides almost everywhere with f. This concludes the proof. □

The next lemma builds upon Lemma 3.2 and gives an integral characterization of convexity, which is more manageable, and follows from the rigidity in the inequality $|\text{Tr} A| \leqq |A|_1$.

Lemma 3.3. *Let $\Omega \subseteq \mathbb{R}^d$ be open and let $f \in L^1_{\text{loc}}(\Omega)$ with bounded Hessian–Schatten variation in Ω. Then*

$$|D_1^2 f|(\Omega) \geqq |\text{Tr} D\nabla f(\Omega)|. \tag{3.4}$$

Assume now that equality in (3.4) holds. Then

- *either $|D_1^2 f|(\Omega) = \text{Tr} D\nabla f(\Omega)$ and then f has a representative which is continuous and convex,*
- *or $|D_1^2 f|(\Omega) = -\text{Tr} D\nabla f(\Omega)$ and then f has a representative which is continuous and concave.*

Proof. We can assume that $\text{Tr} D\nabla f(\Omega) \geqq 0$, otherwise one replaces f by $-f$.

Let now $A \in \mathbb{R}^{d \times d}$ be a symmetric matrix and let $\lambda_1, \ldots, \lambda_d$ denote its eigenvalues. By item i) of Proposition 1.2,

$$|A|_1 = \sum_{i=1}^d |\lambda_i| \geqq \sum_{i=1}^d \lambda_i = \text{Tr} A$$

and equality holds if and only if $\lambda_i \geqq 0$ for all i, which is the same as $A \geqq 0$ as a symmetric matrix.

By Proposition 1.7 (in particular, $|D_1^2 f| \ll |D\nabla f|$ and $\text{Tr} D\nabla f \ll |D\nabla f|$),

$$|D_1^2 f|(\Omega) = \int_\Omega \left| \frac{dD\nabla f}{d|D\nabla f|} \right|_1 d|D\nabla f| \geqq \int_\Omega \text{Tr} \frac{dD\nabla f}{d|D\nabla f|} d|D\nabla f| = \text{Tr} D\nabla f(\Omega),$$

which proves the bound (3.4). If equality holds, then

$$\left| \frac{dD\nabla f}{d|D\nabla f|} \right|_1 = \text{Tr} \frac{dD\nabla f}{d|D\nabla f|} \qquad |D\nabla f|\text{-a.e.}$$

so that

$$\frac{dD\nabla f}{d|D\nabla f|} \geqq 0 \qquad |D\nabla f|\text{-a.e.}$$

which means that $D\nabla f \geqq 0$ as a matrix-valued measure, so that the conclusion then follows by Lemma 3.2. $\square$

3.2. Extremality with respect to spherical averaging

In this section, we consider only the case $d \geqq 2$. This is because this is an auxiliary section for the proof of Theorem 3.1, which holds only for $d \geqq 2$. We start by doing some explicit computation involving the Hessian–Schatten total variation of f^{cone}. First, by Proposition 1.7, $f^{\mathrm{cone}} \in W^{1,1}(\mathbb{R}^d)$ with $\nabla f^{\mathrm{cone}} \in BV(\mathbb{R}^d; \mathbb{R}^d)$. More precisely,

$$\nabla f^{\mathrm{cone}}(x) = -\chi_{B_1}(x)\frac{x}{|x|}.$$

This computation is easily justified by locality, as f^{cone} is smooth on $B_1 \setminus \{0\}$ and on $\mathbb{R}^d \setminus \bar{B}_1$. Now we claim that

$$D\nabla f^{\mathrm{cone}}(x) = -\frac{|x|^2\mathrm{Id} - x \otimes x}{|x|^3}\mathscr{L}^d \llcorner B_1 + (x \otimes x)\mathscr{H}^{d-1} \llcorner \partial B_1. \quad (3.5)$$

Taking into account that $D\nabla f^{\mathrm{cone}}$ does not charge points, this formula is easily justified on $\mathbb{R}^d \setminus \partial B_1$ by locality, as above. For what concerns the singular part, on ∂B_1, it is enough to use the representation formula for the singular part of differentials of vector valued functions of bounded variation, e.g. [3], notice indeed that the unit outer normal to ∂B_1 is x and that the jump of ∇f^{cone} at $x \in \partial B_1$ is exactly x.

Taking traces, we have that

$$\mathrm{Tr}D\nabla f^{\mathrm{cone}}(x) = \frac{(1 - d)}{|x|}\mathscr{L}^d \llcorner B_1 + \mathscr{H}^{d-1} \llcorner \partial B_1,$$

so that

$$\int_{B_r} \mathrm{d}Tr D\nabla f^{\mathrm{cone}} = -d\omega_d r^{d-1}\chi_{(0,1]}(r) \qquad \forall r > 0. \quad (3.6)$$

Recall that by Lemma 1.10, $|D_p^2 f^{\mathrm{rad}}|(\mathbb{R}^d) \leqq |D_p^2 f|(\mathbb{R}^d)$. The next lemma states that this inequality is somehow rigid.

Lemma 3.4. *Let* $p \in [1, \infty]$. *Let* $f \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$ *with bounded Hessian–Schatten variation and assume that*

$$|D_p^2 f^{\mathrm{rad}}|(\mathbb{R}^d) = |D_p^2 f|(\mathbb{R}^d). \quad (3.7)$$

Then, for every $r > 0$ *one has*

$$|D_p^2 f|(B_r) = |D_p^2 f^{\mathrm{rad}}|(B_r), \ \ |D_p^2 f|(\partial B_r) = |D_p^2 f^{\mathrm{rad}}|(\partial B_r)$$
$$\mathit{and} \ |D_p^2 f|(\mathbb{R}^d \setminus \bar{B}_r) = |D_p^2 f^{\mathrm{rad}}|(\mathbb{R}^d \setminus \bar{B}_r). \quad (3.8)$$

Proof. First notice that thanks to Lemma 1.10, for any $\varepsilon > 0$,

$$|D_p^2 f^{\mathrm{rad}}|(B_r) \leqq |D_p^2 f|(B_r), \ |D_p^2 f^{\mathrm{rad}}|(B_{r+\varepsilon} \setminus \bar{B}_{r-\varepsilon}) \leqq |D_p^2 f|(B_{r+\varepsilon} \setminus \bar{B}_{r-\varepsilon})$$
$$\text{and } |D_p^2 f^{\mathrm{rad}}|(\mathbb{R}^d \setminus \bar{B}_r) \leqq |D_p^2 f|(\mathbb{R}^d \setminus \bar{B}_r)$$

so that, by regularity of measures, letting $\varepsilon \searrow 0$,

$$|D_p^2 f^{\mathrm{rad}}|(B_r) \leqq |D_p^2 f|(B_r), \qquad |D_p^2 f^{\mathrm{rad}}|(\partial B_r) \leqq |D_p^2 f|(\partial B_r)$$
$$\text{and } |D_p^2 f^{\mathrm{rad}}|(\mathbb{R}^d \setminus \bar{B}_r) \leqq |D_p^2 f|(\mathbb{R}^d \setminus \bar{B}_r).$$

Then we can compute, by the inequalities above and exploiting (3.7),

$$|D_p^2 f|(\mathbb{R}^d) = |D_p^2 f^{\mathrm{rad}}|(\mathbb{R}^d) = |D_p^2 f^{\mathrm{rad}}|(B_r) + |D_p^2 f^{\mathrm{rad}}|(\partial B_r) + |D_p^2 f^{\mathrm{rad}}|(\mathbb{R}^d \setminus \bar{B}_r)$$
$$\leqq |D_p^2 f|(B_r) + |D_p^2 f|(\partial B_r) + |D_p^2 f|(\mathbb{R}^d \setminus \bar{B}_r) = |D_p^2 f|(\mathbb{R}^d),$$

so that equality holds throughout and therefore we obtain (3.8). $\square$

Now we state and prove the main results of this section, splitting the case $p = 1$ and the case $p \in (1, \infty)$. Recall that $|D_1^2 f^{\mathrm{cone}}|(\mathbb{R}^d \setminus \bar{B}_1) = 0$ according to (3.5).

