Lipschitz continuity, Aleksandrov theorem and characterizations for H-convex functions

Valentino Magnani

Abstract

In the geometries of stratified groups, we show that H-convex functions locally bounded from above are locally Lipschitz continuous and that the class of v-convex functions exactly corresponds to the class of upper semicontinuous H-convex functions. As a consequence, v-convex functions are locally Lipschitz continuous in every stratified group. In the class of step 2 groups we characterize locally Lipschitz H-convex functions as measures whose distributional horizontal Hessian is positive semidefinite. In Euclidean space the same results were obtained by Dudley and Reshetnyak. We prove that a continuous H-convex function is a.e. twice differentiable whenever its second order horizontal derivatives are Radon measures.

Contents

1 Introduction

The notion of convexity plays an important role in several areas of Mathematics, as Calculus of Variations, Differential Geometry, Real Analysis, Optimal Control Theory, Partial Differential Equations and other more. In the setting of stratified groups, different notions of convexity have been recently proposed by Danielli, Garofalo and Nhieu, [9], and by Lu, Manfredi and Stroffolini, [23]. A stratified group is a nilpotent simply connected Lie group $\mathbb{G}$ with a graded Lie algebra $\mathcal{G} = V_1 \oplus \cdots \oplus V_\iota$ and a natural family of dilations $\delta_r : \mathbb{G} \longrightarrow \mathbb{G}$ which are compatible with the group operation, [14]. The sub-Riemannian geometry of a stratified group is determined by its first layer V_1, which yields the so-called "Carnot-Carathéodory distance", [5]. Directions spanned by vector fields of V_1 are called "horizontal directions". Their privileged role appears in the following definition.

A "weakly H-convex" (weakly horizontally convex) function, [9], or a "CC-convex" (Caffarelli-Cabré convex) function, [23], $u : \mathbb{G} \longrightarrow \mathbb{R}$ satisfies the inequality

$$u\left(x\delta_\lambda(x^{-1}y)\right) \leq (1-\lambda)\, u(x) + \lambda u(y) \tag{1}$$

whenever $x, y \in \mathbb{G}$, $0 \leq \lambda \leq 1$ and the geometrical constraint $x^{-1}y \in \mathbb{V}_1 \subset \mathbb{G}$ holds, where $\mathbb{V}_1 = \exp V_1 \subset \mathbb{G}$ is the subset of horizontal directions, see Section 2 for precise definitions. In the sequel, we will refer to functions satisfying (1) simply as "H-convex functions". This notion amounts to require that the restriction of the function to all "horizontal lines" $t \longrightarrow x\delta_t h$ with $h \in \mathbb{V}_1$ is a one-dimensional convex function, as we will precisely illustrate in Proposition 3.9. We note that in groups of step higher than two, horizontal lines introduced in Definition 3.7 are not always lines in the usual sense with respect to graded coordinates. For instance, in Remark 3.8 we show an example of horizontal line that is defined by a parabola with respect to a system of graded coordinates. The most interesting geometric phenomenon related to H-convex functions is the validity of the following principle: the information on the behaviour of a function on the subset of horizontal directions satisfying the Hörmander condition yields a "global" information on the function in terms of the Carnot-Carathéodory distance, which is in turn generated by curves moving along horizontal directions. The explanation of this phenomenon comes from the well known Chow theorem, according to which the Hörmander condition on a set of vector fields, that is a "local" condition, implies the connectedness by horizontal curves, that is a "global condition", see for instance [5]. In a deeper and more general form the previous observations constitute part of the leading themes of [17].

Our first result in this direction is given in Theorem 3.18. Here we prove that every H-convex function which is locally bounded from above is locally Lipschitz with respect to the Carnot-Carathéodory distance. Here we use in a suitable way the following "generating property": let $h_1, h_2, \ldots, h_m$ be elements of $\mathbb{V}_1$ such that $(\exp^{-1} h_1, \ldots, \exp^{-1} h_m)$ is a basis of V_1. Then every direction $v \in \mathbb{G}$ in the closed unit ball can be written as the following ordered product

$$v = \delta_{a_1} h_{i_1} \delta_{a_2} h_{i_2} \cdots \delta_{a_N} h_{i_N},$$

where $N \in \mathbb{N}$ and $(i_1, \ldots, i_N) \in \{1, 2, \ldots, m\}^N$ depend only on $\mathbb{G}$ and the vector $a = (a_1, \ldots, a_N)$ depends on v and it varies in a compact neighbourhood of the origin in $\mathbb{R}^N$. This allows us to extend the one-dimensional Lipschitz property along horizontal lines to all the space. Note that the Lipschitz continuity result of Theorem 9.1 in [9] does not apply directly to H-convex functions which are locally bounded from above in that one needs to prove first that they are at least in L^1_{loc}. In fact, locally summable H-convex functions are locally Lipschitz continuous, after redefinition on a set of measure zero, see Proposition 6.6. Rickly has recently proved that only measurability of H-convex functions suffices to prove Lipschitz continuity, [31].

Another interesting approach is that of [23], according to which a "v-convex function" (convex in the viscosity sense) is an upper semicontinuous function $u : \mathbb{G} \longrightarrow \mathbb{R}$ such that

$$\nabla^2_H u \geq 0 \quad \text{in the viscosity sense.} \tag{2}$$

The symmetrized horizontal Hessian $\nabla^2_H u$ of a function u of class C^2 is the matrix of elements $(X_iX_ju + X_jX_iu)/2$, for every $i, j = 1, 2, \ldots, m$, where $(X_1, X_2, \ldots, X_m)$ is an orthonormal basis of V_1. Condition (2) means that for every $x \in \Omega$ and every $\varphi \in C^2(\mathbb{G})$ such that $\varphi \geq u$ in a neighbourhood of x and $u(x) = \varphi(x)$, we have $\nabla^2_H \varphi(x) \geq 0$. Here it is rather natural wondering whether the class of v-convex functions coincides with that of upper semicontinuous H-convex functions. The second main result of the paper gives a positive answer to this question, see Theorem 4.5. Recently, different proofs of this result have been given in [21], [30] and [36]. As a consequence of Theorem 4.5 and of Corollary 3.19, it follows that v-convex functions are locally Lipschitz continuous in every stratified group. Let us briefly describe our approach. Proposition 4.3 shows that upper semicontinuous H-convex functions are v-convex in every stratified group. The difficult part is to prove the converse to this statement. This is the heart of the proof of Theorem 4.5. Reasoning by contradiction, we assume that u is not H-convex, then we look for a point $\xi \in \mathbb{G}$ and a test function ϕ of class C^2 that touches u from above at ξ and that satisfies the condition $X_1^2\phi(\xi) < 0$. The horizontal direction X_1 corresponds to that direction where the function u fails to be one-dimensional convex. Working in graded coordinates, see Definition 2.2, and performing a left translation, it is not restrictive to assume that $u(0) > \max\{u(\alpha, 0, \ldots, 0), u(\beta, 0, \ldots, 0)\}$, with $\alpha < 0 < \beta$. The general scheme of our proof is that of [4], where remarkable modifications are added, due to the complexity of stratified groups. Let us briefly recall this scheme. We consider suitable smooth functions ψ_ε and open neighbourhoods O_ε of the segment $](\alpha, 0, \ldots, 0), (\beta, 0, \ldots, 0)[$ such that $|X_1^2\psi_\varepsilon|$ restricted to O_ε is less than or equal to a constant independent of ε. The shape of O_ε shrinks around the direction of x_1 as $\varepsilon \to 0^+$, in addition the upper semicontinuity of u allows for proving that $u < \eta_{\varepsilon_0} = -Cx_1^2 + \psi_{\varepsilon_0}$ on $\partial O_{\varepsilon_0}$ for suitably small positive numbers ε_0 and C. The constant C can be chosen so that $X_1^2(-Cx_1^2 + \psi_{\varepsilon_0}) < 0$ on $\overline{O}_{\varepsilon_0}$. Finally, the strict inequality between u and η_{ε_0} on $\partial O_{\varepsilon_0}$ gives a number v_0 such that $\phi = v_0 + \eta_{\varepsilon_0}$ touches u from above at some point $\xi \in \overline{O}_{\varepsilon_0}$ and $X_1^2\phi(\xi) < 0$.

Several obstacles are hidden along this path in the case of general stratified groups. First of all, the group operation is far from being manageable due to the nontrivial Baker-

Campbell-Hausdorff formula, [19]. The first idea is of using a family of test functions containing the polynomial coefficients of the vector field X_1 itself

$$\psi_\varepsilon(x) = \varepsilon^{-2}\left[\sum_{j=2}^{m} x_j^2 + \sum_{s=m+1}^{q}\left(x_s^2 + a_{1s}(x)^2\right)\right], \tag{3}$$

where $X_1 = \partial_{x_1} + \sum_{s=m+1}^{q} a_{1s}(x)\,\partial_{x_s}$. After this choice, the demanding technical part is finding a constant $C > 0$ such that $\sup_{O_\varepsilon} |X_1^2\psi_\varepsilon| < C$ for every $\varepsilon > 0$ suitably small. This requires the study of the second order differential operator X_1^2. Here a nontrivial technical piece appears in Lemma 4.4, where we study some partial derivatives of coefficients of X_1, which are of crucial importance to the estimate of $X_1^2\psi_\varepsilon$ on O_ε. To do this, we systematically use the relation between the homogeneous polynomials appearing in the Baker-Campbell-Hausdorff formula and the coefficients of X_1, as we recall in formula (19). The fact that the set O_ε is defined by

$$O_\varepsilon = \left\{x \in \mathbb{R}^q \,\middle|\, \alpha < x_1 < \beta,\ \psi_\varepsilon(x) < M + 1\right\} \tag{4}$$

permits us to estimate all the factors $\varepsilon^{-2}a_{1j}a_{1s}$ and $\varepsilon^{-2}x_j a_{1s}$ appearing in the expression of $X_1^2\psi_\varepsilon$, see (50) and (51). In fact, due to (3) and (4) we clearly have

$$\sup_{x\in O_\varepsilon}\left\{\max\left\{|x_j\, a_{1s}(x)\,\varepsilon^{-2}|, |a_{1j}(x)\, a_{1s}(x)\,\varepsilon^{-2}|\right\}\right\} \le \frac{(M+1)}{2}.$$

The constant M is the maximum of u on some fixed compact neighbourhood of the segment $[(\alpha, 0, \dots, 0), (\beta, 0, \dots, 0)]$. For ε suitably small the open set O_ε is contained in this neighbourhood, then the definition of O_ε and a suitable rescaling of u, which preserves v-convexity, imply that $u(x) < -Cx_1^2 + \psi_\varepsilon(x)$ for every $x \in \partial O_\varepsilon$, where C only depends on M and on the fixed rescaling of u. This conflicts with v-convexity of u, proving Theorem 4.5.

The third main result of the present paper studies the relationship between H-convexity and the horizontal Hessian. In Euclidean space Bakel'man, [3], proved that the distributional second derivatives of a convex function are signed Radon measures. Reshetnyak established that a locally summable function is equivalent to a convex function if and only if its distributional Hessian is positive semidefinite, [29]. The characterization of distributions with positive semidefinite Hessian as convex functions is due to Dudley, [11]. Clearly the same question can be tackled considering H-convex functions and the distributional horizontal Hessian D_H^2, see Definition 5.3. It has been proved in [9] and [23] that for every locally summable H-convex function u the matrix of distributions $D_H^2 u$ is nonnegative, hence it is formed by Radon measures. The natural way to obtain this result is the approximation by convolutions

$$u_\varepsilon(x) = \int_{\mathbb{G}} \varepsilon^{-Q}\vartheta(\delta_{1/\varepsilon}y)\, u(y^{-1}x)\, dy = \int_{\mathbb{G}} \varepsilon^{-Q}\vartheta\left(\delta_{1/\varepsilon}(xy^{-1})\right)\, u(y)\, dy,$$

where ϑ is a smooth nonnegative function with compact support which satisfies $\int_{\mathbb{G}} \vartheta = 1$ and Q is the Hausdorff dimension of the group. In fact, H-convexity is preserved by left translations, then one easily checks that the H-convexity of u implies the H-convexity of u_ε for every $\varepsilon > 0$ and this fact along with L^1_{loc} convergence of u_ε to u suffices to prove that $D^2_H u \geq 0$. However, this method fails if used to show that locally summable functions u with $D^2_H u \geq 0$ are equivalent to locally Lipschitz H-convex functions. It suffices to check that the equality

$$X_i X_j \left[\vartheta\left(\delta_{1/\varepsilon}(xy^{-1})\right)\right] = \varepsilon^{-2}(X_i X_j \vartheta)\left(\delta_{1/\varepsilon}(xy^{-1})\right)$$

in general does not hold, when y is fixed and the operators X_i differentiate with respect to the variable x. In fact, X_i are not right invariant differential operators. To overcome this problem, one simply defines the different convolution

$$u_\varepsilon(x) = \int_{\mathbb{G}} \varepsilon^{-Q}\vartheta(\delta_{1/\varepsilon}y)\, u(xy^{-1})\, dy = \int_{\mathbb{G}} \varepsilon^{-Q}\vartheta\left(\delta_{1/\varepsilon}(y^{-1}x)\right)\, u(y)\, dy, \tag{5}$$

that satisfies

$$\nabla^2_H u_\varepsilon(x) = \int_{\mathbb{G}} (\nabla^2_H \vartheta_\varepsilon)(y^{-1}x)\; u(y)\, dy, \tag{6}$$

where $\vartheta_\varepsilon(y) = \varepsilon^{-Q}\vartheta(\delta_{1/\varepsilon}y)$. From the assumption $D^2_H u \geq 0$, we would be tempted to infer from (6) that $\nabla^2_H u_\varepsilon(x) \geq 0$, but we are not allowed for this conclusion. In fact, to use the hypothesis $D^2_H u \geq 0$, the horizontal Hessian ∇^2_H inside the integral (6) must differentiate with respect to the variable y. To overcome this difficulty, we seek those stratified groups and those mollificators ϑ such that the following key property holds

$$(\nabla^2_H \vartheta)(y) = (\nabla^2_H \vartheta)(y^{-1}). \tag{7}$$

In Euclidean space (7) becomes trivial for even functions, because ∇^2_H coincides with the usual ∇^2. The situation changes in the case of noncommutative stratified groups, where the form of ∇^2_H depends on the algebraic structure of the group. In Theorem 5.6 we prove (7) for all 2 step stratified groups. The proof of this result relies on a detailed analysis of the operators $X_i X_j$ in the case of 2 step groups. Due to (7) we have

$$\nabla^2_H u_\varepsilon(x) = \int_{\mathbb{G}} (\nabla^2_H \vartheta_\varepsilon)(x^{-1}y)\; u(y)\, dy = \int_{\mathbb{G}} \nabla^2_{yH}\left[\vartheta_\varepsilon(x^{-1}y)\right]\; u(y)\, dy, \tag{8}$$

where the symbol ∇_{yH} specifies that the operator ∇^2_H differentiates with respect to the variable y. This allows us to desume that $\nabla^2_H u_\varepsilon \geq 0$, then u_ε is H-convex. The rest of the proof follows by standard compactness arguments and the estimates (9), which are discussed below. All these considerations can be extended without difficulty to distributions represented by signed Radon measures, instead of locally summable functions. Then we are arrived at our main result, stated in Theorem 5.7. Here we prove that in every step 2 stratified group a distribution T represented by a signed Radon measure is defined by a locally Lipschitz H-convex function if and only if $D^2_H T \geq 0$. This proves the sub-Riemannian versions of Dudley's and Reshetnyak's theorems in all step 2 stratified groups.

