The Heisenberg group and Pansu's Theorem

July 31, 2009

Abstract

The goal of these notes is to introduce the reader to the Heisenberg group with its Carnot-Carathéodory metric and to Pansu's differentiation theorem. As they are very similar, we will first study Rademacher's theorem about Lipschitz maps and then see how the same technique can be applied in the more complex setting of the Heisenberg group.

Contents

1 The Heisenberg group **2**
1.1 Continuous Heisenberg group 2
1.1.1 Using the tools from differential geometry 2
1.1.2 Naive definition 3
1.1.3 Dilatations and Derivatives 4
1.2 Discrete Heisenberg group 4

2 Bilipschitz maps **5**
2.1 Rademacher's theorem 5
2.2 Pansu's theorem 9

3 Proof of Pansu's theorem **10**

1 The Heisenberg group

1.1 Continuous Heisenberg group

[1.A] Definition ***(Real Heisenberg group)***

The Heisenberg group $H_3(\mathbb{R})$ is the group of 3×3 upper triangular matrices of the form $\begin{pmatrix} 1 & X & Z \\ 0 & 1 & Y \\ 0 & 0 & 1 \end{pmatrix}$*, where $X, Y, Z \in \mathbb{R}$.*

The Heisenberg group is mainly studied for the strange and remarkable properties of a particular metric, called Carnot-Carathéodory or subriemannian metric. The idea behind this metric is fairly simple, but slightly counterintuitive : a classical (Riemannian) metric is defined as the length of the minimal path between two given points, over all possible paths on the manifold, or here on the group. The CC metric will be defined almost the same way, but the infimum will not be taken over all the possible paths, but only over some special paths, which will be chosen tangent to a given distribution. A rigorous definition of this metric requires tools from differential geometry, with which the reader might not be familiar. In this case, he should straight away jump to the 'naive' definition, which should give him an idea of how it works without asphyxiating him with too many new concepts.

1.1.1 Using the tools from differential geometry

We can first notice that the first definition of the Heisenberg group is equivalent to the following one

[1.B] Definition ***(Real Heisenberg group 2)***

The Heisenberg group is the simply connected Lie group whose Lie algebra is $\mathbb{R}^3(\xi, \eta, \zeta)$ with $[\xi, \eta] = \zeta$, all the other brackets being zero.

ξ and η will span what we will call the horizontal distribution, and the idea of the Carnot-Carathéodory metric is to just consider curves tangent to this distribution, which are called horizontal curves.

[1.C] Definition ***(Carnot-Carathéodory metric)***

Let $\mathcal{H}$ be the distribution (i.e. a vector subbundle $\subseteq TH_3$ of the tangent bundle of H_3) spanned by left translation of $\{\xi, \eta\}$. A curve on H_3 is called horizontal if it is tangent to $\mathcal{H}$. An arbitrary norm on TH_3 being chosen, the CC metric on H_3 is then defined by, for $p, q \in H_3$ $d_{CC}(p, q) = \inf\{$length of the smooth horizontal curves that connect A and $B\}$, this distance being infinite if there is no such curve.

This definition needs an explanation. It may seem counterintuitive to try to connect two points with only horizontal curves. The same definition on $\mathbb{R}^3$ would give no results : the distance between two points would be the Euclidean one if they're both on the same horizontal plane, and infinite if they're not. The main difference is that, unlike the $\mathbb{R}^3$ case, the horizontal distribution we have defined on H_3 is clearly not a foliation. Recall that the Frobenius theorem states that a subbundle $\mathcal{H}$ of the tangent bundle is integrable if, for any two vector fields X, Y on $\mathcal{H}$, the Lie bracket $[X, Y]$ takes values in $\mathcal{H}$ as well. This is obviously not the case here, and the behavior of the Heisenberg group is best described by the following theorem by Chow, which is the starting point of Carnot Carathéodory geometry

[1.D] Definition ***(Bracket generating distribution)***

Given a collection $\{X_a\}$ of vector fields, form its Lie Hull, the collection of all vector fields $\{X_a, [X_b, X_c], [X_a, [X_b, X_c]]...\}$ generated by Lie brackets of the X_a. We say that the collection $\{X_a\}$ is bracket generating if this Lie hull spans the whole tangent bundle. A distribution $H \subseteq TQ$ is called bracket generating if any local horizontal frame $\{X_a\}$ for the distribution is bracket generating.

[1.E] Theorem ***(Chow)***

If a distribution $\mathcal{H} \subseteq TQ$ is bracket generating then the set of points that can be connected to $A \in Q$ by a horizontal path is the component of Q containing A.

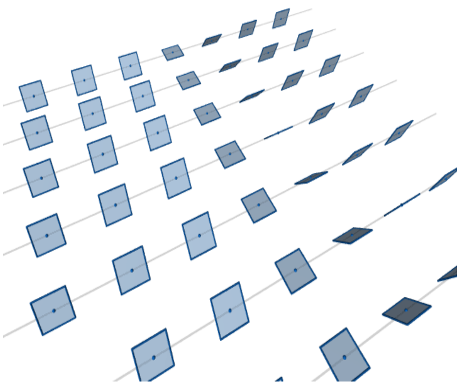
As the horizontal distribution we have defined on the Heisenberg group is clearly bracket generating, this theorem explains why on this particular example, the Carnot Carathéodory metric is always finite.