Proposition 3.5. *Let $f \in L_{\mathrm{loc}}^1(\mathbb{R}^d)$ with bounded Hessian–Schatten variation and assume that*

$$f^{\mathrm{rad}} = f^{\mathrm{cone}} \quad \text{and} \quad |D_1^2 f|(\mathbb{R}^d) = |D_1^2 f^{\mathrm{cone}}|(\mathbb{R}^d). \tag{3.9}$$

Then f is equal to f^{cone} up to a linear term: there exists $\alpha \in \mathbb{R}^d$ such that

$$f(x) = f^{\mathrm{cone}}(x) + \alpha \cdot x \quad \text{for a.e.} x \in \mathbb{R}^d.$$

Proof. Let $r > 0$ and let $U \in SO(\mathbb{R}^d)$. By Lemma 1.10, $f_U := f(U \cdot)$ has finite Hessian–Schatten total variation. Also, for any radial function $g \in C_c^\infty(\mathbb{R}^d)$ one has

$$\int_{\mathbb{R}^d} f_U \Delta g \, d\mathscr{L}^d = \int_{\mathbb{R}^d} f(\Delta g)_{U^t} \, d\mathscr{L}^d = \int_{\mathbb{R}^d} f \Delta g \, d\mathscr{L}^d,$$

so that, integrating both sides with respect to $d\mu_d(U)$ and using Fubini's Theorem,

$$\int_{\mathbb{R}^d} f^{\mathrm{rad}} \Delta g \, d\mathscr{L}^d = \int_{\mathbb{R}^d} f \Delta g \, d\mathscr{L}^d.$$

Then, as $f^{\mathrm{rad}} = f^{\mathrm{cone}}$ and integrating by parts,

$$\int_{\mathbb{R}^d} g \, d Tr D\nabla f^{\mathrm{cone}} = \int_{\mathbb{R}^d} g \, d Tr D\nabla f.$$

Therefore, by an approximation argument, recalling the explicit computation (3.6), we obtain that

$$\int_{B_r} d Tr D\nabla f = -d\omega_d r^{d-1} \chi_{(0,1]}(r) \quad \forall r > 0.$$

In particular, taking into account (3.2) and (3.8) (recall that we can use (3.8) thanks to the standing assumption (3.9))

$$-\mathrm{TrD}\nabla f(B_1) = d\omega_d = |\mathrm{D}_1^2 f^{\mathrm{cone}}|(B_1) = |\mathrm{D}_1^2 f|(B_1).$$

Now Lemma 3.3 can be applied, to obtain that the function f has a continuous and concave representative in B_1 that, without loss of generality, we still denote by f. By (3.8) again, f is affine on $\mathbb{R}^d \backslash \bar{B}_1$, say $f(x) = \alpha \cdot x + \beta$ for $x \in \mathbb{R}^d \backslash \bar{B}_1$, for some $\alpha \in \mathbb{R}^d$ and $\beta \in \mathbb{R}$. Now $f^{\mathrm{rad}} = f^{\mathrm{cone}}$ forces $\beta = 0$.

Setting also $\tilde{f}(x) := f(x) - \alpha \cdot x$, we conclude the proof by showing $\tilde{f} = f^{\mathrm{cone}}$. Notice that still $\tilde{f}$ is continuous and concave on B_1 and $\tilde{f}^{\mathrm{rad}} = f^{\mathrm{cone}}$. Notice that this last fact implies $\tilde{f}(0) = 1$.

Now, for any $\sigma \in \partial B_1$, define $\tilde{f}_\sigma(s) := \tilde{f}(s\sigma)$ for $s \in [0, \infty)$, a function continuous and concave in $[0, 1)$ with $\tilde{f}_\sigma(0) = 1$. Notice that for $\mathscr{H}^{d-1}$-a.e. $\sigma \in \partial B_1$, $\tilde{f}_\sigma \in W^{1,1}_{\mathrm{loc}}((0, \infty))$. This can be seen either with a change of coordinates and the characterization of Sobolev functions on lines or by approximation, using repeatedly integration in polar coordinates. Hence, for $\mathscr{H}^{d-1}$-a.e. $\sigma \in \partial B_1$, the function $\tilde{f}_\sigma$ has a continuous representative in $[1, \infty)$. Now, for $\mathscr{H}^{d-1}$-a.e. $\sigma \in \partial B_1$, $\tilde{f}_\sigma$ vanishes a.e. in $(1, \infty)$ (as $\tilde{f}$ vanishes identically on $\mathbb{R}^d \setminus \bar{B}_1$), therefore this implies $\tilde{f}_\sigma(s) \to 0$ as $s \uparrow 1$ and the continuous representative is the one null in $[1, \infty)$. Then, exploiting continuity and concavity, for $\mathscr{H}^{d-1}$-a.e. $\sigma \in \partial B_1$, $\tilde{f}_\sigma(s) \geqq (1 - s)$ for $s \in [0, 1]$. Then it holds that $\tilde{f} \geqq f^{\mathrm{cone}}$ $\mathscr{L}^d$-a.e. on B_1, whence, being $\tilde{f}^{\mathrm{rad}} = f^{\mathrm{cone}}$, $\tilde{f} = f^{\mathrm{cone}}$ on B_1. □

Proposition 3.6. *Let* $p \in [1, \infty)$. *Let* $f \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$ *with bounded Hessian–Schatten variation and assume that*

$$f^{\mathrm{rad}} = f^{\mathrm{cone}} \quad and \quad |\mathrm{D}_p^2 f|(\mathbb{R}^d) = |\mathrm{D}_p^2 f^{\mathrm{cone}}|(\mathbb{R}^d). \tag{3.10}$$

Then f is equal to f^{cone} up to a linear term: there exists $\alpha \in \mathbb{R}^d$ such that

$$f(x) = f^{\mathrm{cone}}(x) + \alpha \cdot x \quad for\ a.e.x \in \mathbb{R}^d.$$

Proof. We focus on the case $p > 1$ as the case $p = 1$ has already been proved in Proposition 3.5. Let now $g := \frac{1}{2}(f + f^{\mathrm{cone}})$. Recalling (3.8), $|\mathrm{D}_p^2 g|(\mathbb{R}^d \backslash \bar{B}_1) = 0$. Still, $g^{\mathrm{rad}} = f^{\mathrm{cone}}$, so that, by Lemma 1.10 and (3.10),

$$|\mathrm{D}_p^2 f^{\mathrm{cone}}|(\mathbb{R}^d) \leqq |\mathrm{D}_p^2 g|(\mathbb{R}^d) \leqq \frac{1}{2}|\mathrm{D}_p^2 f|(\mathbb{R}^d) + \frac{1}{2}|\mathrm{D}_p^2 f^{\mathrm{cone}}|(\mathbb{R}^d) = |\mathrm{D}_p^2 f^{\mathrm{cone}}|(\mathbb{R}^d),$$

hence equality holds throughout and therefore g satisfies (3.10) in place of f.