The last part of the present paper concerns the a.e. existence of second derivatives of H-convex functions. In the Euclidean setting the celebrated Aleksandrov-Busemann-Feller theorem (shortly, ABF theorem) states that convex functions on a finite dimensional space are a.e. twice differentiable. We precisely consider the version of Theorem 1 in Section 6.4 of [12]. The problem of obtaining ABF theorem on stratified groups has been raised in several recent papers [1], [4], [9], [23]. In particular, Ambrosio and the author pointed out in [1] that the validity of the following L^∞-estimates

$$\sup_{y\in B_{\xi,r}} |u(y)| \leq C \fint_{B_{\xi,cr}} |u(y)|\, dy \qquad \text{and} \qquad \|\nabla_H u\|_{L^\infty(B_{\xi,r})} \leq \frac{C}{r} \fint_{B_{\xi,cr}} |u(y)|\, dy \tag{9}$$

for every H-convex function u, along with the second order differentiability in the L^1 sense

$$\lim_{r\to 0^+} \frac{1}{r^2} \fint_{B_{x,r}} |u(y) - P_{[x]}(y)|\, dy = 0 \tag{10}$$

for a.e. x, where $P_{[x]}$ is a polynomial of homogeneous degree less than or equal to 2, would imply the following second order pointwise differentiability

$$\lim_{y\to x} \frac{|u(y) - P_{[x]}(y)|}{\rho(x,y)^2} = 0 \tag{11}$$

for a.e. x. We give a complete proof of this fact, see Theorem 6.5. Inequalities (9) have been established in [9], [23] and [21]. The validity of (10) for a.e. x has been proved in [1] for the class of functions with locally H-bounded second variation, see also [25]. Gutiérrez and Montanari have proved that H-convex functions have locally H-bounded second variation, filling the gap to obtain Aleksandrov theorem, [16]. Extension of this result to step two Carnot groups has been established by Danielli, Garofalo, Nhieu and Tournier, [10]. Recently, Trudinger has achieved a further improvement for free divergence Hörmander vector fields of step two, [33].

2 Basic materials on stratified groups

In this section we present essential materials needed for the paper. A *stratified group* is a simply connected nilpotent Lie group $\mathbb{G}$ endowed with a graded Lie algebra $\mathcal{G}$, which is decomposed into a direct sum of subspaces V_j subject to the conditions $V_{j+1} = [V_j, V_1]$ for every $j \in \mathbb{N} \setminus \{0\}$ and $V_j = \{0\}$ whenever j is greater than a positive integer. We denote by ι the maximum integer such that $V_\iota \neq \{0\}$ and we call it the *nilpotence degree* of the group or the *step* of the group. The left translation $l_x : \mathbb{G} \longrightarrow \mathbb{G}$ is defined by $l_x(y) = xy$ for every $x, y \in \mathbb{G}$. The assumption that $\mathbb{G}$ is simply connected and nilpotent ensures that the exponential map $\exp : \mathcal{G} \longrightarrow \mathbb{G}$ is a diffeomorphism. The inverse map of exp is denoted by $\ln : \mathbb{G} \longrightarrow \mathcal{G}$. The subset of *horizontal directions in the group* is defined by $\mathbb{V}_1 = \exp V_1$.

The underlying metric of the group is a left invariant Riemannian metric g such that the subspaces V_j are orthogonal each other. We will always refer to these metrics, called *graded metrics*. The Riemannian volume measure on $\mathbb{G}$ with respect to a graded metric will be denoted by v_g. We also write $|A| = v_g(A)$ for every measurable subset $A \subset \mathbb{G}$. It is clear that v_g is left invariant, hence it is the Haar measure of the group. For ease of notation, we will use the symbol dx when the integration is considered with respect to the Riemannian volume v_g. The averaged integral of a summable map $u : A \longrightarrow \mathbb{R}$ is defined as $u_A = \fint_A u = |A|^{-1} \int_A u$.

The grading of $\mathcal{G}$ allows us to define dilations on the group as follows.

Definition 2.1 (Dilations with sign) For every $t > 0$ we consider the family of maps $\tilde{\delta}_t : \mathcal{G} \longrightarrow \mathcal{G}$ defined as $\tilde{\delta}_t(\sum_{j=1}^{\iota} v_j) = \sum_{j=1}^{\iota} t^j v_j$, where the sum $\sum_{j=1}^{\iota} v_j \in \mathcal{G}$ is the unique representation of a vector of $\mathcal{G}$, provided that $v_j \in V_j$ for every $j = 1, \ldots, \iota$. This notion is motivated by the fact that the composition $\delta_t := \exp \circ \tilde{\delta}_t \circ \ln : \mathbb{G} \longrightarrow \mathbb{G}$ is a group homomorphism and satisfies the one parameter group law $\delta_r(\delta_s y) = \delta_{rs} y$ for every $r, s > 0$ and every $y \in \mathbb{G}$. If $t < 0$, then we define $\delta_t y := \delta_{|t|} y^{-1}$ for every $y \in \mathbb{G}$. Denoting by e the unit element of the group we also define $\delta_0 y := e$ for every $y \in \mathbb{G}$.

By virtue of the left invariance of g we can construct a natural left invariant distance on $\mathbb{G}$ in such a way that it is 1-homogeneous with respect to dilations. To do this, we consider the class of *horizontal curves*, e.g. absolutely continuous curves $\gamma : [a, b] \longrightarrow \mathbb{G}$, such that for a.e. $t \in [a, b]$ they satisfy $\gamma'(t) = \sum_{i=1}^{m} c_i(t) X_i(\gamma(t))$, where $\sum_{i=1}^{m} c_i^2(t) \leq 1$ and $(X_1, \ldots, X_m)$ is an orthonormal basis of V_1. The fact that the Lie algebra generated by V_1 coincides with $\mathcal{G}$ ensures that any pair of points of $\mathbb{G}$ can be joined by an horizontal curve. Hence we can define the finite number

$$\rho(x, y) := \inf \left\{ b - a \;\middle|\; \gamma : [a, b] \longrightarrow \mathbb{G} \text{ is horizontal and } \gamma(a) = x,\ \gamma(b) = y \right\}$$

for any $x, y \in \mathbb{G}$. One can verify that d is a distance on $\mathbb{G}$. This is the so-called *Carnot-Carathéodory* distance. Throughout the paper we will always refer to this distance. The usual Euclidean norm is denoted by $|\cdot|$. The distance $\rho(y, e)$, where e is the unit element of the group, is simply denoted by $\rho(y)$, hence $\rho(e) = 0$. The left invariance of Carnot-Carathéodory distance on stratified groups yields the symmetry property $\rho(y^{-1}) = \rho(y)$. This equality and the notion of dilation with sign give the useful formula

$$\rho(\delta_t y) = |t|\, \rho(y) \tag{12}$$

for every $t \in \mathbb{R}$ and every $y \in \mathbb{G}$. The open ball of center $x \in \mathbb{G}$ and radius $r > 0$ with respect to the Carnot-Carathéodory distance is denoted by $B_{x,r}$. Balls of radius r centered at e will be denoted simply by B_r. The symbols $D_{x,r}$ and D_r denote closed balls with analogous meanings. Note that we have

$$|B_{x,r}| = |B_1|\, r^Q \tag{13}$$

for every $x \in \mathbb{G}$ and any $r > 0$. One can check that the integer Q is the Hausdorff dimension of $\mathbb{G}$ with respect to the Carnot-Carathéodory distance and it is strictly greater than the topological dimension q of the group whenever $\mathbb{G}$ is not Abelian. More information on stratified groups can be found for instance in [14] and [32].

Definition 2.2 (Graded coordinates) We define $n_i = \dim V_i$ and $m_i = \sum_{j=1}^{i} n_j$ for any $i = 1, \dots \iota$. We also define $m_0 = 0$ and $m_1 = m$. We say that a basis $(X_1, \dots, X_q)$ of $\mathcal{G}$ is an *adapted basis* if $(X_{m_{j-1}+1}, X_{m_{j-1}+2}, \dots, X_{m_j})$ is a basis of V_j for any $j = 1, \dots \iota$. We say that $(X_1, \dots, X_q)$ is a *graded basis* if it is an adapted and orthonormal basis with respect to a graded metric. The *graded coordinates* with respect to the basis $(X_1, \dots, X_q)$ are given by the diffeomorphism $F : \mathbb{R}^q \longrightarrow \mathbb{G}$ defined by

$$F(x) = \exp\Big(\sum_{j=1}^{q} x_j X_j\Big).$$

The *degree* of the coordinate x_j is the unique integer d_j such that $X_j \in V_{d_j}$.

We will assume throughout that $(X_1, \dots, X_q)$ represents a graded basis of $\mathcal{G}$ and that $(X_1 \dots, X_m)$ is an orthonormal basis of the first layer V_1. The notions of polynomials on stratifed groups and of homogeneous degree will be important tools throughout the paper. Here we briefly recall these notions, referring to Chapter 1.C of [14]. A *polynomial on* $\mathbb{G}$ is function $P : \mathbb{G} \longrightarrow \mathbb{R}$ such that $P \circ F$ is a polynomial on $\mathbb{R}^q$, where F is a system of graded coordinates. For every polynomial $P : \mathbb{G} \longrightarrow \mathbb{R}$ we have the expression

$$P(x) = \sum_{\alpha} c_\alpha \, x^\alpha, \tag{14}$$

where $\alpha \in \mathbb{N}^q$, $x^\alpha = x_1^{\alpha_1} \cdots x_q^{\alpha_q}$ and only a finite number of coefficients $c_\alpha \in \mathbb{R}$ do not vanish. For every $\alpha \in \mathbb{N}^q$ we define the *homogeneous degree* of α as follows

$$d(\alpha) = \sum_{k=1}^{q} d_k \, \alpha_k$$

and the homogeneous degree of a polynomial $P : \mathbb{G} \longrightarrow \mathbb{R}$ with expression (14) by

$$\text{h-deg}(P) := \max\{d(\alpha) \mid c_\alpha \neq 0\}.$$

We define the vector space

$$\mathcal{P}_{H,k}(\mathbb{G}) = \{P \mid P \text{ is a polynomial on } \mathbb{G} \text{ with h-deg}(P) \leq k\}.$$

Throughout the paper the remarkable Baker-Campbell-Hausdorff formula will be needed.

Theorem 2.3 (Baker-Campbell-Hausdorff formula) *Let $X, Y \in \mathcal{G}$, where $\mathcal{G}$ is the nilpotent Lie algebra of a simply connected group $\mathbb{G}$ of step ι and define*

$$\ln\Big(\exp X \exp Y\Big) = X \circledcirc Y\,.$$

Then we have

$$X \circledcirc Y = \sum_{n=1}^{\iota} \frac{(-1)^{n+1}}{n} \sum_{1\le|\alpha|+|\beta|\le\iota} \frac{(\mathrm{Ad}X)^{\alpha_1}(\mathrm{Ad}Y)^{\beta_1}\cdots(\mathrm{Ad}X)^{\alpha_n}(\mathrm{Ad}Y)^{\beta_n-1}(Y)}{\alpha!\beta!\,|\alpha+\beta|}, \tag{15}$$

where for any $Z \in \mathcal{G}$ the map $\mathrm{Ad}Z : \mathcal{G} \longrightarrow \mathcal{G}$ is the linear operator defined by $\mathrm{Ad}Z(W) = [Z, W]$ and for any $\alpha \in \mathbb{N}^n$ we have assumed the convention $\alpha! = \prod_{l=1}^n \alpha_l$ and $|\alpha| = \sum_{l=1}^n \alpha_l$.

A proof of this important formula can be found for instance in [34]. In order to obtain manageable expressions of vector fields $(X_1, \ldots, X_m)$ with respect to graded coordinates we will also use a less explicit form of (15). In fact, there are uniquely defined homogeneous polynomials $P_s : \mathbb{R}^q \times \mathbb{R}^q \longrightarrow \mathbb{R}$ with h-deg$(P) = d_s$, such that

$$\exp\Big(\sum_{j=1}^q x_j X_j\Big) \exp\Big(\sum_{j=1}^q y_j X_j\Big) = \exp\Big(\sum_{s=1}^q P_s(x, y)\, X_s\Big). \tag{16}$$

Recall that a *homogeneous polynomial* P with h-deg$(P) = k$ satisfies the homogeneity formula $P(\delta_r x) = r^k P(x)$ for every $r > 0$ and every $x \in \mathbb{G}$. By definition of graded coordinates with respect to $(X_1, \ldots, X_q)$ we have

$$F^{-1}\left(F(x)F(y)\right) = \sum_{s=1}^q P_s(x, y)\, e_s := \mathcal{Q}(x, y), \tag{17}$$

where $x, y \in \mathbb{R}^q$ and $(e_1, e_2, \ldots, e_q)$ is the canonical basis of $\mathbb{R}^q$. The vector field X_j with respect to graded coordinates is defined by $\tilde{X}_j = F_*^{-1} X_j$, where for every $p \in N$ the formula

$$f_* X(p) = df(f^{-1}(p))\left(X\left(f^{-1}(p)\right)\right)$$

defines the *image of X under f*, whenever $f : M \longrightarrow N$ is a C^1 diffeomorphism of differentiable manifolds and X is a vector field of M. The vector field $\tilde{X}_j$ is left invariant with respect to left translations on $\mathbb{R}^q$ defined by $y \longrightarrow \mathcal{Q}(x, y)$, then

$$\tilde{X}_j(x) = \left(\partial_{y_j} \mathcal{Q}\right)(x, 0)$$

and a careful calculation leads to the formula

$$\tilde{X}_j = \partial_{x_j} + \sum_{s=m+1}^q a_{js}(x_1, x_2, \ldots, x_{s-1})\, \partial_{x_s} \tag{18}$$

for every $j = 1, \dots, m$. The functions a_{js} are homogeneous polynomials with degree $d_s - d_j$ and are defined by the formula

$$a_{js}(x) = (\partial_{y_j} P_s)(x, 0), \tag{19}$$

see p.621 of [32] for more details. Throughout the paper, we will often identify vector fields $\tilde{X}_j$ on $\mathbb{R}^q$ with vector fields X_j on $\mathbb{G}$.

Definition 2.4 (Horizontal gradient) Let Ω be an open set of $\mathbb{G}$ and let $\xi \in \Omega$. The *horizontal gradient* of $u \in C^1(\Omega)$ at ξ is the vector $\nabla_H u(\xi) = (X_1 u(\xi), \dots, X_m u(\xi))$, where $(X_1, X_2, \dots, X_m)$ is an orthonormal basis of V_1.

The notion of horizontal Hessian naturally appears when one considers the Taylor expansion with respect to the horizontal coordinates. To see this in a more rigorous form, let us consider $u : \Omega \longrightarrow \mathbb{R}$ of class C^2 and fix $\xi \in \Omega$. We wish to obtain the Taylor expansion of u at ξ in the horizontal submanifold $\xi\mathbb{V}_1$. Let $(X_1, X_2, \dots, X_m)$ be an orthonormal basis of V_1 and consider the following function of class C^2

$$h \longrightarrow f(h) = u\Big(\xi \exp\Big(\sum_{j=1}^{m} h_j X_j\Big)\Big)$$

along with its Taylor expansion

$$f(h) = f(0) + \langle \nabla f(0), h \rangle + \frac{1}{2} \langle \nabla^2 f(0) h, h \rangle + o(|h|^2) \tag{20}$$

It is easy to check that $\nabla f(0) = \nabla_H u(\xi)$. Consider the canonical basis $(e_1, \dots, e_m)$ of $\mathbb{R}^m$ and note that

$$\frac{d^2}{dt^2} f\left(t(e_i + e_j)\right)_{|t=0} = (X_i + X_j)^2 u(\xi) = \left(\partial_{x_i} + \partial_{x_j}\right)^2 f(0).$$

Then we obtain the formula

$$f_{x_i x_j}(0) = \frac{1}{2}\left(X_i X_j u(\xi) + X_j X_i u(\xi)\right). \tag{21}$$

The previous relation motivates the following definition.

Definition 2.5 (Horizontal Hessian) Let Ω be an open set of $\mathbb{G}$ and let $\xi \in \Omega$. Consider $u \in C^2(\Omega)$ and an orthonormal basis $(X_1, X_2, \dots, X_m)$ of V_1. The *horizontal Hessian* of u at ξ is defined by the matrix

$$\left(\nabla_H^2 u(\xi)\right)_{ij} = \frac{1}{2}\left(X_i X_j u(\xi) + X_j X_i u(\xi)\right) \tag{22}$$

where $i, j = 1, \dots, m$.

In view of previous definition and the Taylor expansion (20) we get

$$u\Big(\xi \exp\Big(\sum_{j=1}^{m} h_j X_j\Big)\Big) = u(\xi) + \langle \nabla_H u(\xi), h \rangle + \frac{1}{2} \langle \nabla_H^2 u(\xi) h, h \rangle + o(|h|^2). \tag{23}$$

3 Lipschitz continuity of H-convex functions

In this section we recall the notion of H-convex function and of H-convex set, see [9], [23]. The main result is stated in Theorem 3.18, where we prove that H-convex functions locally bounded from above are locally Lipschitz continuous.

Definition 3.1 (H-convex set) We say that $C \subset \mathbb{G}$ is H-convex if for every $x, y \in C$ such that $x^{-1}y \in \mathbb{V}_1$ we have $x\delta_t(x^{-1}y) \in C$ for every $t \in [0,1]$.

The following proposition is a straightforward consequence of Definition 3.1

Proposition 3.2 *Let $C \subset \mathbb{G}$ be H-convex. Then for every $x \in \mathbb{G}$ and every $r > 0$ the set $x\,(\delta_r C)$ is H-convex.*

Convex sets in Euclidean space are authomatically connected. On the contrary, this is not true for H-convex sets in stratified groups, as the following example shows.

Example 3.3 Consider the Heisenberg group $\mathbb{H}^3$ with coordinates (x, y, t) and the group operation $(x, y, t)(\xi, \eta, \tau) = (x+\xi, y+\eta, t+\tau+x\eta-y\xi)$. Let us define the closed disconneted subset $C = L \cup M$ with $L = [0,1] \times \{0\} \times \{0\}$ and $M = [0,1] \times \{0\} \times \{1\}$. It is immediate to check that C is H-convex, in that for every $(x, y) \in L \times M$ we have $x^{-1}y \notin \mathbb{V}_1$, due to the expression $\mathbb{V}_1 = \{(x, y, 0) \mid x, y \in \mathbb{R}\}$. It is also easy to see that the individual subsets L and M are H-convex, hence C is H-convex, but it is not connected.

Throughout the section the open set Ω will be assumed to be H-convex.

Definition 3.4 (H-convex function) We say that a function $u : \Omega \longrightarrow \mathbb{R}$ is H-convex if for every $x, y \in \Omega$ such that $x^{-1}y \in \mathbb{V}_1$ and every $0 \leq \lambda \leq 1$ we have

$$u\left(x\delta_\lambda(x^{-1}y)\right) \leq (1-\lambda)\,u(x) + \lambda u(y). \tag{24}$$

Example 3.5 Let $\mathbb{H}^3$ be the Heisenberg group and choose graded coordinates (x, y, t) with the group operation $(x, y, t)(\xi, \eta, \tau) = (x+\xi, y+\eta, t+\tau+-2x\eta+2y\xi)$. Then the left invariant gauge defined by

$$N(x) = \left[(x_1^2 + x_2^2)^2 + 16\,x_3^2\right]^{1/4} \tag{25}$$

is homogeneous and it satisfies the triangle inequality with respect to the group operation. This function is clearly not convex in the usual sense. However, it has been proved in [9] that this function is H-convex in the more general class of H-type groups.

The H-convexity of the gauge provides us an easy way of constructing an example of H-convex open set which is not connected.

Example 3.6 We consider the metric balls $B_{x,r}$ in $\mathbb{H}^3$ defined using the gauge (25). Due to Proposition 7.4 of [9] these balls are open H-convex sets. Let us define $z_h = (0,0,h)$ with $h>0$. Using the group operation $(x,y,t)(\xi,\eta,\tau) = (x+\xi, y+\eta, t+\tau+-2x\eta+2y\xi)$, one can check that the intersections

$$\{x^{-1}\cdot z_h \cdot y \mid x,y \in B_\varepsilon\} \cap \mathbb{V}_1 \quad \text{and} \quad B_\varepsilon \cap B_{z_h,\varepsilon}$$

are empty for $h>0$ sufficiently large and $\varepsilon>0$ sufficiently small. Then the union $B_\varepsilon \cup B_{z_h,\varepsilon}$ is open, H-convex and not connected.

Definition 3.7 (Horizontal line) For every $x\in\Omega$ and every $h\in\mathbb{V}_1$ we say that the function $l_{x,h}:\mathbb{R}\longrightarrow\mathbb{G}$ defined by $l_{x,h}(t)=x\delta_t h$ is an *horizontal line* of direction h.

Remark 3.8 Note that in groups of step higher than 2 horizontal lines do not appear as "Euclidean lines" when read through a system of graded coordinates. Let us consider the Engel group $\mathbb{E}^4$ with the only nontrivial bracket relations

$$[X_1,X_2]=X_3, \quad [X_1,X_3]=X_4 \tag{26}$$

where the Lie algebra $\mathfrak{e}^4 = V_1\oplus V_2\oplus V_3$ is formed by $V_1=\mathrm{span}\{X_1,X_2\}$, $V_2=\mathrm{span}\{X_3\}$ and $V_3=\mathrm{span}\{X_4\}$. Using graded coordinates (x_1,x_2,x_3,x_4), the Baker-Campbell-Hausdorff formula (15) for 3 step groups shows that

$$e_2\cdot te_1 = (0,1,0,0)\cdot(t,0,0,0) = (t,1,-t/2,t^2/12)$$

and $F(e_2\cdot te_1) = (\exp X_2)\delta_t(\exp X_1)$ where $\exp X_1\in\mathbb{V}_1$. In other words, the horizontal line $t\longrightarrow(\exp X_2)\delta_t(\exp X_1)\in\mathbb{E}^4$ becomes a parabola if read in graded coordinates.

Proposition 3.9 (Characterization) *A function $u:\Omega\longrightarrow\mathbb{R}$ is H-convex if and only if for every $(x,h)\in\Omega\times\mathbb{V}_1$ the composition $u\circ l_{x,h}$ is convex on all disjoint open intervals of $l_{x,h}^{-1}(\Omega)\subset\mathbb{R}$.*

PROOF. Assume that u is H-convex. By continuity of $l_{x,h}$, the set $l_{x,h}^{-1}(\Omega)$ is a family of disjoint intervals $\{J \mid J\in\mathcal{I}\}$, where $\mathcal{I}$ is countable or finite. Let us choose an interval $J\in\mathcal{I}$. We wish to prove that $J\ni t\longrightarrow u(l_{x,h}(t))$ is convex. We fix two points $t,\tau\in J$. The assumption $h\in\mathbb{V}_1$ implies that

$$\delta_{t+\tau}h = \exp((t+\tau)\ln h) = \exp(t\ln h+\tau\ln h) = \exp(t\ln h)\exp(\tau\ln h) = \delta_t h\,\delta_\tau h,$$

where we have used the trivial equality $[\ln h,\ln h]=0$ and the Baker-Campbell-Hausdorff formula (15). We have proved that

$$\delta_{t\lambda+(1-\lambda)\tau}h = \delta_{\tau+\lambda(t-\tau)}h = \delta_\tau h\,\delta_{\lambda\,(t-\tau)}h = \delta_\tau h\,\delta_\lambda\left((\delta_\tau h)^{-1}\delta_t h\right) = \delta_\tau h\;\delta_\lambda\big(l_{x,h}(\tau)^{-1}l_{x,h}(t)\big)$$

for every $\lambda \in [0,1]$. It follows that

$$l_{x,h}(t\lambda + (1-\lambda)\tau) = x\delta_{t\lambda+(1-\lambda)\tau}h = l_{x,h}(\tau)\ \delta_\lambda\big(l_{x,h}(\tau)^{-1}l_{x,h}(t)\big)\,.$$

Clearly $l_{x,h}(\tau)^{-1}l_{x,h}(t) \in \mathbb{V}_1$, then the definition of H-convexity gives us the inequality

$$u(l_{x,h}(t\lambda + (1-\lambda)\tau) \le (1-\lambda)\, u(l_{x,h}(\tau)) + \lambda\, u(l_{x,h}(t)).$$

This proves the convexity of $u \circ l_{x,h}$ on J. Conversely, assume that $u \circ l_{x,h}$ is convex on all the intervals where it is defined and for every choice $(x,h) \in \Omega \times \mathbb{V}_1$. Choose $x, y \in \Omega$ such that $x^{-1}y \in \mathbb{V}_1$. By H-convexity of Ω, defining $h = x^{-1}y$, we have $l_{x,h}(\lambda) \in \Omega$ for every $\lambda \in [0,1]$, then it follows that $u\left(x\delta_\lambda(x^{-1}y)\right) = u \circ l_{x,h}(\lambda) \le \lambda\ u(y) + (1-\lambda)\ u(x)$. This completes the proof. $\Box$

Remark 3.10 Note that H-convexity expressed in terms of one-dimensional convexity of restrictions to horizontal lines does not require that the open set Ω is necessarily H-convex. In the sequel, we will refer to this notion when H-convex functions are considered on an arbitrary open set.

Lemma 3.11 *Let $u : \Omega \longrightarrow \mathbb{R}$ be an H-convex function and let $L = \sup_{z\in\partial B_{\xi,3r}\cup\partial B_{\xi,R}} |u(z)|$ be a finite number, where $D_{\xi,R} \subset \Omega$ and $0 < 3r < R$. Then for every $x, y \in B_{\xi,r}$ such that $x^{-1}y \in \mathbb{V}_1$, defining $M_{r,R} = 2L/(R-3r)$, we have*

$$|u(x) - u(y)| \le M_{r,R}\ \rho(x,y). \tag{27}$$

Proof. Let us fix two arbitrary points $x, y \in B_{\xi,r}$ such that $x^{-1}y \in \mathbb{V}_1 \setminus \{e\}$ and define $h = x^{-1}y$. Notice that $h \in B_{2r}$ and that the horizontal line $l_{x,h}(t)$ is contained in $B_{\xi,3r}$ for every $t \in [0,1]$. In particular we have $l_{x,h}(0) = x$ and $l_{x,h}(1) = y$. For elementary topological reasons the horizontal line $l_{x,h}$ meets the boundaries $\partial B_{\xi,3r}$ and $\partial B_{\xi,R}$, then there exist numbers $t_2 < t_1 < 0 < 1 < T_1 < T_2$ such that $l_{x,h}(t_2) \in \partial B_{\xi,R}$, $l_{x,h}(t_1) \in \partial B_{\xi,3r}$, $l_{x,h}(T_1) \in \partial B_{\xi,3r}$ and $l_{x,h}(T_2) \in \partial B_{\xi,R}$. We can find an open interval $I \subset l_{x,h}^{-1}(\Omega)$ containing the subset $\{t_i, T_i \mid i = 1,2\}$ and by Proposition 3.9 the restriction of the function $\varphi = u \circ l_{x,h}$ to I is convex. From convexity of φ we reach the inequality

$$\frac{|\varphi(t) - \varphi(t')|}{|t-t'|} \le \max\left\{\frac{|\varphi(t_2) - \varphi(t_1)|}{|t_2 - t_1|}, \frac{|\varphi(T_2) - \varphi(T_1)|}{|T_2 - T_1|}\right\}. \tag{28}$$

for every $t, t' \in [t_1, T_1]$ such that $t \ne t'$. Now a delicate step appears: due to the condition $h \in \mathbb{V}_1$, we have the equality

$$\rho(x\delta_t h, x\delta_{t'}h) = |t-t'|\,\rho(h) = |t-t'|\,\rho(x,y). \tag{29}$$

In fact, if we consider $h = \exp v$ with $v \in V_1$, then we have $\delta_t h = \exp tv$ and

$$(\delta_t h)^{-1}\delta_{t'}h = \exp(-t)v\ \exp t'v.$$

Since the vectors $(-t)v$ and $t'v$ are proportional, the Baker-Campbell-Hausdorff formula becomes trivial giving, $(\delta_t h)^{-1}\delta_{t'}h = \exp(t'-t)v$. The condition $v \in V_1$ also implies that $\delta_{(t'-t)} \exp v = \exp(t'-t)v$, then (29) follows. Here we have used dilations with sign and relation (12). We divide inequality (28) by $\rho(x,y)$, then formula (29) yields

$$\frac{|\varphi(t)-\varphi(t')|}{\rho\left((l_{x,h}(t), l_{x,h}(t')\right)} \leq \max\left\{\frac{|\varphi(t_2)-\varphi(t_1)|}{\rho\left((l_{x,h}(t_2), l_{x,h}(t_1)\right)}, \frac{|\varphi(T_2)-\varphi(T_1)|}{\rho\left((l_{x,h}(T_2), l_{x,h}(T_1)\right)}\right\}. \tag{30}$$

Taking into account that $\mathrm{dist}(\partial B_{\xi,3r}, \partial B_{\xi,R}) \geq R-3r > 0$ and considering (30) with $t=1$ and $t'=0$, it follows that

$$\frac{|u(y)-u(x)|}{\rho(x,y)} \leq \frac{1}{R-3r} \max\left\{|u(l_{x,h}(t_2))-u(l_{x,h}(t_1))|, |u(l_{x,h}(T_2))-u(l_{x,h}(T_1))|\right\}.$$

By hypothesis, the previous inequality becomes

$$|u(y)-u(x)| \leq \frac{2L}{R-3r}\,\rho(x,y). \tag{31}$$

This ends the proof. □

In order to extend the Lipschitz property (27) to all points $x, y \in B_{\xi,r}$ without the geometric constraint $x^{-1}y \in \mathbb{V}_1$, the following proposition will be of crucial importance. It will be also applied in Theorem 3.17, in order to obtain boundedness of H-convex functions bounded from above.