We next decompose $\mathrm{D}\nabla f$ in absolutely continuous and singular part, use that the singular one has a rank one density with respect to the total variation, and show that the absolutely continuous one is proportional to the one of $\mathrm{D}\nabla f^{\mathrm{cone}}$. We are going to use the theory of functions of bounded variation throughout, see e.g. [3]. The superscript s denotes the singular part of a measure with respect to $\mathscr{L}^d$. We have a $\mathscr{L}^d$-negligible Borel set $N \subseteq B_1$ such that $|\mathrm{D}_1^2 f|^s \llcorner B_1 = |\mathrm{D}_1^2 f| \llcorner N$. Also $|\mathrm{D}_1^2 g|^s \llcorner B_1 = |\mathrm{D}_1^2 g| \llcorner N$, being $|\mathrm{D}_1^2 f^{\mathrm{cone}}| \llcorner B_1 \ll \mathscr{L}^d$, by (3.5). In addition

$$|D_p^2 g| \llcorner N \leqq \frac{1}{2}|D_p^2 f| \llcorner N + \frac{1}{2}|D_p^2 f^{\text{cone}}| \llcorner N = \frac{1}{2}|D_p^2 f| \llcorner N$$

$$\leqq |D_p^2 g| \llcorner N + \frac{1}{2}|D_p^2 f^{\text{cone}}| \llcorner N = |D_p^2 g| \llcorner N$$

hence equality holds throughout and in particular, $|D_p^2 g| \llcorner N = \frac{1}{2}|D_p^2 f| \llcorner N$. Now, recall that $|D_p^2 f| \llcorner (B_1 \setminus N) \ll \mathcal{L}^d$ and $|D_p^2 g| \llcorner (B_1 \setminus N) \ll \mathcal{L}^d$, also $|D_p^2 f^{\text{cone}}| \llcorner B_1 \ll \mathcal{L}^d$, by (3.5). Therefore, by Proposition 1.7,

$$|D_p^2 g|(B_1) = |D_p^2 g|(N) + \int_{B_1 \setminus N} \left| \frac{dD\nabla g}{d|D\nabla g|} \right|_p d|D\nabla g|$$

$$= |D_p^2 g|(N) + \int_{B_1 \setminus N} \left| \frac{dD\nabla g}{d\mathcal{L}^d} \right|_p d\mathcal{L}^d$$

$$= |D_p^2 g|(N) + \frac{1}{2} \int_{B_1 \setminus N} \left| \frac{dD\nabla f}{d\mathcal{L}^d} + \frac{dD\nabla f^{\text{cone}}}{d\mathcal{L}^d} \right|_p d\mathcal{L}^d$$

$$\leqq \frac{1}{2}|D_p^2 f|(N) + \frac{1}{2} \int_{B_1 \setminus N} \left| \frac{dD\nabla f}{d\mathcal{L}^d} \right|_p + \left| \frac{dD\nabla f^{\text{cone}}}{d\mathcal{L}^d} \right|_p d\mathcal{L}^d$$

$$\leqq \frac{1}{2}|D_p^2 f|(B_1) + \frac{1}{2}|D_p^2 f^{\text{cone}}|(B_1) = |D_p^2 g|(B_1),$$

where we also used (3.10) for f and g and (3.8) in the last equality. Hence equality holds throughout, so that

$$\left| \frac{dD\nabla f}{d\mathcal{L}^d} + \frac{dD\nabla f^{\text{cone}}}{d\mathcal{L}^d} \right|_p = \left| \frac{dD\nabla f}{d\mathcal{L}^d} \right|_p + \left| \frac{dD\nabla f^{\text{cone}}}{d\mathcal{L}^d} \right|_p \qquad \mathcal{L}^d\text{-a.e. on } B_1.$$

By strict convexity of the p-Schatten norm (item vi) of Proposition 1.2), and the fact (by (3.5)) that the density of $D\nabla f^{\text{cone}}$ with respect to $\mathcal{L}^d$ is nonzero $\mathcal{L}^d$-a.e. on B_1, we have that for some Borel map $t : B_1 \to [0, \infty)$,

$$\frac{dD\nabla f}{d\mathcal{L}^d} = t \frac{dD\nabla f^{\text{cone}}}{d\mathcal{L}^d} \qquad \mathcal{L}^d\text{-a.e. on } B_1. \tag{3.11}$$

Now, by (3.5), for $q \in [1, \infty]$,

$$\left| \frac{dD\nabla f^{\text{cone}}}{d\mathcal{L}^d}(x) \right|_q = \left| -\frac{|x|^2 \text{Id} - x \otimes x}{|x|^3} \right|_q = \frac{(d-1)^{1/q}}{|x|} \qquad \mathcal{L}^d\text{-a.e. on } B_1. \tag{3.12}$$

Then, by (3.11) and (3.12) (with $q = 1, p$),

$$\left| \frac{dD\nabla f}{d\mathcal{L}^d}(x) \right|_p = t(x)\frac{(d-1)^{1/p}}{|x|} = (d-1)^{1/p-1}t(x)\frac{d-1}{|x|}$$

$$= (d-1)^{1/p-1}\left| \frac{dD\nabla f}{d\mathcal{L}^d}(x) \right|_1 \qquad \mathcal{L}^d\text{-a.e. on } B_1.$$

Therefore, by Proposition 1.7,

$$|\mathrm{D}_p^2 f|(B_1 \setminus N) = (d-1)^{1/p-1}|\mathrm{D}_1^2 f|(B_1 \setminus N). \tag{3.13}$$

On the singular set N, by Proposition 1.7 and Alberti's rank 1 Theorem together with item v) of Proposition 1.2,

$$|\mathrm{D}_p^2 f|(N) = \int_N \left|\frac{\mathrm{d}\mathrm{D}\nabla f}{\mathrm{d}|\mathrm{D}\nabla f|}\right|_p \mathrm{d}|\mathrm{D}\nabla f| = \int_N \left|\frac{\mathrm{d}\mathrm{D}\nabla f}{\mathrm{d}|\mathrm{D}\nabla f|}\right|_1 \mathrm{d}|\mathrm{D}\nabla f| = |\mathrm{D}_1^2 f|(N). \tag{3.14}$$

Therefore, by (3.13), (3.14) and (3.8), taking into account that $d \geqq 2$ and $p \geqq 1$ (hence $1 \leqq (d-1)^{1-1/p}$),

$$\begin{aligned}
|\mathrm{D}_1^2 f|(B_1) &= |\mathrm{D}_1^2 f|(B_1 \setminus N) + |\mathrm{D}_1^2 f|(N) \\
&= (d-1)^{1-1/p}|\mathrm{D}_p^2 f|(B_1 \setminus N) + |\mathrm{D}_p^2 f|(N) \\
&\leqq (d-1)^{1-1/p}\big(|\mathrm{D}_p^2 f|(B_1 \setminus N) + |\mathrm{D}_p^2 f|(N)\big) \tag{3.15} \\
&= (d-1)^{1-1/p}|\mathrm{D}_p^2 f|(B_1) \\
&= (d-1)^{1-1/p}|\mathrm{D}_p^2 f^{\mathrm{cone}}|(B_1) = |\mathrm{D}_1^2 f^{\mathrm{cone}}|(B_1)
\end{aligned}$$

where the last equality follows from (3.2). Recalling (3.8) and arguing exactly as for (3.14) for the first and third equalities,

$$|\mathrm{D}_1^2 f|(\partial B_1) = |\mathrm{D}_p^2 f|(\partial B_1) = |\mathrm{D}_p^2 f^{\mathrm{cone}}|(\partial B_1) = |\mathrm{D}_1^2 f^{\mathrm{cone}}|(\partial B_1). \tag{3.16}$$

Then, by (3.8), exploiting (3.15) and (3.16)

$$\begin{aligned}
|\mathrm{D}_1^2 f|(\mathbb{R}^d) &= |\mathrm{D}_1^2 f|(B_1) + |\mathrm{D}_1^2 f|(\partial B_1) \leqq |\mathrm{D}_1^2 f^{\mathrm{cone}}|(B_1) \\
&+ |\mathrm{D}_1^2 f^{\mathrm{cone}}|(\partial B_1) = |\mathrm{D}_1^2 f^{\mathrm{cone}}|(\mathbb{R}^d).
\end{aligned}$$

Recalling Lemma 1.10 together with (3.10), the inequality above yields that f satisfies (3.9), so that the conclusion follows from Proposition 3.5. □