Proposition 3.12 (Generating property) *Let $h_1, h_2, \ldots, h_m$ be elements of $\mathbb{V}_1$ such that $\ln h_1, \ldots, \ln h_m$ is a basis of V_1. Then there exists a positive integer N, a vector of integers $(i_1, \ldots, i_N) \in \{1,2,\ldots,m\}^N$ and an open bounded neighbourhood of the origin $O \subset \mathbb{R}^N$ such that the following set*

$$\left\{\prod_{s=1}^{N} \delta_{a_s} h_{i_s} \;\middle|\; (a_1, a_2, \ldots, a_N) \in O\right\}, \tag{32}$$

is an open neighbourhood of $e \in \mathbb{G}$, where the product of elements is understood respecting their numbering order.

The proof of Proposition 3.12 is contained in Lemma 1.40 of [14], see also the proof of Corollary 3.3 in [28]. Using notation of this proposition, we define the map $\mathcal{F}: \mathbb{R}^N \longrightarrow \mathbb{G}$,

$$\mathcal{F}(a_1, a_2, \ldots, a_N) = \prod_{s=1}^{N} \delta_{a_s} h_{i_s}. \tag{33}$$

According to the previous proposition we observe that there exists $r_0 > 0$ such that $\partial B_{r_0} = \mathcal{F}(C_0)$, where C_0 is compact set of $\mathbb{R}^N$ which is contained in O. For every $r > 0$ the map $\mathcal{F}$ satisifes the homogeneity property

$$\mathcal{F}(ra_1, \ldots, ra_N) = \delta_r\left(\prod_{s=1}^{N} \delta_{a_s} h_{i_s}\right) = \delta_r\left(\mathcal{F}(a_1, \ldots, a_N)\right). \tag{34}$$

This immediately implies that $\mathcal{F}$ is surjective and that the compact set $C_1 = r_0^{-1} C_0$ satisfies the condition

$$\mathcal{F}(C_1) = \partial B_1. \tag{35}$$

Theorem 3.13 *Let $u : \Omega \longrightarrow \mathbb{R}$ be an H-convex function such that $\sup_{D_{\xi,R}} |u| < \infty$, where $D_{\xi,R} \subset \Omega$. Then there exists $c > 0$ depending only on the group such that for every $0 < r < R/3c$ and every $x, y \in B_{\xi,r}$ we have*

$$|u(x) - u(y)| \leq \sup_{D_{\xi,R}} |u| \left(\frac{c}{R - 3\,c\,r} \right) \rho(x, y). \tag{36}$$

Proof. Let us consider $\mathcal{F} : \mathbb{R}^N \longrightarrow \mathbb{G}$ defined in (33) and let $C_1 \subset \mathbb{R}^N$ denote the compact subset which satisfies condition (35). Let us define the numbers

$$\upsilon_0 = \max_{b \in C_1} |b|\,, \quad \upsilon_1 = \sum_{s=1}^{N} \rho(h_{i_s}) \quad \text{and} \quad c = 2\,\upsilon_0\,\upsilon_1 \tag{37}$$

We arbitrarily choose two different points $x, y \in B_{\xi,r}$ and we define the number $\tau = \rho(x,y) > 0$. By (34) we can write $y = x\mathcal{F}(\tau b)$ for some $b \in C_1$, then condition $3\,c\,r < R$ implies that for every $k = 1, 2, \ldots, N$ we have

$$x \prod_{s=1}^{k} \delta_{\tau b_s} h_{i_s} \in B_{\xi, c\,r}. \tag{38}$$

In view of Lemma 3.11 there exists a constant $M_{c\,r,R}$ such that for every $z, w \in B_{\xi,c\,r}$ with $z^{-1}w \in \mathbb{V}_1$ we have

$$|u(z) - u(w)| \leq M_{c\,r,R}\, \rho(z, w). \tag{39}$$

We define the points $x_0 = x$ and $x_k = x \prod_{s=1}^{k} \delta_{\tau b_s} h_{i_s}$ for every $k = 1, \ldots, N$, observing that $x_{k-1}^{-1} x_k = \delta_{\tau b_k} h_{i_k} \in \mathbb{V}_1$ and $x_N = y$. Thus, from conditions (38) and (39) we conclude that

$$|u(x_k) - u(x_{k-1})| \leq\, M_{c\,r,R}\, \rho(x_k, x_{k-1}) \leq\, M_{c\,r,R}\, |b_k|\, \rho(h_{i_k})\, \rho(x, y)$$

for every $k = 1, 2, \ldots, N$. Due to (37) we arrive at the following estimate

$$|u(x_k) - u(x_{k-1})| \leq\, M_{c\,r,R}\, \upsilon_0\, \rho(h_{i_k})\; \rho(x, y).$$

The expression of $M_{c\,r,R}$ given in Lemma 3.11 and the triangle inequality yield

$$|u(x_N) - u(x_0)| = |u(y) - u(x)| \leq \frac{2L\,\upsilon_0\,\upsilon_1}{R - 3\,c\,r}\, \rho(x, y) = \frac{c\,L}{R - 3\,c\,r}\, \rho(x, y), \tag{40}$$

where $L = \sup_{z \in \partial B_{\xi,3cr} \cup \partial B_{\xi,R}} |u(z)|$. The arbitrary choice of $x, y \in B_{\xi,r}$ leads us to the conclusion. □

Remark 3.14 Note that the factor $\sup_{D_{\xi,R}}|u|$ in (36) could be precisely replaced by $\sup_{z\in\partial B_{\xi,3cr}\cup\partial B_{\xi,R}}|u(z)|$

Corollary 3.15 *Let $u:\Omega\longrightarrow\mathbb{R}$ be an H-convex function such that $\sup_{D_{\xi,4cr}}|u|<\infty$ and $D_{\xi,4cr}\subset\Omega$, where c is defined in Theorem 3.13. Then u is locally Lipschitz and for every $x,y\in B_{\xi,r}$ we have*

$$|u(x)-u(y)|\leq\frac{1}{r}\,\|u\|_{L^\infty(D_{\xi,4cr})}\,\rho(x,y). \tag{41}$$

Remark 3.16 The Lipschitz condition (41) implies the a.e. intrinsic differentiability of u, see [28] for the general result. As a consequence, we have

$$|\nabla_H u(x)|\leq\frac{1}{r}\,\|u\|_{L^\infty(D_{\xi,4cr})}. \tag{42}$$

Theorem 3.17 (Boundedness) *Let $u:\Omega\longrightarrow\mathbb{R}$ be an H-convex function and assume that it is locally bounded from above. Then it is locally bounded.*

PROOF. Let C_1 be the compact set in (35) and define the compact set $\cup_{|t|\leq1}t\,C_1=\tilde{C}_1$. By formulae (34) and (35) one can easily check that $\mathcal{F}(\tilde{C}_1)=D_1$. Let us define the number $\alpha_j=1+\max_{a\in\tilde{C}_1}|a_j|$ and the interval $I_j=[-\alpha_j,\alpha_j]$ for every $j=1,\ldots,N$. We set $K_0=\prod_{j=1}^N I_j\subset\mathbb{R}^N$ and we note that $D_1\subset\mathcal{F}(K_0)=K\subset\mathbb{G}$. Let us assume that $K\subset\Omega$. We choose h_{i_1} defined in (33). By Proposition 3.9, the convexity of the function $I_1\ni a_1\longrightarrow u(\delta_{a_1}h_{i_1})$ gives

$$\mu_1=2u(e)-M\leq 2u(e)-u(\delta_{-a_1}h_{i_1})\leq u(\delta_{a_1}h_{i_1}) \tag{43}$$

where $M=\sup_{z\in K}u(z)$ is finite by hypothesis and $a_1\in I_1$. We denote by e the unit element of $\mathbb{G}$. We fix $a_1\in I_1$ and consider the convex function $I_2\ni a_2\longrightarrow u(\delta_{a_1}h_{i_1}\delta_{a_2}h_{i_2})$, where h_{i_2} is defined in (33). It follows that

$$\mu_2=2\mu_1-M\leq 2u(\delta_{a_1}h_{i_1})-u(\delta_{a_1}h_{i_1}\delta_{-a_2}h_{i_2})\leq u(\delta_{a_1}h_{i_1}\delta_{a_2}h_{i_2}) \tag{44}$$

for every $(a_1,a_2)\in I_1\times I_2$. One can clearly repeat this argument N times, achieving

$$\mu_N\leq 2\,u\left(\prod_{s=1}^{N-1}\delta_{a_s h_{i_s}}\right)-u\left(\prod_{s=1}^{N-1}\delta_{a_s h_{i_s}}\delta_{-a_N}h_{i_N}\right)\leq u\left(\prod_{s=1}^{N}\delta_{a_s h_{i_s}}\right) \tag{45}$$

where $\mu_j=2\,\mu_{j-1}-M$ for every $j=2,\ldots,N$ and $(a_1,a_2,\ldots,a_N)\in K_0$. We have proved that $\mu_N\leq\inf_{z\in K}u(z)$, hence u is bounded on K. Let $\xi\in\Omega$ and $r>0$ such that $\xi\delta_r K\subset\Omega$. Then the function $K\ni y\longrightarrow u(\xi\delta_r y)$ is H-convex and it is bounded on K, namely u is bounded on $\xi\delta_r K\supset D_{\xi,r}$. This proves that u is locally bounded. $\square$

Theorem 3.18 (Lipschitz continuity) *An H-convex function $u : \Omega \longrightarrow \mathbb{R}$ which is locally bounded from above is locally Lipschitz.*

Proof. We first note that Theorem 3.17 ensures the local boundedness of u. Let $\xi \in \Omega$ and choose a closed ball $D_{\xi,R}$ contained in Ω. Let us fix $2r = R/4c$, so that by Corollary 3.15 u is Lipschitz on $B_{\xi,2r}$. The arbitrary choice of ξ immediately yields the continuity of u. Let K be a compact subset of Ω and let $\mathcal{C} = \{B_{\xi_j,r_j} \mid j = 1, \ldots, \nu\}$ be a finite open covering of K such that u is Lipschitz on $B_{\xi_j,2r_j}$ with Lipschitz constant $L_j = (2r_j)^{-1}\|u\|_{L^\infty(D_{\xi,8cr_j})}$ for every $j = 1, \ldots, \nu$. We define $\mathcal{C}_1 \subset \mathcal{C} \times \mathcal{C}$ as the subfamily of couple of balls $B, B' \in \mathcal{C}$ such that $\mathrm{dist}(B, B') > 0$, then we fix the number

$$M = \max\left\{\max\left\{r_p^{-1} \mid p = 1, 2, \ldots, \nu\right\}, \max\{\mathrm{dist}(B, B')^{-1} \mid (B, B') \in \mathcal{C}_1\}\right\}. \tag{46}$$

We denote by M_1 the maximum of $|u|$ on K. Now we choose two arbitrary points $x, y \in K$ and we consider the following possible cases. If x, y belong to the same ball, then $|u(x) - u(y)| \leq \max\{L_j \mid j = 1, 2, \ldots, \nu\}\, \rho(x, y)$. If x, y belong to different balls B, B' with $\mathrm{dist}(B, B') > 0$ then $|u(x) - u(y)| \leq 2\, M\, M_1\, \rho(x, y)$. The last case occurs when $x \in B_{\xi_p,r_p}$, $y \in B_{\xi_k,r_k}$, $B_{\xi_p,r_p} \neq B_{\xi_k,r_k}$ and $\mathrm{dist}(B_{\xi_p,r_p}, B_{\xi_k,r_k}) = 0$. If $y \in B_{\xi_p,2r_p}$, then $|u(x) - u(y)| \leq \max\{L_j \mid j = 1, 2, \ldots, \nu\}\, \rho(x, y)$, otherwise $y \notin B_{\xi_p,2r_p}$ and we have

$$\rho(y, x) \geq \rho(y, \xi_p) - \rho(x, \xi_p) \geq 2\, r_p - \rho(x, \xi_p) > r_p.$$

In this case we obtain $|u(x) - u(y)| < 2\, M_1\, r_p^{-1} \rho(x, y) \leq 2\, M_1\, M\, \rho(x, y)$. Joining the estimates obtained in all possible cases we conclude that

$$|u(x) - u(y)| \leq \Big(2MM_1 + \max\{L_j \mid j = 1, 2, \ldots, \nu\}\Big)\, \rho(x, y)$$

for every $x, y \in K$. This finishes the proof. □

Corollary 3.19 *Every upper semicontinuous H-convex function is locally Lipschitz.*

Proof. It suffices to observe that upper semicontinuous functions are locally bounded from above, then Theorem 3.18 concludes the proof. □

4 H-convexity coincides with v-convexity

In this section we compare the notion of H-convexity with that of v-convexity. The notion of convexity in the "viscosity sense" has been introduced by Lu, Manfredi and Stroffolini in [23]. We will show that v-convexity and H-convexity are equivalent notions for upper semicontinuous functions.

Definition 4.1 (v-convex function) An upper semicontinuous function $u : \Omega \longrightarrow \mathbb{R}$ is *v-convex* if for every $\phi \in C^2(\Omega)$ that touches u from above at ξ, namely $u(\xi) = \phi(\xi)$ and $u \leq \phi$ on a neighbourhood of ξ, we have $\nabla_H^2 \phi(\xi) \geq 0$. In a more concise way we write

$$\nabla_H^2 u \geq 0 \quad \text{in the viscosity sense.} \tag{47}$$

The following proposition shows that v-convexity is preserved under intrinsic dilations and left translations.

Proposition 4.2 *Let $u : \Omega \longrightarrow \mathbb{R}$ be v-convex. Then for every $x \in \Omega$ and every $r > 0$ the function $u_{x,r} : \Omega_{x,r} \longrightarrow \mathbb{R}$ is v-convex, where $u_{x,r}(y) = u(\delta_{1/r}(x^{-1}y))$ and $\Omega_{x,r} = x\,\delta_r\Omega$.*

PROOF. Suppose that $\phi \in C^2(\Omega)$ touches $u_{x,r}$ from above at $\xi \in \Omega_{x,r}$. Then the function $\phi^{x,r} = \phi \circ l_x \circ \delta_r$ touches u from above at $\eta = \delta_{1/r}(x^{-1}\xi) \in \Omega$ and by hypothesis we have $\nabla^2_H \phi^{x,r}(\eta) \geq 0$. Observing that $\nabla^2_H \phi^{x,r}(\eta) = r^2\, \nabla^2_H \phi(\xi)$ the thesis follows. $\Box$

Proposition 4.3 *Every upper semicontinuous H-convex function $u : \Omega \longrightarrow \mathbb{R}$ is v-convex.*

PROOF. By contradiction, suppose that u is not v-convex. Then there exists $\phi \in C^2(\Omega)$ that touches u from above at $\xi \in \Omega$ and $\nabla^2_H \phi(\xi)$ is not nonnegative. Thus, we have at least one direction $\overline{h} = (h_1, \ldots, h_m)$ such that

$$\frac{d^2}{dt^2}\, \phi(\xi\delta_t h)_{|t=0} = \langle \nabla^2_H \phi(\xi)\overline{h}, \overline{h}\rangle < 0.$$

where $h = \exp\left(\sum_{j=1}^m h_j X_j\right) \in \mathbb{V}_1$. By continuity of second derivatives of u we have that $\frac{d^2}{dt^2}\, \phi(\xi\delta_t h)_{|t=\tau} < 0$ for every $\tau \in [-\alpha, \alpha]$, where $\alpha > 0$ and $\xi\delta_t h \in \Omega$ for every $t \in [-\alpha, \alpha]$. Then the function $[-\alpha, \alpha] \ni t \longrightarrow \phi(\xi\delta_t h)$ is strictly concave in $[-\alpha, \alpha]$. We can choose α suitably small, such that $u(\xi\delta_t h) \leq \phi(\xi\delta_t h)$ for every $t \in [-\alpha, \alpha]$. By strict concavity of $t \longrightarrow \phi(\xi\delta_t h)$ we have that

$$\frac{1}{2}\left(u(\xi\delta_{-\alpha}h) + u(\xi\delta_\alpha h)\right) \leq \frac{1}{2}\left(\phi(\xi\delta_{-\alpha}h) + \phi(\xi\delta_\alpha h)\right) < \phi(\xi) = u(\xi),$$

hence u cannot be H-convex due to Proposition 3.9. $\Box$

The next technical lemma will be used in the proof of Theorem 4.5, which is the main result of this section. We will rely on the notions of homogeneous polynomial and on the explicit formula for vector fields X_j recalled in Section 2.

Lemma 4.4 *Let $\sum_{d(\lambda)=d_j-2} c_\lambda\, x^\lambda$ denote the expression of $\partial_{x_1} a_{1j}$, where a_{1j} satisfies (18) and $j = m+1, \ldots, q$. Then we have $c_{(d_j-2,0,\ldots,0)} = 0$ and in the case $d_j = 2$ the homogeneous polynomial $\partial_{x_1} a_{1j}$ vanishes.*

PROOF. For every fixed $j \geq m+1$ the homogeneous polynomial P_j of (16) can be written as follows

$$P_j(x, y) = \sum_{d(\alpha)+d(\beta)=d_j} \gamma_{\alpha,\beta}\, x^\alpha\, y^\beta,$$

where h-deg$(P_j) = d_j$, and the expression (19) yields

$$a_{1j}(x) = \partial_{y_1} P_j(x, 0) = \sum_{d(\alpha)=d_j-1} \gamma_{\alpha,e_1}\, x^\alpha$$

where $e_1 = (1, 0, \dots, 0) \in \mathbb{N}^q$. The partial derivative of a_{1j} with respect to x_1 is written as follows

$$\partial_{x_1} a_{1j}(x) = \sum_{d(s,\sigma)=d_j-1} s\, \gamma_{(s,\sigma),e_1}\, x^{(s-1,\sigma)} = \sum_{d(\lambda)=d_j-2} c_\lambda\, x^\lambda$$

where $\sigma = (\sigma_2, \dots, \sigma_q) \in \mathbb{N}^{q-1}$ and $c_{(d_j-2,0,\dots,0)} = (d_j - 1)\, \gamma_{(d_j-1)e_1,e_1}$. The polynomial P_j can be written in the following form

$$P_j(x, y) = \sum_{d(\alpha)=d_j-1} \gamma_{\alpha,e_1}\, x^\alpha\, y_1 + R(x, y). \tag{48}$$

As an immediate application of the Baker-Campbell-Hausdorff formula (15), we observe that $P_j(x_1e_1, y_1e_1) = 0$ for every $x_1, y_1 \in \mathbb{R}$, then (48) yields

$$P_j(x_1e_1, y_1e_1) = \nu_0 x_1^{d_j} + \gamma_{(d_j-1)e_1,e_1} x_1^{d_j-1} y_1 + \sum_{s=2}^{d_j} \nu_s\, x_1^{d_j-s}\, y_1^s = 0$$

for every $x_1, y_1 \in \mathbb{R}$, then in particular $\gamma_{(d_j-1)e_1,e_1} = c_{(d_j-2,0,\dots,0)} = 0$. In the case $d_j = 2$ we achieve

$$\partial_{x_1} a_{1j}(x) = \sum_{d(s,\sigma)=1} s\, \gamma_{(s,\sigma),e_1}\, x^{(s-1,\sigma)} = \gamma_{e_1,e_1} = 0.$$

This concludes the proof. □

Theorem 4.5 (H-convexity equals v-convexity) *Let $u : \Omega \longrightarrow \mathbb{R}$ be an upper semicontinuous function. Then u is H-convex if and only if it is v-convex.*

Proof. The first implication follows by Proposition 4.3. We have to prove that v-convexity implies H-convexity. By contradiction, assume that u is not H-convex. In view of Proposition 3.9, we can find $p \in \mathbb{G}$, $h \in \mathbb{V}_1$ and $\alpha, \beta \in \mathbb{R}$ such that

$$p\,[\delta_\alpha h, \delta_\beta h] := p\,\{\delta_t h \mid t \in [\alpha, \beta]\} \subset \Omega,$$

$\alpha < 0 < \beta$ and $u(p) > \max\{u(p\delta_\alpha h), u(p\delta_\beta h)\}$. By virtue of Proposition 3.2 left translations preserve H-convex sets and by Proposition 4.2 they also preserve v-convexity, then we can translate p to the the unit element of the group e and assume that $u(e) > \max\{u(\delta_\alpha h, \delta_\beta h)\}$ and $[\delta_\alpha h, \delta_\beta h] \subset \Omega$. Up to rescaling u by $u \circ \delta_r$, with a suitable $r > 0$, we can find a graded basis $(X_1, \dots, X_q)$ of $\mathcal{G}$ such that $h = \exp X_1$. By virtue of Proposition 4.2, this rescaling preserves v-convexity. The function u will be considered with respect to the graded coordinates $F : \mathbb{R}^q \longrightarrow \mathbb{G}$ associated to the previously fixed graded basis of $\mathcal{G}$. With this convention we can assume $\Omega \subset \mathbb{R}^q$, $[\alpha, \beta] \times \{0\} \subset \Omega$ and $u(0) > \max\{u(\alpha e_1), u(\beta e_1)\}$. Here we have denoted by $(e_1, \dots, e_q)$ the canonical basis of $\mathbb{R}^q$. Adding a constant to u and multiplying u by a suitable large positive number the H-convexity is preserved and we can suppose that $u(0) = 0$ and

$$\max\{u(\alpha e_1), u(\beta e_1)\} < -1. \tag{49}$$

Let $K \subset \Omega$ be a compact neighbourhood of $[\alpha, \beta] \times \{0\}$. The upper semicontinuity of u implies that there exists $M = \max_K u \geq 0$. Define the function

$$\psi_\varepsilon(x) = \varepsilon^{-2} \left[\sum_{j=2}^{m} x_j^2 + \sum_{j=m+1}^{q} \left(x_j^2 + a_{1j}(x)^2 \right) \right],$$

where $\varepsilon > 0$ and $x = (x_1, \ldots, x_q) \subset \mathbb{R}^q$. The polynomial functions a_{1j} appear in the representation of the vector field X_1 with respect to the coordinate system F, namely, $X_1 = \partial_{x_1} + \sum_{j=m+1}^{q} a_{1j}(x)\, \partial_{x_j}$. We define the open set

$$O_\varepsilon(\alpha, \beta, M) = \left\{ x \in \mathbb{R}^q \,\middle|\, \alpha < x_1 < \beta,\ \psi_\varepsilon(x) < M + 1 \right\}$$

for every $0 < \varepsilon < 1$. The vector field $\tilde{X}_1 = F_*^{-1} X_1$ in $\mathbb{R}^q$ has the form (18). For ease of notation, we will use the same symbol X_1 to denote it. Due to (18), by a direct computation we obtain

$$X_1^2 = \partial_{x_1}^2 + \sum_{j=m+1}^{q} \left(\partial_{x_1} a_{1j}\ \partial_{x_j} + 2\, a_{1j}\ \partial_{x_1} \partial_{x_j} \right) + \sum_{l,j=m+1}^{q} a_{1l} \left(\partial_{x_l} a_{1j}\ \partial_{x_j} + a_{1j}\ \partial_{x_l} \partial_{x_j} \right),$$

then it follows that

$$X_1^2 \psi_\varepsilon = \sum_{j=m+1}^{q} \partial_{x_1} a_{1j}\ \partial_{x_j} \psi_\varepsilon + \sum_{l,j=m+1}^{q} a_{1l}\, \partial_{x_l} a_{1j}\ \partial_{x_j} \psi_\varepsilon + \sum_{l,j=m+1}^{q} a_{1l}\ a_{1j}\ \partial_{x_l} \partial_{x_j} \psi_\varepsilon .$$

We have $\partial_{x_q} \psi_\varepsilon = \varepsilon^{-2} 2 x_q$ and

$$\partial_{x_j} \psi_\varepsilon = \varepsilon^{-2} \Big\{ 2 x_j + \sum_{s=j+1}^{q} 2\, a_{1s}\ \partial_{x_j} a_{1s} \Big\}$$

for every $j = 1, \ldots, q-1$. We will assume that the formal expression $\sum_{s=j+1}^{q} \theta(s)$ vanishes for every function θ whenever $j = q$. With this convention we have

$$\begin{aligned} \frac{X_1^2 \psi_\varepsilon}{2} &= \sum_{j=m+1}^{q} \left(x_j\ \partial_{x_1} a_{1j}\ \varepsilon^{-2} \right) + \sum_{j=m+1}^{q} \sum_{s=j+1}^{q} \left(a_{1s}\ \partial_{x_1} a_{1j}\ \varepsilon^{-2} \right)\ \partial_{x_j} a_{1s} \\ &+ \sum_{l,j=m+1}^{q} \left(x_j\ a_{1l}\ \varepsilon^{-2} \right)\ \partial_{x_l} a_{1j} + \sum_{l,j=m+1}^{q} \sum_{s=j+1}^{q} \left(a_{1s}\ a_{1l}\ \varepsilon^{-2} \right)\ \partial_{x_l} a_{1j}\ \partial_{x_j} a_{1s} \qquad (50) \\ &+ \sum_{j=m+1}^{q} \left(a_{1j}^2\ \varepsilon^{-2} \right) + \sum_{l,j=m+1}^{q} \sum_{s=j+1}^{q} \left(a_{1l}\, a_{1j}\, \varepsilon^{-2} \right) \left[\partial_{x_l} a_{1s} \partial_{x_j} a_{1s} + a_{1s} \partial_{x_l} \partial_{x_j} a_{1s} \right]. \qquad (51) \end{aligned}$$

We wish to prove that in the previous expression all products inside the brackets $(\cdots)$ and restricted to the open subset $O_\varepsilon(\alpha,\beta,M)$ are bounded by a constant depending only on M and γ, where we have defined $\gamma=\max\{|\beta|,|\alpha|\}$. For every $x\in O_\varepsilon(\alpha,\beta,M)$ we have

$$\max\left\{|x_j\,a_{1s}(x)\,\varepsilon^{-2}|,|a_{1j}(x)\,a_{1s}(x)\,\varepsilon^{-2}|\right\}\leq\frac{(M+1)}{2}$$

whenever $j,s=m+1,\ldots,q$. Let us consider $\partial_{x_1}a_{1j}=\sum_{d(\lambda)=d_j-2}c_\lambda\,x^\lambda$. By Lemma 4.4 we have $c_{(d_j-2,0,\ldots,0)}=0$ and $\partial_{x_1}a_{1j}$ is identically zero when $d_j=2$. In the case $d_j>2$ every monomial $c_\lambda\,x^\lambda$ with $c_\lambda\neq 0$ contains at least one factor $x_l^{\lambda_l}$ with $l>1$ and $\lambda_l\geq 1$. Then for every $x\in O_\varepsilon(\alpha,\beta,M)$ the estimate

$$|\partial_{x_1}a_{1j}(x)|\leq\varepsilon\sum_{d(\lambda)=d_j-2}|c_\lambda|\,\gamma^{\lambda_1}\,(M+1)^{\sum_{s=2}^q\lambda_s}=\varepsilon\,\omega_j(\gamma,M)$$

holds whenever $0<\varepsilon<1$. For every $t,\tau\geq 0$ we define $\omega_j(t,\tau)=0$ whenever $d_j=2$. This gives in turn the estimate

$$\max\left\{|x_j\,\partial_{x_1}a_{1s}(x)\,\varepsilon^{-2}|,|a_{1j}(x)\,\partial_{x_1}a_{1s}(x)\,\varepsilon^{-2}|\right\}\leq\omega_s(\gamma,M)\,\sqrt{M+1}$$

for every $x\in O_\varepsilon(\alpha,\beta,M)$. The function $\omega_j(t,\tau)$ is nondecreasing with respect to $t\geq 0$. As a consequence, we can find a positive function $C(t,\tau)$ nondecreasing with respect to t such that

$$|X_1^2\psi_\varepsilon(x)|\leq C(\gamma,M)\tag{52}$$

for every $x\in O_\varepsilon(\alpha,\beta,M)$ and every $0<\varepsilon<1$. The monotonicity of $C(\cdot,M)$ implies the existence of $t_0>0$ such that