3.3. Proof of the main result

Proof of Theorem 3.1. Let f_1 and f_2 be as in the statement and recall (3.3), so that we can define f_i^{rad} for $i = 1, 2$. As f^{cone} is already a radial function, we still have $\lambda f_1^{\mathrm{rad}} + (1-\lambda) f_2^{\mathrm{rad}} = f^{\mathrm{cone}}$. Now we compute, using Lemma 1.10 and the assumption,

$$\begin{aligned}
|\mathrm{D}_p^2 f^{\mathrm{cone}}|(\mathbb{R}^d) &= |\mathrm{D}_p^2 (\lambda f_1^{\mathrm{rad}} + (1-\lambda) f_2^{\mathrm{rad}})|(\mathbb{R}^d) \\
&\leqq \lambda|\mathrm{D}_p^2 f_1^{\mathrm{rad}}|(\mathbb{R}^d) + (1-\lambda)|\mathrm{D}_p^2 f_2^{\mathrm{rad}}|(\mathbb{R}^d) \\
&\leqq \lambda|\mathrm{D}_p^2 f_1|(\mathbb{R}^d) + (1-\lambda)|\mathrm{D}_p^2 f_2|(\mathbb{R}^d) \\
&= \lambda|\mathrm{D}_p^2 f^{\mathrm{cone}}|(\mathbb{R}^d) + (1-\lambda)|\mathrm{D}_p^2 f^{\mathrm{cone}}|(\mathbb{R}^d) \\
&= |\mathrm{D}_p^2 f^{\mathrm{cone}}|(\mathbb{R}^d),
\end{aligned}$$

hence equality holds throughout. Therefore,

$$|D_p^2 f_i^{\mathrm{rad}}|(\mathbb{R}^d) = |D_p^2 f_i|(\mathbb{R}^d) \qquad \text{for } i = 1, 2,$$

and

$$|D_p^2(\lambda f_1^{\mathrm{rad}} + (1-\lambda) f_2^{\mathrm{rad}})|(\mathbb{R}^d) = |D_p^2(\lambda f_1^{\mathrm{rad}})|(\mathbb{R}^d) + |D_p^2((1-\lambda) f_2^{\mathrm{rad}})|(\mathbb{R}^d)$$

so that, by Lemma 1.12,

$$|D_p^2 f^{\mathrm{cone}}| = \lambda |D_p^2 f_1^{\mathrm{rad}}| + (1-\lambda)|D_p^2 f_2^{\mathrm{rad}}| \qquad (3.17)$$

as measures on $\mathbb{R}^d$. As f_1^{rad} and f_2^{rad} are radial functions with bounded Hessian–Schatten variation, by Proposition 1.13, $f_i^{\mathrm{rad}}(x) = g_i(|x|)$ for $g_i \in W_{\mathrm{loc}}^{1,1}((0,\infty))$. Similarly, $f^{\mathrm{cone}}(x) = g^{\mathrm{cone}}(|x|) = (1 - |x|)_+$, notice that $\lambda g_1 + (1-\lambda)g_2 = g^{\mathrm{cone}}$. Then, using repeatedly the representation formula of Proposition 1.13 and (3.17),

$$
\begin{aligned}
|D_p^2 f^{\mathrm{cone}}|(B_1) &= d\omega_d \int_0^1 \|(0, g'_{\mathrm{cone}}, \dots, g'_{\mathrm{cone}})\|_{\ell^p} s^{d-2} ds \\
&\leqq d\omega_d \left(\lambda \int_0^1 \|(0, g'_1, \dots, g'_1)\|_{\ell^p} s^{d-2} ds \right. \\
&\quad \left. + (1-\lambda) \int_0^1 \|(0, g'_2, \dots, g'_2)\|_{\ell^p} s^{d-2} ds \right) \\
&\leqq \lambda |D_p^2 f_1^{\mathrm{rad}}|(B_1) + (1-\lambda)|D_p^2 f_2^{\mathrm{rad}}|(B_1) = |D_p^2 f^{\mathrm{cone}}|(B_1),
\end{aligned}
$$

hence equality holds throughout. In particular, as we have obtained

$$d\omega_d \int_0^1 \|(0, g'_i, \dots, g'_i)\|_{\ell^p} s^{d-2} ds = |D_p^2 f_i^{\mathrm{rad}}|(B_1) \qquad \text{for } i = 1, \ 2,$$

exploiting the representation formula of Proposition 1.13, we have that g'_1 and g'_2 are constant on $(0, 1)$. Also, by (3.17), and the representation formula of Proposition 1.13 again, g'_1 and g'_2 vanish identically on $(1, \infty)$. Recall also that $g_i \in W_{\mathrm{loc}}^{1,1}((0, \infty))$, so that g_i has a continuous representative, for $i = 1, 2$. Hence, there exist $\alpha_1, \alpha_2 \in \mathbb{R}$ and $\beta_1, \beta_2 \in \mathbb{R}$ such that

$$g_i(s) = \alpha_i(1 - s)_+ + \beta_i.$$

Now, $\lambda g_1 + (1-\lambda)g_2 = g^{\mathrm{cone}}$ forces $\lambda \alpha_1 + (1-\lambda)\alpha_2 = 1$, whereas

$$|\alpha_i||D_p^2 f^{\mathrm{cone}}|(\mathbb{R}^d) = |D_p^2 f_i^{\mathrm{rad}}|(\mathbb{R}^d) = |D_p^2 f^{\mathrm{cone}}|(\mathbb{R}^d) \qquad \text{for } i = 1, 2$$

forces $|\alpha_1| = |\alpha_2| = 1$. Hence, $\alpha_1 = \alpha_2 = 1$.

Therefore, to sum up, we have, for $i = 1, 2$,

$$f_i^{\mathrm{rad}} = f^{\mathrm{cone}} + \beta_i,$$

so that

$$|D_p^2 f_i^{\mathrm{rad}}|(\mathbb{R}^d) = |D_p^2 f^{\mathrm{cone}}|(\mathbb{R}^d) = |D_p^2 f_i|(\mathbb{R}^d).$$

Notice that $f_i^{\mathrm{rad}} - \beta_i = (f_i - \beta_i)^{\mathrm{rad}}$. Now we use Proposition 3.6 to infer that

$$f_i(x) - \beta_i = f^{\mathrm{cone}}(x) + a_i \cdot x \qquad \text{for a.e. } x \in \mathbb{R}^d,$$

hence the proof is concluded with $L_i(x) := \alpha_i \cdot x + \beta_i$. $\quad\square$

4. Solutions of the Minimization Problem

In this section we stick to the two dimensional case $d = 2$. Recall that, by Proposition 1.11, functions with bounded Hessian–Schatten variation are continuous, as we are in dimension 2 and hence the evaluation functionals in (4.1) below are meaningful (we will implicitly take the continuous representative, whenever it is possible).

Fix $\Omega \subseteq \mathbb{R}^2$ open, and fix $x_1, \ldots, x_N \in \Omega$ distinct test points and fix also $y_1, \ldots, y_N \in \mathbb{R}$. For $\lambda \in [0, \infty]$ and $p, q \in [1, \infty]$ we consider the functional

$$\mathcal{F}_\lambda^{p,q} : L^1_{\text{loc}}(\Omega) \to [0, \infty] \qquad \text{defined as}$$
$$\mathcal{F}_\lambda^{p,q}(f) := |\mathrm{D}^2_p f|(\Omega) + \lambda \|(f(x_i) - y_i)_{i=1,\ldots,N}\|_{\ell^q}, \tag{4.1}$$

where we adopt the convention that $\infty \cdot 0 = 0$. Notice that if $p = q = 1$, we have that $\mathcal{F}_\lambda^{1,1} = \mathcal{F}_\lambda$, where $\mathcal{F}_\lambda$ is defined in (0.4) in the Introduction.

Our aim is to establish conditions under which $\mathcal{F}_\lambda^{p,q}$ has minimizers, i.e. we want to ensure the existence of a minimizer of

$$\inf_{f \in L^1_{\text{loc}}(\Omega)} \mathcal{F}_\lambda^{p,q}(f).$$

It turns out that for many values of λ, p, q, minimizers indeed exist. Here we state our main results in this direction.

Theorem 4.1. *Let $p, q \in [1, \infty]$ and let $\lambda \in [0, 2^{1/p-1}4\pi]$. Then there exists a minimizer of $\mathcal{F}_\lambda^{p,q}$.*

Theorem 4.2. *Let $\lambda \in [0, \infty]$. Then there exists a minimizer of $\mathcal{F}_\lambda^{1,1}$.*

Theorems 4.1 and 4.2 will follow easily from the results of Section 4.1. We defer their proof of to Section 4.2.