$$C(t,M)\,t^2<1\quad\text{for every}\quad 0\leq t\leq t_0.\tag{53}$$

The compact set K is also a neighbourhood of 0, then we can find $\mu_1>1$ such that $\delta_{1/\mu_1}K\subset K$ and $\gamma/\mu_1\leq t_0$. Define the v-convex function $u_1(x)=u(\delta_{\mu_1}x)$ on the open subset $\delta_{1/\mu_1}\Omega$ and the numbers $\alpha_1=\alpha/\mu_1$, $\beta_1=\beta/\mu_1$ and $\gamma_1=\gamma/\mu_1$. Fix $0<\varepsilon_1<1$ so that $O_\varepsilon(\alpha_1,\beta_1,M)\subset K_1=\delta_{1/\mu_1}K$ for every $0<\varepsilon<\varepsilon_1$. Clearly we have $\max_{K_1}u_1=M$ and by (52) the inequality

$$|X_1^2\psi_\varepsilon(x)|\leq C(\gamma_1,M)\tag{54}$$

holds for every $x\in O_\varepsilon(\alpha_1,\beta_1,M)$ and every $0<\varepsilon<\varepsilon_1$. Now we define

$$\varphi_\varepsilon(x)=-C(\gamma_1,M)x_1^2+\psi_\varepsilon(x).$$

Our next claim is to prove that $\varphi_\varepsilon(x)>u_1(x)$ for every $x\in\partial O_\varepsilon(\alpha_1,\beta_1,M)$ when $\varepsilon>0$ is chosen suitably small. The inequality (49) implies that $\max\{u_1(\alpha_1e_1),u_1(\beta_1e_1)\}<-1$, therefore the upper semicontinuity of u_1 yields $0<\varepsilon_0<\varepsilon_1$ such that

$$\max\left\{u\Big(\alpha_1e_1+\sum_{s=2}^q x_s\,e_s\Big),u\Big(\beta_1e_1+\sum_{s=2}^q x_s\,e_s\Big)\right\}<-1\tag{55}$$

whenever $\psi_{\varepsilon_0}\Big(\sum_{s=2}^q x_s\, e_s\Big) < M+1$. We will utilize the following topological formula $\partial(A\cap B)\subset(\overline{A}\cap\partial B)\cup(\overline{B}\cap\partial A)$ for any couple of subsets A and B of a topological space. Now we observe that

$$O_{\varepsilon_0}(\alpha_1,\beta_1,M)=\Big\{x\in\mathbb{R}^q\,\Big|\,\alpha_1<x_1<\beta_1\Big\}\bigcap\Big\{x\in\mathbb{R}^q\,\Big|\,\psi_{\varepsilon_0}(x)<M+1\Big\}$$

Let $x\in\partial O_{\varepsilon_0}(\alpha_1,\beta_1,M)$ and consider the case $x\in\{(\alpha_1,y),(\beta_1,y)\}$ with $y=\sum_{s=2}^q y_s e_s$. We have

$$\varphi_{\varepsilon_0}(x)\geq -C(\gamma_1,M)\gamma_1^2+\psi_{\varepsilon_0}(y)\geq -C(\gamma_1,M)\gamma_1^2>-1.$$

If $\psi_{\varepsilon_0}(y)<M+1$, then

$$\max\{u_1(\alpha_1,y),u_1(\beta_1,y)\}<-1<\varphi_{\varepsilon_0}(x). \tag{56}$$

In the case $\psi_{\varepsilon_0}(y)=M+1$ and $\alpha_1\leq x_1\leq\beta_1$ we have

$$\varphi_{\varepsilon_0}(x)\geq -C(\gamma_1,M)\gamma_1^2+M+1>M\geq u_1(x). \tag{57}$$

Estimates (56) and (57) prove that

$$\varphi_{\varepsilon_0}(x)>u_1(x)\quad \text{for every }\ x\in\partial O_{\varepsilon_0}(\alpha_1,\beta_1,M). \tag{58}$$

Due to inequality (54) we also have

$$X_1^2\varphi_{\varepsilon_0}(x)=-2\,C(\gamma_1,M)+X_1^2\psi_{\varepsilon_0}(x)\leq -C(\gamma_1,M)<0 \tag{59}$$

for every $x\in O_{\varepsilon_0}(\alpha_1,\beta_1,M)$. Let us define the number

$$v_0=\inf\Big\{t\in\mathbb{R}\,\Big|\,t+\varphi_{\varepsilon_0}(x)\geq u_1(x)\,\text{for every}\,x\in\overline{O}_{\varepsilon_0}(\alpha_1,\beta_1,M)\Big\}$$

Note first that $0\leq v_0<\infty$ in that $\varphi_{\varepsilon_0}(0)=u_1(0)=0$ and $u_1(x)\leq M$ for every $x\in\overline{O}_{\varepsilon_0}(\alpha_1,\beta_1,M)\subset K_1$. It is easy to check that $v_0+\varphi_{\varepsilon_0}(x)\geq u_1(x)$ for every $x\in\overline{O}_{\varepsilon_0}(\alpha_1,\beta_1,M)$. By definition of v_0 we can find $\xi\in\overline{O}_{\varepsilon_0}(\alpha_1,\beta_1,M)$ where the equality is attained. In view of (58) the point ξ belongs to the open set $O_{\varepsilon_0}(\alpha_1,\beta_1,M)$. As a result, the function $\phi=v_0+\varphi_{\varepsilon_0}$ touches u_1 from above at ξ, but $X_1^2\phi(\xi)<0$. This conflicts with v-convexity of u_1 and concludes the proof. □

Corollary 4.6 *In every stratified group v-convex functions are locally Lipschitz continuous.*

PROOF. By Theorem 4.5 and Corollary 3.19 the proof immediately follows. □

5 Distributional characterization of H-convexity

The main result of this section is Theorem 5.7, where we prove that in 2 step stratified groups every distribution represented by a Radon measure is defined by a locally Lipschitz H-convex function if and only if its distributional horizontal Hessian is positive semidefinite. Throughout the section the symbol Ω will denote an H-convex open set of $\mathbb{G}$. We start with the following simple characterization of H-convexity in the case of regular functions, see also [9] and [23].

Proposition 5.1 *Every $u \in C^2(\Omega)$ is H-convex if and only if $\nabla_H^2 u(x) \geq 0$ for every $x \in \Omega$.*

Proof. By Proposition 3.9 H-convexity is characterized by convexity of $t \longrightarrow u(x\delta_t h)$ for every $x \in \Omega$ and every $h \in \mathbb{V}_1$. Defining $h = \exp\left(\sum_{j=1}^m h_j X_j\right) \in \mathbb{V}_1$ and using formula (23) we get

$$\frac{d^2}{dt^2} u(x\delta_t h)_{|t=\tau} = \langle \nabla_H^2 u(x\delta_\tau h)\overline{h}, \overline{h}\rangle \geq 0. \tag{60}$$

where $\overline{h} = (h_1, \ldots, h_m) \in \mathbb{R}^m$. Formula (60) proves our claim. □

Definition 5.2 (Convolution) Let $u, v \in L^1_{loc}(\mathbb{G})$ where v has compact support. The *convolution* of u and v is defined by

$$u * v(x) = \int_{\mathbb{G}} u(y)\; v(y^{-1}x)\, dy \tag{61}$$

Under assumptions of Definition 5.2 the convolution $u*v$ is a well defined locally summable function. Note that this convolution does not commute, see [14] for more information.

Definition 5.3 (Distributional horizontal Hessian) Let $T \in \mathcal{D}(\Omega)'$ be a distribution. The *horizontal Hessian* of T is the symmetric matrix of distributions defined as follows

$$\langle D_H^2 T, \varphi\rangle = \langle T, \nabla_H^2 \varphi\rangle \tag{62}$$

for every $\varphi \in C_c^\infty(\Omega)$. We say that the horizontal Hessian of T is nonnegative and write $D_H^2 T \geq 0$ if for every $\varphi \in C_c^\infty(\Omega)$ such that $\varphi \geq 0$ the matrix $\langle T, \nabla_H^2 \varphi\rangle$ is nonnegative.

As observed in [9] and [23], the distributional horizontal Hessian of a locally summable H-convex function is a positive semidefinite matrix of Radon measures. For the reader's convenience, in Proposition 5.5 we briefly recall this fact.

Lemma 5.4 *Let $T \in \mathcal{D}(\Omega)'$ be a distribution with $D_H^2 T \geq 0$. Then $D_H^2 T$ is a matrix of Radon measures.*

PROOF. By hypothesis, writing any $X \in V_1$ as $\sum_{j=1}^m \xi_j X_j$ where $(X_1, \dots, X_m)$ is a basis of V_1, we have $X^2T = \sum_{i,j=1}^m \xi_i\xi_j\,(X_iX_j + X_jX_i)T \geq 0$, then X^2T is a Radon measure, see Theorem 2.1.7 of [20] and Theorem 1.54 of [2], hence also

$$X_iX_jT + X_jX_iT = (X_i + X_j)^2T - X_i^2T - X_j^2T$$

is a measure for every $i, j = 1, \dots, m$. $\square$

Proposition 5.5 *Let $u : \Omega \longrightarrow \mathbb{R}$ be a locally summable H-convex function. Then $D_H^2 u$ is a matrix of Radon measures and $D_H^2 u \geq 0$.*

PROOF. Let us choose $\vartheta \in C_c^\infty(B_1)$ such that $\vartheta \geq 0$ and $\int_{\tilde{B}_1} \vartheta = 1$. For every $y \in \mathbb{G}$ we define $\vartheta_\varepsilon(y) = \varepsilon^{-Q}\vartheta(\delta_{1/\varepsilon}y)$, where $\varepsilon > 0$. Let $\varphi \in C_c^\infty(\Omega)$ be a nonnegative function and let $\Omega' \Subset \Omega$ be an open subset containing $\operatorname{supp}\varphi$. We can find $\kappa > 0$ such that

$$\max_{x \in \overline{\Omega'}} \rho(y^{-1}x, x) < \operatorname{dist}(\Omega', \Omega^c) \tag{63}$$

whenever $\rho(y) \leq \kappa$. Then the convolution

$$u_\varepsilon(x) = \vartheta\varepsilon * u(x) = \int_\Omega \vartheta_\varepsilon(y)\, u(y^{-1}x)\, dy$$

is smooth, H-convex and well defined in Ω' for every $\varepsilon \leq \kappa$. The function u_ε is convex along horizontal lines, according to Proposition 3.9 and Remark 3.10. Thus, we can apply Proposition 5.1, that gives $\nabla_H^2 u_\varepsilon \geq 0$ and integrating by parts we achieve

$$\int_{\Omega'} u_\varepsilon\, \nabla_H^2\varphi = \int_{\mathbb{G}} u_\varepsilon\, \nabla_H^2\varphi = \int_{\mathbb{G}} \varphi\, \nabla_H^2 u_\varepsilon \geq 0 \tag{64}$$

for every $\varepsilon \leq \kappa$. The convergence of u_ε to u in $L^1_{loc}(\Omega)$ and Lemma 5.4 conclude the proof. $\square$

Theorem 5.6 *Let $\mathbb{G}$ be a 2 step stratified group. Then there exists a nonnegative function $\vartheta \in C_c^\infty(B_1)$ such that $\int_{B_1} \vartheta = 1$ and*

$$(\nabla_H^2\vartheta)(y) = (\nabla_H^2\vartheta)(y^{-1}) \quad \textit{for every } y \in \mathbb{G}. \tag{65}$$

PROOF. Recall from (18) the form of the horizontal vector field X_j for every $j = 1, \dots, m$ with respect to a system of graded coordinates $F : \mathbb{R}^q \longrightarrow \mathbb{G}$

$$\tilde{X}_j = \partial_{x_j} + \sum_{s=m+1}^{q} a_{js}\, \partial_{x_s}. \tag{66}$$

For ease of notation we will simply write X_j instead of $\tilde{X}_j$. The fact that $\mathbb{G}$ is of step 2 implies that a_{js} has homogeneous degree equal to one for every $j = 1, \ldots, m$ and every $s = m+1, \ldots, q$. Let us consider the second order operator

$$X_i X_j = \partial_{x_i}\partial_{x_j} + \sum_{s=m+1}^{q} (\partial_{x_i} a_{js}\ \partial_{x_s} + a_{js}\ \partial_{x_i}\partial_{x_s}) + \sum_{l=m+1}^{q} a_{il}\ \partial_{x_l}\partial_{x_j}$$
$$+ \sum_{l,s=m+1}^{q} (a_{il}\ \partial_{x_l} a_{js}\ \partial_{x_s} + a_{il}\ a_{js}\ \partial_{x_l}\partial_{x_s})$$

for every $i, j = 1, \ldots, m$. Now we choose two smooth even functions with compact support $\theta_1 : \mathbb{R}^m \longrightarrow [0, +\infty[$ and $\theta_2 : \mathbb{R}^{q-m} \longrightarrow [0, +\infty[$ such that, defining $\theta(\xi, \eta) = \theta_1(\xi) + \theta_2(\eta)$ with $\xi = (x_1, \ldots, x_m)$ and $\eta = (x_{m+1}, \ldots, x_q)$, the support of θ is contained in $F^{-1}(B_1)$ and $\vartheta = \theta \circ F^{-1} : \mathbb{G} \longrightarrow [0, +\infty[$ satisfies $\int_{B_1} \vartheta = 1$. We clearly have $\partial_{x_l}\partial_{x_j}\theta = 0$ for every $l = m+1, \ldots, q$ and every $j = 1, \ldots, m$. In addition, the polynomial a_{js} cannot contain the variable x_l for every $l = m+1, \ldots, q$ in that it has homogeneous degree equal to one, then $\partial_{x_l} a_{js}$ also vanishes. It follows that

$$X_i X_j \theta = \partial_{x_i}\partial_{x_j}\theta + \sum_{s=m+1}^{q} \partial_{x_i} a_{js}\ \ \partial_{x_s}\theta + \sum_{l,s=m+1}^{q} a_{il}\ a_{js}\ \partial_{x_l}\partial_{x_s}\theta.$$

The fact that θ is even easily implies that $(\partial_{x_i}\partial_{x_j}\theta)(x) = (\partial_{x_i}\partial_{x_j}\theta)(-x)$. The homogeneous polynomial a_{js} has homogeneous degree equal to one, then it has the form $\sum_{k=1}^{m} c_k x_k$. Thus, the products $a_{il}a_{js}$ are even functions and we obtain

$$\sum_{l,s=m+1}^{q} a_{il}(x)\ a_{js}(x)\ (\partial_{x_l}\partial_{x_s}\theta)(x) = \sum_{l,s=m+1}^{q} a_{il}(-x)\ a_{js}(-x)\ (\partial_{x_l}\partial_{x_s}\theta)(-x).$$

The factors $\partial_{x_j} a_{js}$ are constants then $\partial_{x_j} a_{js}\ \partial_{x_s}\theta$ is an odd function for every $j = 1, \ldots, m$ and every $s = m+1, \ldots, q$. At this point the symmetrization of $X_i X_j$ will help us. Consider

$$\frac{X_i X_j \theta + X_j X_i \theta}{2} = \partial_{x_i}\partial_{x_j}\theta + \sum_{s=m+1}^{q} \frac{(\partial_{x_i} a_{js} + \partial_{x_j} a_{is})}{2}\ \ \partial_{x_s}\theta + \sum_{l,s=m+1}^{q} a_{il}\ a_{js}\ \partial_{x_l}\partial_{x_s}\theta.$$

We aim to show that $\partial_{x_i} a_{js} + \partial_{x_j} a_{is} = 0$ for every $i, j = 1, \ldots, m$ and every $s = m+1, \ldots, q$. Once this is obtained we immediately achieve

$$(\nabla_H^2 \theta)(x) = (\nabla_H^2 \theta)(-x) \tag{67}$$

for every $x \in \mathbb{R}^q$. In the case of 2 step groups the Baker-Campbell-Hausdorff formula (15) gives a rather manageable expression of the group operation. We have

$$\exp\Big(\sum_{k=1}^{q} x_k X_k\Big) \exp\Big(\sum_{l=1}^{q} y_l X_l\Big) = \exp\Big(\sum_{k=1}^{q} (x_k + y_k) X_k + \frac{1}{2} \sum_{1 \le k < l \le m} (x_k y_l - y_k x_l)\,[X_k, X_l]\Big)$$

clearly $[X_k, X_l] \in V_2$ then

$$[X_k, X_l] = \sum_{s=m+1}^{q} c_{kl}^s X_s$$

for some coefficients c_{kl}^s. Let $\{P_s(x,y)\}_{s=1,\dots,q}$ be the family of homogeneous polynomials satisfying (16). Then the previous relations yield

$$P_s(x,y) = x_s + y_s + \sum_{1 \le k < l \le m} c_{kl}^s \,(x_k y_l - y_k x_l)$$

where $s = m+1, \dots, q$. Differentiating with respect to x_j and y_j we get

$$\partial_{x_i}\partial_{y_j} P_s = \sum_{1 \le k < l \le m} c_{kl}^s \,(\delta_{ik}\delta_{jl} - \delta_{jk}\delta_{il}) = \begin{cases} c_{ij}^s & \text{if } \ i < j \\ 0 & \text{if } \ i = j \\ -c_{ji}^s & \text{if } \ i > j \end{cases} ,$$

then we have proved that

$$\partial_{x_i} a_{js} = \partial_{x_i}\partial_{y_j} P_s = -\partial_{x_j}\partial_{y_i} P_s = -\partial_{x_j} a_{is}$$

for every $i, j = 1, \dots, m$ and every $s = m+1, \dots, q$. As a consequence, formula (67) holds. In order to show rigorously that (67) implies (65) we return to notation of (66). This permits us to stress that the vector field $\tilde{X}_j$ is represented with respect to graded coordinates. Then we write $X_j = F_* \tilde{X}_j$ to indicate the corresponding vector field on $\mathbb{G}$. We have

$$X_j \vartheta = X_j(\theta \circ F^{-1}) = F_* \tilde{X}_j (\theta \circ F^{-1}) = \tilde{X}_j \theta,$$

hence applying X_i to $X_j \vartheta$ and using the previous relations we get

$$(X_i X_j \vartheta)(F(x)) = \tilde{X}_i \tilde{X}_j \theta(x) \quad \text{for every } x \in \mathbb{R}^q.$$

Finally, in view of (67) we conclude that

$$\nabla_H^2 \vartheta(F(x)) = \widetilde{\nabla}_H^2 \theta(x) = \widetilde{\nabla}_H^2 \theta(-x) = \nabla_H^2 \vartheta(F(-x)) = \nabla_H^2 \vartheta(F(x)^{-1}),$$

hence ending the proof. □

Theorem 5.7 (Dudley-Reshetnyak) *Let $\mathbb{G}$ be a stratified group of step 2 and let $T \in \mathcal{D}(\Omega)'$ be a locally finite measure. Then T is defined by a locally Lipschitz H-convex function if and only if $D_H^2 T \ge 0$.*

Proof. In view of Proposition 5.5 we have to prove that the condition $D_H^2 T \ge 0$ implies the existence of a locally Lipschitz H-convex function $u : \Omega \longrightarrow \mathbb{R}$ such that $\langle T, \varphi \rangle = \int_\Omega u\varphi$ for every $\varphi \in C_c^\infty(\Omega)$. The crucial part of the proof is to find a suitable $\vartheta \in C_c^\infty(B_1)$ with $\vartheta \ge 0$ and $\int_{B_1} \vartheta = 1$ such that

$$(\nabla_H^2 \vartheta)(y) = (\nabla_H^2 \vartheta)(y^{-1}) \quad \text{for every } y \in \mathbb{G}. \tag{68}$$

Theorem 5.6 shows that a function ϑ satisfying (68) is available in 2 step groups. Then we define $\vartheta_\varepsilon(y)=\varepsilon^{-Q}\vartheta(\delta_{1/\varepsilon}y)$ and consider the convolution

$$u_\varepsilon(x)=T*\vartheta_\varepsilon=\int_\Omega \vartheta_\varepsilon(y^{-1}x)\,d\mu(y),$$

where μ is a signed and locally finite Radon measure in Ω and $\mathrm{dist}(x,\Omega^c)>\varepsilon$. The left invariance of the second order operator ∇^2_H gives $\nabla^2_{xH}\left[\vartheta_\varepsilon(y^{-1}x)\right]=\left(\nabla^2_H\vartheta_\varepsilon\right)(y^{-1}x)$, where the symbol ∇^2_{xH} specifies that ∇^2_H differentiates with respect to the variable x. As a consequence, we have

$$\nabla^2_H u_\varepsilon(x)=\int_\Omega \left(\nabla^2_H\vartheta_\varepsilon\right)(y^{-1}x)\,d\mu(y),$$

in view of the key property (68) we obtain

$$\nabla^2_H u_\varepsilon(x)=\int_\Omega \left(\nabla^2_H\vartheta_\varepsilon\right)(x^{-1}y)\,d\mu(y)=\int_\Omega \nabla^2_{yH}\left[\vartheta_\varepsilon(x^{-1}y)\right]\,d\mu(y)\geq 0$$

where the last inequality follows by hypothesis. By Proposition 5.1, the smooth function u_ε is H-convex. Consider an arbitrary compact set $K\subset\Omega$ and choose $\lambda_0>0$ such that $K_{\lambda_0}=\{y\in\mathbb{G}\mid \mathrm{dist}(y,K)\leq\lambda_0\}\subset\Omega$. For every $0<\varepsilon<\lambda_0$ we have

$$\int_K |u_\varepsilon(x)|\,dx\leq\int_\Omega\left(\int_K|\vartheta_\varepsilon(y^{-1}x)|\,dx\right)\,d|\mu|(y)\leq|\mu|(K_{\lambda_0})<\infty\,, \tag{69}$$

where $|\mu|$ is the total variation of μ, see [2]. Let us fix an infinitesimal sequence $(\varepsilon_j)\in]0,\lambda_0[$. The H-convexity of u_{ε_j} for every $j\in\mathbb{N}$ allows us to apply estimates (71) to u_{ε_j} restricted to any compact ball of Ω with suitably small radius. The inequality (69), with K replaced by a compact ball, along with (71) and (72), yields a uniform bound on the L^∞ norm and on the Lipschitz constant of u_{ε_j} restricted to the fixed compact ball. By a standard argument, using Ascoli-Arzelà theorem we get a subsequence $u_j=u_{\varepsilon_{k_j}}$ which uniformly converges to a continuous function $u:\Omega\longrightarrow\mathbb{R}$ on compact sets of Ω. The condition (24) is preserved in the limit and implies the H-convexity of u, hence u is a locally Lipschitz H-convex function. In view of the L^1_{loc} convergence of u_j to u and of the convergence of $y\longrightarrow\int_\Omega\varphi(x)\,\vartheta_{\varepsilon_{k_j}}(y^{-1}x)\,dx$ to φ uniformly on compact sets of Ω, we achieve

$$\begin{aligned}\int_\Omega u_j(x)\,\varphi(x)\,dx&=\int_\Omega\left(\int_\Omega\vartheta_{\varepsilon_{k_j}}(y^{-1}x)\,d\mu(y)\right)\varphi(x)\,dx\\ &=\int_\Omega\left(\int_\Omega\varphi(x)\,\vartheta_{\varepsilon_{k_j}}(y^{-1}x)\,dx\right)\,d\mu(y)\longrightarrow\langle T,\varphi\rangle\quad\text{as}\quad j\to\infty,\end{aligned}$$

hence we have shown that

$$\langle T,\varphi\rangle=\int_\Omega\varphi(y)\,d\mu(y)=\int_\Omega u(y)\,\varphi(y)\,dy$$

for every $\varphi\in C^\infty_c(\Omega)$. This ends the proof. □

6 Aleksandrov-Busemann-Feller theorem

In this section we deal with the existence of pointwise second derivatives of H-convex functions. In the rest of the section, Ω will be assumed to be open and H-convex.

Definition 6.1 **(H-BV^2 function)** Let $u \in L^1(\Omega)$ and let $(X_1, \ldots, m)$ be a basis of V_1. We say that u has *H-bounded second variation* (in symbols H-BV^2) and write $u \in BV_H^2(\Omega)$ if the distributional derivatives $X_i u$, $X_i X_j u$ are finite Radon measures for every $i, j = 1, \ldots, m$. If $u \in L^1_{loc}(\Omega)$ and $X_i u$, $X_i X_j u$ are Radon measures we say that u has *locally H-bounded second variation* (in symbols locally H-BV^2) and write $u \in BV^2_{H,loc}(\Omega)$.

The following result corresponds to Theorem 3.9 of [1].

Theorem 6.2 **(L^1-differentiability)** *Let $u \in BV^2_{H,loc}(\Omega)$. Then for a.e. $x \in \Omega$ there exists a polynomial $P_{[x]} \in \mathcal{P}_{H,2}(\mathbb{G})$ such that*

$$\lim_{r \to 0^+} \frac{1}{r^2} \fint_{B_{x,r}} |u(y) - P_{[x]}(y)|\, dy = 0 \tag{70}$$

The next result has been proved in [9], [23] and [21].

Theorem 6.3 **(L^∞-estimates)** *Let $u : \Omega \longrightarrow \mathbb{R}$ be a continuous H-convex function. Then for every $\xi \in \Omega$ there exists a radius $R > 0$ with $D_{\xi,R} \subset \Omega$ and a constant $C > 0$ depending on ξ such that for every $r < R/15$ the following estimates hold*

$$\sup_{y \in B_{\xi,r}} |u(y)| \le C \fint_{B_{\xi,5r}} |u(y)|\, dy \qquad \text{and} \qquad \|\nabla_H u\|_{L^\infty(B_{\xi,r})} \le \frac{C}{r} \fint_{B_{\xi,15r}} |u(y)|\, dy. \tag{71}$$

Lemma 6.4 *Let $v : \Omega \longrightarrow \mathbb{R}$ be a locally Lipschitz function. Then $\nabla_H u \in L^\infty_{loc}(\Omega)^m$ and for every closed ball $D_{x,3s} \subset \Omega$ and every $z, y \in D_{x,s}$ we have*

$$|v(z) - v(y)| \le \|\nabla_H v\|_{L^\infty(D_{x,3s})}\, \rho(z, y). \tag{72}$$

The validity of this lemma follows by both Theorem 1.3 and Theorem 2.7 of [15], which hold in the more general Carnot-Carathéodory spaces. For the reader's convenience we sketch its proof in the simpler case of stratified groups. Let us fix a basis $(X_1, \ldots, X_m)$ of the first layer $V_1 \subset \mathcal{G}$. The local Lipschitz condition and the a.e. horizontal differentiability of v ensure that $X_j v$ exists in the distributional sense for every $j = 1, \ldots, m$

$$\int_\Omega v(y)\, X_j \varphi(y)\, dy = -\int_\Omega \varphi(y)\, X_j v(y)\, dy$$

where $\varphi \in C^\infty_c(\Omega)$ and $X_j v \in L^\infty_{loc}(\Omega)$. To see this, it suffices to observe that

$$y \longrightarrow \frac{v(y \exp h X_j) - v(y)}{h}$$

is uniformly bounded with respect to $h \in]-\varepsilon, \varepsilon[\setminus\{0\}$ and converges to $X_j v(y)$ for a.e. y, see for instance Theorem 3.2 of [26]. Then we use the weak compactness of a bounded family of functions in $L^p(K_i)$ for some fixed $p > 1$ and all compact sets $K_i \subset \Omega$ such that $\bigcup_{i=1}^\infty K_i = \Omega$. Let $\gamma : [0, T] \longrightarrow \mathbb{R}$ be a subunit curve joining z with y, with $z, y \in D_{x,s}$, i.e. an absolutely continuous function γ such that $\gamma'(t) = \sum_{j=1}^m a_j(t)\, X_j(\gamma(t))$, with $a_j \in L^\infty(0, T)$ and $\sum_{j=1}^m a_j(t)^2 \leq 1$ for a.e. $t \in (0, T)$. By definition of Carnot-Carathéodory distance $\rho(z, y)$ we can find a subunit curve γ with $T < \rho(z, y)+h$, then $\rho(z, \gamma(t)) \leq t \leq T < \rho(z, y) + h$ and $\gamma([0, T]) \subset B_{z,\rho(z,y)+h}$, where $h > 0$ is arbitrarily small. As a consequence, by convolution with a family of smooth kernels ϕ_ε we readily obtain $|v_\varepsilon(z) - v_\varepsilon(y)| < \|\nabla_H v_\varepsilon\|_{L^\infty(B_{z,\rho(z,y)+h})}\, (\rho(z, y) + h)$ where $v_\varepsilon = v * \phi_\varepsilon$. Taking the limit as $h \to 0^+$ we obtain

$$|v_\varepsilon(z) - v_\varepsilon(y)| \leq \|\nabla_H v_\varepsilon\|_{B_{z,\rho(z,y)}}\, \rho(z, y) \leq \|\nabla_H v\|_{L^\infty(B_{z,\rho(z,y)})}\, \rho(z, y). \tag{73}$$

The continuity of v ensures that v_ε uniformly converges to v on compact sets of Ω, then taking the limit in (73) as $\varepsilon \to 0^+$ we have

$$|v(z) - v(y)| \leq \|\nabla_H v\|_{L^\infty(B_{z,\rho(z,y)})}\, \rho(z, y) \leq \|\nabla_H v\|_{L^\infty(D_{x,3s})}\, \rho(z, y).$$