4.1. Auxiliary results

For the next lemma, we recall again that functions with bounded Hessian–Schatten variation in dimension 2 are automatically continuous. Hence, the evaluation (at 0) functional in the infimum above is meaningful. The spirit of this lemma is to provide us with "bump" functions whose Hessian–Schatten total variation is almost optimal.

Lemma 4.3. *Let $p \in [1, \infty]$. Then it holds that*

$$\inf \left\{ |\mathrm{D}^2_p f|(\mathbb{R}^2) : f \in L^1_{\text{loc}}(\mathbb{R}^2) \text{ with compact support and } f(0) = 1 \right\} = 2^{1+1/p}\pi.$$
$$\tag{4.2}$$

In particular, thanks to (3.2), the infimum is attained by the cut cone $x \mapsto (1-|x|)^+$ when $p = 1$.

Proof. For $\varepsilon \in (0, 1)$, define $f_\varepsilon(x) = (1 - |x|^\varepsilon) \vee 0$. By Proposition 1.13,

$$|D_p^2 f_\varepsilon|(\mathbb{R}^2) = 2\pi \left(\int_0^1 s^{\varepsilon-1} \|(\varepsilon(\varepsilon - 1), \varepsilon)\|_{\ell^p} ds + \varepsilon \right) \to 2^{1+1/p}\pi \quad \text{as } \varepsilon \searrow 0,$$

so that we have $\leqq$ in (4.2).

We prove now the opposite inequality in (4.2). Take then $f \in L^1_{loc}(\mathbb{R}^2)$, compactly supported, with bounded Hessian–Schatten variation and such that $f(0) = 1$. We have to prove that $|D_p^2 f|(\mathbb{R}^2) \geqq 2^{1+1/p}\pi$. Using Lemma 1.9, Lemma 1.10, we see that we can assume with no loss of generality that $f \in C_c^\infty(\mathbb{R}^2)$ and f is radial, say $f(x) = g(|x|)$, with $g(0) = 1$ and $g'_+(0) = 0$. Now, by Proposition 1.7 and the inequality $(|a| + |b|) \leqq 2^{1-1/p}(|a|^p + |b|^p)^{1/p}$, we obtain that

$$|D_p^2 f|(\mathbb{R}^2) \geqq 2^{1/p-1}|D_1^2 f|(\mathbb{R}^2).$$

Hence, it is enough to show the claim in the case $p = 1$, i.e. we have to show that $|D_1^2 f|(\mathbb{R}^2) \geqq 4\pi$. We compute now that

$$\int_0^\infty s|g''|ds \geqq \int_0^\infty sg''ds = -\int_0^\infty g'ds = 1$$

$$\text{and} \quad \int_0^\infty |g'|ds \geqq -\int_0^\infty g'ds = 1,$$

so that, by by Proposition 1.13,

$$|D_1^2 f|(\mathbb{R}^2) = 2\pi \int_0^\infty s|g''| + |g'|ds \geqq 4\pi.$$

$\square$

The existence of "good bump functions" granted by Lemma 4.3 allows us to prove, in Proposition 4.4 below, that for λ large enough the infimum of $\mathcal{F}_\lambda^{p,q}$ does not depend on λ, namely that minimizing $\mathcal{F}_\lambda^{p,q}$ asymptotically promotes the perfect fit with the data.

Proposition 4.4. *Let $p, q \in [1, \infty]$ and let $\lambda \in [2\pi 2^{1/p} N^{1-1/q}, \infty]$. Then*

$$\inf_{f \in L^1_{loc}(\Omega)} \mathcal{F}_\lambda^{p,q}(f) = \inf_{f \in L^1_{loc}(\Omega)} \mathcal{F}_\infty^{p,q}(f).$$

In particular, in this range of λ, the infima are also independent of q.

Proof. We let $r \in (0, \infty)$ small enough so that $\text{dist}(x_i, x_j) > 3r$ if $i \neq j$. Let $\varepsilon \in (0, 1)$. For $i = 1, \ldots, N$, by Lemma 4.3 and a scaling argument, we take $g_i \in C_c(\mathbb{R}^2)$ with $g(x_i) = 1$, $\text{supp } g_i \subseteq B_r(x_i)$ and $|D_p^2 g_i|(\mathbb{R}^2) \leqq 2^{1+1/p}\pi + \varepsilon$.

Then we consider $f \in L^1_{loc}(\Omega)$ and we set

$$\tilde{f} := f - \sum_i (f(x_i) - y_i)g_i. \tag{4.3}$$

Notice $\tilde{f}(x_i) = y_i$ for every $i = 1, \ldots, N$ and that

$$|D_p^2 \tilde{f}|(\Omega) \lesseqgtr |D_p^2 f|(\Omega) + (2^{1+1/p}\pi + \varepsilon) \sum_{i=1}^{N} |f(x_i) - y_i|$$

$$= |D_p^2 f|(\Omega) + (2^{1+1/p}\pi + \varepsilon)\|(f(x_i) - y_i)_i\|_{\ell^1}$$

$$\lesseqgtr |D_p^2 f|(\Omega) + (2^{1+1/p}\pi + \varepsilon)N^{1-1/q}\|(f(x_i) - y_i)_i\|_{\ell^q}.$$

Therefore, being $\varepsilon \in (0, 1)$ arbitrary and $f \in L^1_{\mathrm{loc}}(\Omega)$ arbitrary, we have that

$$\inf_{f \in L^1_{\mathrm{loc}}(\Omega)} \mathcal{F}^{p,q}_\infty(f) \leq \inf_{f \in L^1_{\mathrm{loc}}(\Omega)} \mathcal{F}^{p,q}_\lambda(f) \quad \text{whenever } \lambda \geq 2\pi 2^{1/p}N^{1-1/q}.$$

As also $\mathcal{F}^{p,q}_\infty(\tilde{f}) \geq \mathcal{F}^{p,q}_\lambda(f)$, we have proved the claim, thanks to our choice of λ.
□

The following lemma estimates how much the evaluation functional at x differs from the average functional on $B_r(x)$, hence allows us to quantify the error we make replacing the evaluation functional with another functional that has the advantage of being continuous with respect to weaker notion of convergence.

Lemma 4.5. Let $f \in L^1_{\mathrm{loc}}(\Omega)$ with bounded Hessian–Schatten variation in Ω. Let also $B = B_r(x) \subseteq \Omega$ such that $2B := B_{2r}(x) \subseteq \Omega$. Then, if $p \in [1, \infty]$,

$$\left| f(x) - \fint_B f \right| \lesseqgtr 2^{1-1/p} \left(\frac{1}{4\pi}|D_p^2 f|(B) + \frac{1}{2\pi}|D_p^2 f|(2B \setminus B) \right). \quad (4.4)$$

Proof. We can assume with no loss of generality that $x = 0$. By approximation of r from below, we can also assume that $|D_1^2 f|(\partial B) = 0$. Hence, using Proposition 1.8 and Lemma 1.10, we can assume in addition that f is radial and $f \in C^\infty(2B)$, say $f(\cdot) = g(|\cdot|)$. Notice that $g'_+(0) = 0$. We then compute

$$f(0) - \fint_B f = g(0) - \frac{2}{r^2}\int_0^r sg(s)\mathrm{d}s = \frac{2}{r^2}\int_0^r s(g(0) - g(s))\mathrm{d}s$$

$$= -\frac{2}{r^2}\int_0^r s\int_0^s g'(\tau)\mathrm{d}\tau\mathrm{d}s = -\frac{2}{r^2}\int_0^r g'(\tau)\left(\int_\tau^r s\mathrm{d}s\right)\mathrm{d}\tau,$$

so that

$$\left| f(0) - \fint_B f \right| \leq \frac{2}{r^2}\int_0^r |g'|r^2/2\mathrm{d}s = \int_0^r |g'|\mathrm{d}s. \quad (4.5)$$

We stick for the moment to the case $p = 1$. We use Proposition 1.13 to compute

$$|D_1^2 f|(2B \setminus B) = 2\pi \int_r^{2r} s|g''| + |g'|\mathrm{d}s,$$

$$|D_1^2 f|(B) = 2\pi \int_0^r s|g''| + |g'|\mathrm{d}s \quad (4.6)$$

and we take $\xi \in (r, 2r)$ such that

$$r|g'|(\xi) \leqq \int_r^{2r} |g'| \mathrm{d}s. \tag{4.7}$$

Now we write $\{g' > 0\} \cap (0, \xi) = \bigcup_k I_k$ and $\{g' < 0\} \cap (0, \xi) = \bigcup_k J_k$, where I_k and J_k are countably many pairwise disjoint open intervals. Notice that if $p \in \partial I_k$ for some k, then either $p = \xi$ or $g'(p) = 0$. Then, if we take I_k such that $\xi \in \partial I_k$,

$$\int_{I_k} s|g''| \mathrm{d}s \geqq -\int_{I_k} sg'' \mathrm{d}s = \int_{I_k} g' \mathrm{d}s - \xi g'(\xi) = \int_{I_k} |g'| \mathrm{d}s - \xi|g'|(\xi),$$

whereas if we take I_k such that $\xi \notin \partial I_k$,

$$\int_{I_k} s|g''| \mathrm{d}s \geqq -\int_{I_k} sg'' \mathrm{d}s = \int_{I_k} g' \mathrm{d}s = \int_{I_k} |g'| \mathrm{d}s.$$

Similar inequalities hold in the case of an interval of the type J_k. Therefore, summing over all intervals I_k and J_k,

$$\int_0^{2r} s|g''| \mathrm{d}s \geq \int_0^\xi s|g''| \mathrm{d}s \geq \int_0^\xi |g'| \mathrm{d}s - \xi|g'|(\xi),$$

so that, by the choice of ξ due to (4.7),

$$\int_0^r |g'| \mathrm{d}s \leqq \int_0^\xi |g'| \mathrm{d}s \leqq \int_0^{2r} s|g''| \mathrm{d}s + \xi|g'|(\xi) \leqq \int_0^{2r} s|g''| \mathrm{d}s + 2\int_r^{2r} |g'| \mathrm{d}s.$$

Then, using also (4.5) and (4.6),

$$2\left|f(0) - \fint_B f\right| \leqq 2\int_0^r |g'| \mathrm{d}s \leqq \int_0^r |g'| \mathrm{d}s + \int_0^r s|g''| \mathrm{d}s$$
$$+ \int_r^{2r} s|g''| \mathrm{d}s + 2\int_r^{2r} |g'| \mathrm{d}s$$
$$\leqq \frac{1}{2\pi}|D_1^2 f|(B) + \frac{1}{\pi}|D_1^2 f|(2B \setminus B),$$

whence the claim for $p = 1$. For the general case, simply notice that $|D_1^2 f|(B) \leqq 2^{1-1/p}|D_p^2 f|(B)$ and the same holds for $2B \setminus B$, by $\ell_1 - \ell_p$ inequality and Proposition 1.7. □

Remark 4.6. Notice that the constant $1/(4\pi)$ in front of $|D_p^2 f|(B)$ in (4.4) is somehow optimal. We can realize this considering the sequence of functions f_ε used to prove Lemma 4.3. ∎

By Lemma 4.5, there is no surprise in knowing that, given a weakly convergent sequence $f_k \rightharpoonup f$, in duality with the space $L_c^\infty(\Omega)$ of L^∞ function with compact (essential) support, we can estimate how much the evaluation functional fails to converge in terms of concentration of Hessian–Schatten total variation at x.

Lemma 4.7. *Let $f \in L^1_{\text{loc}}(\Omega)$ and let $(f_k) \subseteq L^1_{\text{loc}}(\Omega)$ such that $f_k \rightharpoonup f$ in duality with $L^\infty_c(\Omega)$ with $\sup_k |D^2_p f_k|(A) < \infty$ for any open set $A \Subset \Omega$. Then, f has locally bounded Hessian–Schatten variation in Ω and for any $x \in \Omega$ one has*

$$\limsup_k |f(x) - f_k(x)| \leqq \frac{2^{1-1/p}}{4\pi} \lim_{r \searrow 0} \limsup_k |D^2_p f_k|(B_r(x)). \qquad (4.8)$$

Proof. First, take a non relabelled subsequence so that $\lim_k |f(x) - f_k(x)|$ exists and equals the $\limsup_k$ at the left hand side of (4.8).

We assume that there exists $r_1 > 0$ small enough so that $B_{r_1}(x) \subseteq \Omega$ and moreover that $\limsup_k |D^2 f_k|(B_{r_1}(x)) < \infty$, otherwise there is nothing to show. By lower semicontinuity this implies that f has bounded Hessian–Schatten variation in $B_{r_1}(x)$. We extract a further non relabelled subsequence such that, for some finite measure μ on $B_{r_1}(x)$, $|D^2_p f_k| \rightharpoonup \mu$ in duality with $C_c(B_{r_1}(x))$.

Let now $r \in (0, r_1/2)$. Then,

$$\left| f(x) - f_k(x) \right| \leqq \left| f(x) - \fint_{B_r(x)} f \right| + \left| \fint_{B_r(x)} f - \fint_{B_r(x)} f_k \right| + \left| f_k(x) - \fint_{B_r(x)} f_k \right|.$$

Now notice that by continuity of f the first summand converges to 0 as $r \searrow 0$, whereas, by the convergence assumption the second summand converges to 0 as $k \to \infty$. Also, by Lemma 4.5, we bound the third summand as follows

$$\left| f_k(x) - \fint_{B_r(x)} f_k \right| \leqq 2^{1-1/p} \left(\frac{1}{4\pi} |D^2_p f_k|(B_r(x)) + \frac{1}{2\pi} |D^2_p f_k|(B_{2r}(x) \setminus B_r(x)) \right).$$

To conclude, it is enough notice that that

$$\limsup_{r \searrow 0} \limsup_k |D^2_p f_k|(B_{2r}(x) \setminus B_r(x)) \leqq \lim_{r \searrow 0} \mu(\bar{B}_{2r}(x) \setminus B_r(x)) = 0.$$

$\square$

By using the results above, we can prove the lower semicontinuity of $\mathcal{F}^{p,q}_\lambda$. In the case $q = 1$, notice that the argument used in the proof of Proposition 4.4 together with the next result can be used to show that $\mathcal{F}^{p,1}_\lambda$ is precisely the relaxed functional of $\mathcal{F}^{p,1}_\infty$ when $\lambda = 2^{1+1/p}\pi$.

Lemma 4.8. *Let $p, q \in [1, \infty]$ and let $\lambda \in [0, 2^{1/p-1}4\pi]$. Then $\mathcal{F}^{p,q}_\lambda$ is lower semicontinuous with respect to weak convergence in duality with $L^\infty_c(\Omega)$.*

Proof. Let $(f_k) \subseteq L^1_{\text{loc}}(\Omega)$ be such that $f_k \rightharpoonup f$ in duality with $L^\infty_c(\Omega)$, for some $f \in L^1_{\text{loc}}(\Omega)$. We have to prove that

$$\mathcal{F}^{p,q}_\lambda(f) \leqq \liminf_k \mathcal{F}^{p,q}_\lambda(f_k).$$

First, extract a non relabelled subsequence such that $\mathcal{F}^{p,q}_\lambda(f_k)$ has a limit, as $k \to \infty$, which equals the right hand side of the inequality above. Then, we can

assume that $\liminf_k |D_p^2 f_k|(\Omega) < \infty$, otherwise there is nothing to show. Hence f has bounded Hessian–Schatten variation in Ω and, up to the extraction of a non relabelled subsequence, we can assume that $|D_p^2 f_k| \rightharpoonup \mu$ in duality with $C_c(\Omega)$ for some finite measure μ on Ω. Even though μ depends on p, we do not make this dependence explicit. Also, we extract a non relabelled subsequence such that for every $i = 1, \ldots, N$, $|f(x_i) - f_k(x_i)|$ has a (finite) limit as $k \to \infty$.