Theorem 6.5 (Aleksandrov-Busemann-Feller) *Let $(X_1, \ldots, X_m)$ be a basis of V_1 and let $u : \Omega \longrightarrow \mathbb{R}$ be a measurable H-convex continuous function such that its distributional derivative $X_i X_j u$ is a Radon measure for every $i, j = 1, \ldots, m$. Then for a.e. $x \in \Omega$ there exists a unique polynomial $P_{[x]} \in \mathcal{P}_{H,2}(\mathbb{G})$ such that the following limit holds*

$$\lim_{y \to x} \frac{|u(y) - P_{[x]}(y)|}{\rho(x, y)^2} = 0. \tag{74}$$

PROOF. From results of [31] the function u is locally bounded above, then Theorem 3.18 implies that u is locally Lipschitz continuous. As a result, Lemma 6.4 implies that $\nabla_H u \in L^\infty_{loc}(\Omega)^m$. Thus, by hypothesis we have that $u \in BV^2_{H,loc}(\Omega)$. In view of Theorem 6.2 for a.e. $x \in \Omega$ there exists $P_{[x]} \in \mathcal{P}_{H,2}(\mathbb{G})$ such that

$$\lim_{r \to 0^+} \frac{1}{r^2} \fint_{B_{x,r}} |u(y) - P_{[x]}(y)|\, dy = 0. \tag{75}$$

Let us fix $x \in \Omega$ satisfying this condition and define the map $v(y) = u(y) - P_{[x]}(y)$. We can write the polynomial $P_{[x]}$ as the sum of $L \in \mathcal{P}_{H,1}(\mathbb{G})$ and $R \in \mathcal{P}_{H,2}(\mathbb{G})$ such that $R(x) = 0$ and $X_j R(x) = 0$ for every $j = 1, 2, \ldots, m$. Notice that L has the form $L(\xi) = c + \sum_{j=1}^m \alpha_j\, \xi_j$, where $c, \alpha_j \in \mathbb{R}$ and ξ_j is a coordinate of homogeneous degree equal to one for every $j = 1, 2, \ldots, m$. It follows that both L and $-L$ are H-convex and that the sum $w = u - L$ is H-convex. Let us write $v = w - R$ and notice that conditions $X_j R \in \mathcal{P}_{H,1}(\mathbb{G})$ and $X_j R(x) = 0$ for every $j = 1, 2, \ldots, m$ give us a constant $C_1 > 0$ such that for every $r > 0$ we have the estimate

$$\sup_{B_{x,r}} |\nabla_H R| \leq C_1\, r \qquad \text{and} \qquad \sup_{B_{x,r}} |R| \leq C_1\, r^2. \tag{76}$$

In view of the gradient estimate (71) applied to the H-convex function w we obtain a number $r_0 > 0$ and a constant $C > 0$ such that the inequality

$$\|\nabla_H v\|_{L^\infty(B_{x,r})} \le C\, r^{-1} \fint_{B_{x,15r}} |w(y)|\, dy + \sup_{B_{x,r}} |\nabla_H R|$$

holds for every $0 < r < r_0$, where $\overline{B}_{x,15r_0} \subset \Omega$. From the previous inequality, we infer that

$$\|\nabla_H v\|_{L^\infty(B_{x,r})} \le C\, r^{-1} \fint_{B_{x,15r}} |v(y)|\, dy + \sup_{B_{x,r}} |\nabla_H R| + C\, r^{-1} \fint_{B_{x,15r}} |R(y)|\, dy.$$

Due to the estimates (76), the previous inequality yields

$$\|\nabla_H v\|_{L^\infty(B_{x,r})} \le C\, r^{-1} \fint_{B_{x,15r}} |v(y)|\, dy + (1 + C)\, C_1\, r. \tag{77}$$

Now we arbitrarily fix $\varepsilon \in]0, 1/2[$ and $\tau \in]0, \varepsilon^Q[$. The limit (75) and the definition of v give the estimate

$$\left| \left\{ y \in B_{x,r} \,\middle|\, |v(y)| \ge \varepsilon\, r^2 \right\} \right| \le\ (\varepsilon\, r^2)^{-1} \int_{B_{x,r}} |v(y)|\, dy = \varepsilon^{-1}\, o(r^Q) \qquad \text{as} \qquad r \to 0^+.$$

Then we can fix $r_1 < r_0$ depending on ε and τ such that

$$\left| \left\{ y \in B_{x,r} \,\middle|\, |v(y)| \ge \varepsilon\, r^2 \right\} \right| < \tau\, |B_{x,r}| \tag{78}$$

for every $0 < r < r_1$. We choose $y \in B_{x,r/2}$ and note that $B_{y,\tau^{1/Q} r} \subset B_{x,r}$, then there exists $z_r \in B_{y,\tau^{1/Q} r}$ such that $|v(z_r)| < \varepsilon\, r^2$ for every $r < r_1$. In fact, if this were not true we would have

$$B_{y,\tau^{1/Q} r} \subset \left\{ y \in B_{x,r} \,\middle|\, |v(y)| \ge \varepsilon\, r^2 \right\}$$

that contradicts the inequality (78). We have proved that for every $r < r_1$ the inequality

$$|v(y)| < \varepsilon r^2 + |v(z_r) - v(y)| \tag{79}$$

holds for every $y \in B_{x,r/2}$ and for some $z_r \in B_{y,\tau^{1/Q} r}$ depending on r and y. In view of (75) and (77) there exists $r_2 < r_1/3$ such that $\|\nabla_H v\|_{L^\infty(B_{x,3r})} \le Cr + 3(1 + C)\, C_1\, r = C_2\, r$ for every $r < r_2$. Thus, from Lemma 6.4 and inequality (79) we conclude for every $y \in B_{x,r/2}$ the final estimates

$$|v(y)| \le \varepsilon r^2 + C_2\, r\, \rho(z_r, y) < \varepsilon r^2 + C_2\, \tau^{1/Q} r^2 < \varepsilon\, (1 + C_2)\, r^2 \tag{80}$$

where in the last inequality we have used our initial choice $\tau^{1/Q} < \varepsilon$ and C_2 is a geometrical constant. This ends the proof. □

Proposition 6.6 *Let $u : \Omega \longrightarrow \mathbb{R}$ be a locally summable H-convex function. Then there exists a locally Lipschitz H-convex function $v : \Omega \longrightarrow \mathbb{R}$ such that $v(x) = u(x)$ for a.e. $x \in \Omega$.*

PROOF. We fix $\vartheta \in C_c^\infty(B_1)$ such that $\vartheta \geq 0$ and $\int_{B_1} \vartheta = 1$ and we define $\vartheta_\varepsilon(y) = \varepsilon^{-Q}\vartheta(\delta_{1/\varepsilon}y)$ for every $y \in \mathbb{G}$. For every couple of relatively compact sets $\Omega' \Subset \Omega'' \Subset \Omega$ we can find $\kappa > 0$ such that $\max_{x\in\overline{\Omega'}} \rho(y^{-1}x, x) < \mathrm{dist}(\Omega', (\Omega'')^c)$ whenever $\rho(y) \leq \kappa$. Then the convolution

$$u_\varepsilon(x) = \vartheta\varepsilon * u(x) = \int_\Omega \vartheta_\varepsilon(y)\, u(y^{-1}x)\, dy$$

is well defined, smooth and H-convex on Ω' for every $\varepsilon \leq \kappa$ in that $y^{-1}x \in \Omega'' \subset \Omega$ for every $x \in \Omega'$ and every $y \in B_\varepsilon$. For every $B_{\xi,15r} \subset \Omega'$ we also have the uniform estimate

$$\int_{B_{\xi,15r}} |u_\varepsilon(x)|\, dx \leq \int_{\Omega''} |u(y)|\, dy = C_0 < \infty \tag{81}$$

for every $\varepsilon \leq \kappa$. The H-convexity of u_ε allows us to apply estimates (71), obtaining

$$\sup_{y\in B_{\xi,r}} |u_\varepsilon(y)| \leq \frac{C\,C_0}{|B_{\xi,5r}|} \qquad \text{and} \qquad \|\nabla_H u_\varepsilon\|_{L^\infty(B_{\xi,r})} \leq \frac{C\,C_0}{|B_{\xi,15r}|} \tag{82}$$

whenever, $B_{\xi,15r} \subset \Omega'$. Then by standard arguments, using (72) and the Ascoli-Arzelá compactness theorem, we can find a sequence u_{ε_j} converging on compact subsets of Ω to a locally Lipschitz continuous H-convex function $v : \Omega \longrightarrow \mathbb{R}$. In fact, the pointwise convergence preserves the H-convexity and the Lipschitz property on compact subsets, then the proof is finished. □

Remark 6.7 Note that in Proposition 6.6 we are referring to an individual function u and not to the equivalence class of functions which differ from u on a set of measure zero. Then it makes sense to assume that u satisfies the pointwise notion of Definition 3.4 for a.e. point. Proposition 6.6 implies the validity of Theorem 6.5 for locally summable H-convex functions, after a suitable redefinition on a set of measure zero.

From results of [16], [10] and [31], Theorem 6.5 holds for H-convex functions on two step stratified groups, with no additional assumptions on the function.

References

[1] L.AMBROSIO, V.MAGNANI, *Weak differentiability of BV functions on stratified groups*, Math. Z., **245**, 123-153, (2003)

[2] L.AMBROSIO, N.FUSCO, D.PALLARA, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford University Press, (2000).

[3] I.Ya.Bakel'man, *Geometric methods of solutions of elliptic equations*, (in Russian), Nauka, Moscow, (1965)

[4] Z.Balogh, M.Rickly, *Regularity of convex functions on Heisenberg groups*, preprint (2003)

[5] A.Bellaïche, J.J.Risler eds, *Sub-Riemannian geometry*, Progress in Mathematics, **144**, Birkhäuser Verlag, Basel, 1996.

[6] G.Bianchi, A.Colesanti, C.Pucci, *On the second order differentiability of convex surfaces*, Geom. Dedicata, **60**, 39-48, (1996)

[7] T.Bieske, *On ∞-harmonic functions on the Heisenberg group*, Comm. in Partial Differential Equations, **27**, n.3,4, 727-762, (2002)

[8] L.A.Caffarelli, X.Cabré, *Fully nonlinear elliptic equations*, AMS Colloquium Publications, **43**, AMS, Providence, RI, (1995)

[9] D.Danielli, N.Garofalo, D.M. Nhieu, *Notions of convexity in Carnot groups*, Comm. Anal. Geom., **11**, n.2, 263-341, (2003)

[10] D.Danielli, N.Garofalo, D.M. Nhieu, F.Tournier *The theorem of Busemann-Feller-Alexandrov in Carnot groups*, Comm. Anal. Geom. **12**, n.4, 853-886, (2004)

[11] R.M.Dudley, *On second derivatives of convex functions*, Math. Scand. **41**, n.1, 159-174, (1977)

[12] L.C.Evans, R.F.Gariepy, *Measure theory and fine properties of functions*, Studies in Advanced Mathematics, CRC Press, Boca Raton, (1992)

[13] H.Federer, *Geometric Measure Theory*, Springer, (1969)

[14] G.B.Folland, E.M. Stein, *Hardy Spaces on Homogeneous groups*, Princeton University Press, (1982)

[15] N.Garofalo, D.M.Nhieu, *Lipschitz continuity, global smooth approximation and extension theorems for Sobolev functions in Carnot-Carathéodory spaces*, Jour. Anal. Math., **74**, 67-97 (1998)

[16] C.E. Gutirrez, A. Montanari, *On the second order derivatives of convex functions on the Heisenberg group*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **3**, n.2, 349-366, (2004)

[17] M.Gromov, *Carnot-Carathéodory spaces seen from within*, Subriemannian Geometry, Progress in Mathematics, **144**. ed. by A.Bellaiche and J.Risler, Birkhauser Verlag, Basel, 1996.

[18] P.Hajlasz, P.Koskela, *Sobolev met Poincaré*, Mem. Amer. Math. Soc. **145**, (2000)

[19] G.Hochschild, *The structure of Lie groups*, Holden-Day, (1965)

[20] L.Hörmander, *The Analysis of linear partial differential operators*, Springer-Verlag, (1990)

[21] P.Juutinen, G.Lu, J.Manfredi, B.Stroffolini, *Convex functions in Carnot groups*, preprint (2005)

[22] N.V.Krylov, *Nonlinear elliptic and parabolic equations of the second order*, Mathematics and its Applications, Reidel, (1987)

[23] G.Lu, J.Manfredi, B.Stroffolini, *Convex functions on the Heisenberg group*, Calc. Var. Partial Differential Equations, **19**, n.1, 1-22, (2004)

[24] G.Lu, J.Manfredi, B.Stroffolini, *Convex functions in Carnot groups*, forthcoming

[25] V.Magnani, *Differentiability from Sobolev-Poincaré inequality*, Studia Math. **168**, n.3, p.251-272, (2005)

[26] R.Monti, F.Serra Cassano, *Surface measures in Carnot-Carathédory spaces*, Calc. Var., **13**, n.3, 339-376, (2001)

[27] A.Nagel, E.M.Stein, S.Wainger, *Balls and metrics defined by vector fields I: Basic Properties*, Acta Math., **155**, 103-147, (1985)

[28] P.Pansu, *Métriques de Carnot-Carathéodory quasiisométries des espaces symétriques de rang un*, Ann. Math., **129**, 1-60 (1989)

[29] Yu.G.Reshetnyak, *Generalized derivatives and differentiability almost everywhere*, Mat. Sb. **75**, (1968), 323-334 (in Russian), Math. USSR-Sb. **4**, 293-302 (English translation)

[30] M.Rickly, *On questions of existence and regularity related to notions of convexity in Carnot groups*, Ph.D. thesis (2005)

[31] M.Rickly, *First order regularity of convex functions on Carnot groups*, preprint (2005)

[32] E.M.Stein, *Harmonic Analysis*, Princeton Univ. Press, (1993)

[33] N.Trudinger, *On Hessian measures for non-commuting vector fields*, ArXiv preprint: math.AP/0503695, (2005)

[34] V.S.Varadarajan, *Lie groups, Lie algebras and their representation*, Springer-Verlag, New York, 1984.

[35] N.Th.Varopoulos, L.Saloff-Coste, T.Coulhon, *Analysis and Geometry on Groups*, Cambridge University Press, Cambridge, (1992)

[36] C.Wang, *Viscosity convex functions on Carnot groups*, **133**, n.4, 1247-1253, (2005)

[37] C.Wang, *The Euler Equations of absolutely minimizing Lipschitz extensions for vector fields satisfying Hörmander condition*, preprint (2005)