Notice that for every $z \in \Omega$ one has

$$\mu(\{z\}) \leqq \lim_{r \searrow 0} \limsup_k |D_p^2 f_k|(\bar{B}_r(z)) \leqq \lim_{r \searrow 0} \mu(\bar{B}_r(z)) = \mu(\{z\}). \qquad (4.9)$$

We compute, as $|D_p^2 f|(\{z\}) = 0$ for every $z \in \Omega$,

$$\mathcal{F}_\lambda^{p,q}(f) = |D_p^2 f|(\Omega) + \lambda \|(f(x_i) - y_i)_i\|_{\ell^q}$$

$$= \lim_{r \searrow 0} |D_p^2 f|\left(\Omega \setminus \bigcup_{i=1}^N \bar{B}_r(x_i)\right) + \lambda \|(f(x_i) - y_i)_i\|_{\ell^q}. \qquad (4.10)$$

By lower semicontinuity,

$$|D_p^2 f|\left(\Omega \setminus \bigcup_{i=1}^N \bar{B}_r(x_i)\right) \leqq \liminf_k |D_p^2 f_k|\left(\Omega \setminus \bigcup_{i=1}^N \bar{B}_r(x_i)\right)$$

so that by (4.9)

$$\lim_{r \searrow 0} |D_p^2 f|\left(\Omega \setminus \bigcup_{i=1}^N \bar{B}_r(x_i)\right) \leqq \liminf_k |D_p^2 f_k|(\Omega) - \sum_{i=1}^N \mu(\{x_i\}). \qquad (4.11)$$

Also, by Lemma 4.7 and (4.9),

$$\lim_k \|(f(x_i) - f_k(x_i))_i\|_{\ell^q} \leqq \lim_k \|(f(x_i) - f_k(x_i))_i\|_{\ell^1} \leqq \frac{2^{1-1/p}}{4\pi} \sum_{i=1}^N \mu(\{x_i\}),$$

so that

$$\|(f(x_i) - y_i)_i\|_{\ell^q} \leqq \frac{2^{1-1/p}}{4\pi} \sum_i \mu(\{x_i\}) + \liminf_k \|(f_k(x_i) - y_i)_i\|_{\ell^q}. \qquad (4.12)$$

Inserting (4.11) and (4.12) into (4.10) we obtain, by the super additivity of the lim inf,

$$\mathcal{F}_\lambda^{p,q}(f) \leqq \liminf_k \mathcal{F}_\lambda^{p,q}(f_k) + \left(\lambda \frac{2^{1-1/p}}{4\pi} - 1\right) \sum_i \mu(\{x_i\}),$$

whence the claim by the choice of λ. □

Weak relative compactness of minimizing sequences for $\mathcal{F}_\lambda^{p,q}$ is obtained through a classical argument, the only (slight) technical difficulty relies in possibly irregular domains Ω.

Lemma 4.9. *Let p, $q \in [1, \infty]$ and let $\lambda \in [0, \infty]$. Then there exist a minimizing sequence (f_k) for $\mathcal{F}_\lambda^{p,q}$ and a function $f \in L^1_{\mathrm{loc}}(\Omega)$ such that $f_k \rightharpoonup f$ in duality with $L^\infty_c(\Omega)$.*

Proof. We assume $\lambda > 0$, the case $\lambda = 0$ being trivial. We also assume that Ω is connected, as we can do the modifications independently in each connected component of Ω. Let now $(f_k) \subseteq L^1_{\mathrm{loc}}(\Omega)$ be a minimizing sequence for $\mathcal{F}_\lambda^{p,q}$. In particular, the sequence $(|\mathrm{D}^2 f_k|(\Omega))$ is bounded as well as the sequence $(|f_k(x_i)|)$, for every $i = 1, \ldots, N$. Now we are going to modify (f_k) to obtain a new sequence $(\tilde{f}_k) \subseteq L^1_{\mathrm{loc}}(\Omega)$ that is still minimizing but in Ω is locally uniformly bounded.

There are two cases to be considered:

(a) $N \geq 3$ and there are three points $x_{i_1}, x_{i_2}, x_{i_3} \in \{x_1, \ldots, x_N\}$ such that $x_{i_2} - x_{i_1}$ and $x_{i_3} - x_{i_1}$ are linearly independent.
(b) either $N = 0$ or all the points x_i are on a line $\{tv + c : t \in \mathbb{R}\} \subseteq \mathbb{R}^2$, for some $v \in \mathbb{R}^2 \setminus \{0\}$ and $c \in \mathbb{R}$.

We treat the two cases separately.

Case (a). In this case no modification is needed, indeed we show that (f_k) is locally uniformly bounded in Ω. Take a compact set $K \subseteq \Omega$. For $\varepsilon := \frac{1}{2}\mathrm{dist}(K, \partial\Omega)$ we select points $q_0, q_1, \ldots, q_M \in K$ such that $K \subseteq \cup_j B_\varepsilon(q_j)$, then curves $\gamma_j \subseteq \Omega$ joining q_j to q_0, and finally curves $\hat{\gamma}_i \subseteq \Omega$ joining x_i to q_0. Let

$$K' := \bigcup_{j=0}^M \overline{B}_\varepsilon(q_j) \cup \bigcup_{j=1}^M \gamma_j \cup \bigcup_{i=1}^N \hat{\gamma}_i.$$

Then $\cup_i \{x_i\} \cup K \subseteq K' \subseteq \Omega$, and K' is compact and connected. Therefore, to prove uniform boundedness of (f_k) on K, we can assume with no loss of generality that all points x_i belong to K and that K is connected.

Now we take $\delta \in (0, 1)$ small enough so that $\Omega' := B_{2\delta}(K)$ satisfies $\overline{\Omega'} \subseteq \Omega$. Hence Ω' is a connected domain. We show now that Ω' is a (bounded) John domain, then Ω' satisfies Poincaré inequalities, by [7, Lemma 3.1 and Theorem 5.1] and the trivial inequality

$$\left(\fint_{\Omega'} \left| f - \fint_{\Omega'} f \right|^q \mathrm{d}\mathscr{L}^2 \right)^{1/q}$$

$$\leqq 2 \left(\fint_{\Omega'} |f - a|^q \mathrm{d}\mathscr{L}^2 \right)^{1/q} \qquad \text{for every } a \in \mathbb{R}$$

that holds for every $f \in L^1(\Omega')$ and $q \in [1, \infty)$. Fix any $p_0 \in K$. We have to show that there exist $0 < \alpha \leqq \beta$ such that for every $p \in \Omega'$, there exists a rectifiable curve $\gamma : [0, l(\gamma)] \to \Omega'$, parametrized by arc length, joining p to p_0 and such that $l(\gamma) \leqq \beta$ and

$$\mathrm{dist}(\gamma(t), \partial\Omega') \geqq \frac{\alpha t}{l(\gamma)} \qquad \text{for every } t \in [0, l(\gamma)]. \tag{4.13}$$

To prove this, notice first that there exists $\beta' > 0$ such that for every $p \in K$, there exists rectifiable curve γ, parametrized by arc length, joining p to p_0, with

image contained in $B_\delta(K) \subseteq \Omega'$ and length bounded by β'. This follows from the connectedness of $B_\delta(K)$ and the compactness of K (simply take a finite covering of K of balls of radius δ and centre in K and consider the rectifiable curves with image in $B_\delta(K)$ joining the centres of these balls); also, γ satisfies (4.13) with $\alpha := \delta$. Then the claim for arbitrary $p \in \Omega'$ follows: indeed, for any $p \in \Omega' \backslash K$, $p \in B_{2\delta}(q)$ with $q \in K$, then we join the radial curve connecting p to q to the curve connecting q to p_0 obtained as before and we have that $l(\gamma) \leqq 2\delta + \beta' =: \beta$ and moreover γ still satisfies (4.13) (with $\alpha = \delta$ as before): indeed, for $t \in [0, |p - q|]$,

$$\text{dist}(\gamma(t), \partial\Omega') \geqq 2\delta - |p - q| + t \geqq 2\delta \frac{t}{|p - q|} \geqq 2\delta \frac{t}{l(\gamma)},$$

whereas for $t \in [|p - q|, l(\gamma)]$, (4.13) follows as before.

Take also $\psi \in C_c^\infty(\mathbb{R}^2)$ such that $\text{supp}\,\psi \subseteq \Omega'$ and $\psi = 1$ on a neighbourhood of K. By Proposition 1.11 and standard calculus rules, the sequence $(|D^2(\psi \hat{f}_k)|(\mathbb{R}^2))$ is bounded, where $\hat{f}_k = f_k - g_k$ with g_k suitable affine perturbation. Therefore, by [11, Proposition 3.1] and the compactness of support of $\psi \hat{f}_k$, we have that $\psi \hat{f}_k$ are uniformly bounded in $L^\infty(\mathbb{R}^2)$, in particular $\hat{f}_k$ are uniformly bounded in $L^\infty(K)$. Now, as $|g_k(x_i)| = |\hat{f}_k(x_i) - f(x_i)|$ are bounded for every $i = i_1, i_2, i_3$, it is easy to infer, by the assumption in (a) that the perturbations g_k are uniformly bounded. Hence $\|f_k\|_{L^\infty(K)}$ is bounded and, since K is arbitrary, the claim follows by weak compactness.

Case (b). If $N \leqq 2$, there is an affine function f_* with $f_*(x_i) = y_i$ for all i, and therefore $\mathcal{F}_\lambda^{p,q}(f_*) = 0$. We can therefore assume $N \geqq 3$. Let $v^\perp$ be a unit vector orthogonal to v, and choose $\varepsilon \in (0, 1)$ sufficiently small that $x_0 := x_1 + \varepsilon v^\perp \in \Omega$. Define

$$\tilde{f}_k(x) := f_k(x) - \frac{1}{\varepsilon} f_k(x_0)(x - x_1) \cdot v^\perp.$$

As $\mathcal{F}_\lambda^{p,q}(\tilde{f}_k) = \mathcal{F}_\lambda^{p,q}(f_k)$, this is also a minimizing sequence, with the additional property that $\tilde{f}_k(x_0) = 0$ for all k. The conclusion follows then from the argument of the previous case.   □

4.2. Proof of the main results

Having proved the results in Section 4.1, Theorems 4.1 and 4.2 follow in a immediate, classical way.

Proof of Theorem 4.1. The statement is proved by the direct method of calculus of variations, by Lemma 4.8 and Lemma 4.9.   □

Proof of Theorem 4.2. Let $\lambda_c := 4\pi$. We argue as in Proposition 4.4, starting from a minimizer f of $\mathcal{F}_{\lambda_c}^{1,1}$ granted by Theorem 4.1. We modify f subtracting $\sum_i (f(x_i) - y_i)g_i$ where this time g_i are rescaled cut cones (see (4.3)), in such a way that

$$\tilde{f} := f - \sum_i (f(x_i) - y_i)g_i$$

has a perfect fit with the data. Since $|D_1^2 g_i|(\mathbb{R}^2) = 4\pi$ (recall e.g. Lemma 4.3), one has

$$\mathcal{F}_\infty^{1,1}(\tilde{f}) \leqq |D_1^2 \tilde{f}|(\Omega) \leqq |D_1^2 f|(\Omega) + \sum_i |D_1^2 g_i|(\mathbb{R}^2)|f(x_i) - y_i|$$

$$= \mathcal{F}_{\lambda_c}^{1,1}(f).$$

This, taking the inequality $\mathcal{F}_\lambda^{1,1} \leqq \mathcal{F}_\infty^{1,1}$ into account, proves that $\tilde{f}$ is a minimizer of $\mathcal{F}_\lambda^{1,1}$ for any $\lambda \geqq \lambda_c$.   □

Data availability The only data relevant to this paper are contained in the paper itself.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Alberti, G.: Rank one property for derivatives of functions with bounded variation. *Proc. R. Soc. Edinb. Sect. A Math.* **123**(2), 239–274, 1993
2. Ambrosio, L., Aziznejad, S., Brena, C., Unser, M.: Linear inverse problems with Hessian–Schatten total variation. Preprint arXiv:2210.04077 (2022)
3. Ambrosio, L., Fusco, N., Pallara, D.: Functions of Bounded Variation and Free Discontinuity Problems. Clarendon Press, Oxford (2000)
4. Arora, R., Basu, A., Mianjy, P., Mukherjee, A. Understanding deep neural networks with rectified linear units. arXiv preprint arXiv:1611.01491 (2016)

5. AZIZNEJAD, S, CAMPOS, J, UNSER, M: Measuring complexity of learning schemes using Hessian–Schatten total variation. arXiv preprint arXiv:2112.06209 (2021)
6. BERGOUNIOUX, M., PIFFET, L.: A second-order model for image denoising. *Set-Valued Var. Anal.* **18**(3–4), 277–306, 2010
7. BOJARSKI, B: Remarks on Sobolev imbedding inequalities. In: LAINE, I., SORVALI, T., RICKMAN, S. (Eds.) Complex Analysis Joensuu 1987, pp. 52–68. Springer, Berlin (1988)
8. BOYER, C., CHAMBOLLE, A., DE CASTRO, Y., DUVAL, V., DE GOURNAY, F., WEISS, P.: On representer theorems and convex regularization. *SIAM J. Optim.* **29**(2), 1260–1281, 2019
9. BREDIES, K., CARIONI, M.: Sparsity of solutions for variational inverse problems with finite-dimensional data. *Calc. Var. Part. Differ. Equ.* **59**(1), 1–26, 2020
10. CAMPOS, J., AZIZNEJAD, S., UNSER, M.: Learning of continuous and piecewise-linear functions with Hessian total-variation regularization. *IEEE Open J. Signal Process.* **3**, 36–48, 2021
11. DEMENGEL, F.: Fonctions à hessien borné. *Annales de l'Institut Fourier* **34**(2), 155–190, 1984
12. DE GIORGI, E., LETTA, G.: Une notion générale de convergence faible pour des fonctions croissantes d'ensemble. Annali della Scuola Normale Superiore di Pisa - Classe di Scienze, 4e série **4**(1), 61–99 (1977)
13. EVANS, L.C., GARIEPY, R.F.: Measure Theory and Fine Properties of Functions. CRC Press, Boca Raton (2015)
14. HINTERBERGER, W., SCHERZER, O.: Variational methods on the space of functions of bounded Hessian for convexification and denoising. *Computing* **76**(1–2), 109–133, 2006
15. KNOLL, F., BREDIES, K., POCK, T., STOLLBERGER, R.: Second order total generalized variation (TGV) for MRI. *Magn. Reson. Med.* **65**(2), 480–491, 2011
16. LEFKIMMIATIS, S., UNSER, M.: Poisson image reconstruction with Hessian–Schatten-norm regularization. *IEEE Trans. Image Process.* **22**(11), 4314–4327, 2013
17. LEFKIMMIATIS, S., WARD, J.P., UNSER, M.: Hessian–Schatten-norm regularization for linear inverse problems. *IEEE Trans. Image Process.* **22**(5), 1873–1888, 2013
18. POURYA, M., GOUJON, A., UNSER, M.: Delaunay-triangulation-based learning with Hessian total-variation regularization. arXiv preprint arXiv:2208.07787 (2022)
19. UNSER, M.: A unifying representer theorem for inverse problems and machine learning. *Found. Comput. Math.* **21**(4), 941–960, 2021
20. UNSER, M., AZIZNEJAD, S.: Convex optimization in sums of Banach spaces. *Appl. Comput. Harmon. Anal.* **56**, 1–25, 2022

Luigi Ambrosio · Camillo Brena
Scuola Normale Superiore,
Piazza dei Cavalieri 7,
56126 Pisa
Italy.
e-mail: luigi.ambrosio@sns.it; camillo.brena@sns.it

and

Sergio Conti
Institut für Angewandte Mathematik,
Universität Bonn,
53115 Bonn
Germany.
e-mail: sergio.conti@uni-bonn.de