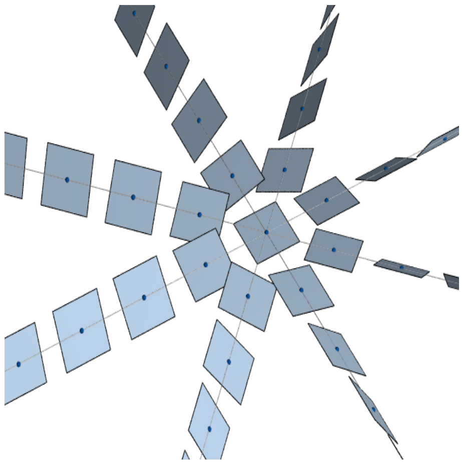
1.1.2 Naive definition

We will see how the CC metric works by studying it in comparison with the usual Euclidean metric. In the usual $\mathbb{R}^3$ space, one can always connect two distinct points with the straight line joining them. However, if one tries to connect two arbitrary points with horizontal lines, one is doomed to fail, as the z-coordinate stays constant on horizontal lines. This is because the $\mathbb{R}^3$ group law is extremely regular and symetric : it is $(x, y, z), (x', y', z') \mapsto (x + x', y + y', z + z')$. Hence, the horizontal plane, which is spanned by $((1, 0, 0), (0, 1, 0))$, is extremely regular too, and matches the intuitive idea of a plane.

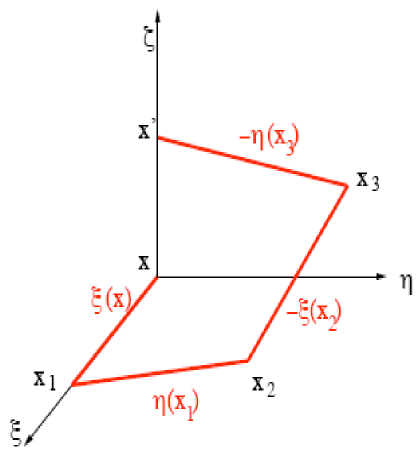
Let us see what happens with the Heisenberg group now. The main difference is that the group law is not as regular, and is twisted : $((X, Y, Z), (X', Y', Z')) \mapsto (X + X', Y + Y', Z' + Z + XY')$; in particular, if $X, Y' \neq 0$, $(X, Y, 0) + (X', Y', 0)$ has a nonzero z-coordinate. If we define the horizontal distribution (we use this term rather than plane, because it is not one) as the one spanned by $((1, 0, 0), (0, 1, 0))$, this shows why you can connect any arbitrary pair of points using only horizontal curves, i.e. curves tangent to the horizontal distribution ! The CC metric is then defined fairly easily, as the length of the minimal horizontal path between two points.

The sceptic reader might appreciate some pictures





These pictures, courtesy of Patrick Massot, show what the horizontal distribution looks like, in cartesian and cylindrical coordinates. Horizontal curves are curves which always stay tangent to the little planes drawn on the picture. The 'twist' can be easily seen on both pictures : the horizontal planes are not horizontal. This is why horizontal curves can reach the entire group. To illustrate this, the last picture, courtesy of Pierre Pansu, shows how to connect $x = (0, 0, 0)$ and $x' = (0, 0, z)$ with horizontal curves.



A good way to familiarize oneself with the CC metric is to compare the first picture with this one to check that the used curves are indeed horizontal.

1.1.3 Dilatations and Derivatives

The Heisenberg group can be provided with a dilatation δ_t, which is the analog of a classical homothetic transformation $\delta_\lambda : x \mapsto \lambda x$ in the Euclidean $\mathbb{R}^n$ space. We define $\delta_t(X,Y,Z) = (tX, tY, t^2Z)$, so that $[\delta_t(\xi), \delta_t(\eta)] = \delta_t(\zeta)$ and $d_{CC}(\delta_t(x), \delta_t(y)) = t(d_{CC}(x,y))$. This allows us to define a derivative on the Heisenberg group : as the classical derivative at 0 is defined by $\lim_{t\to 0} \frac{f(tx)}{t} = \lim_{\lambda\to 0} (\delta_\lambda)^{-1}(f(\delta_\lambda x))$, one can define a derivative on the Heisenberg group : $dF(g)(h) = \lim_{t\to 0} (\delta_t^{-1})((F(g))^{-1}F(g\delta_t h))$. As in the Euclidean case, where we expect the derivative to be linear, Pansu's derivative needs to be an homomorphism. Pansu's theorem, which will be proved in the third part and is the main result of this lecture, is that Lipschitz maps (for the Carnot-Carathéodory metric) are differentiable almost everywhere on the Heisenberg group.

1.2 Discrete Heisenberg group

We now introduce the discrete Heisenberg group, which is the same as the real one except that the entries are now taken in $\mathbb{Z}$.

[1.F] Definition ***(Discrete Heisenberg group)***

The Heisenberg group $H_3(\mathbb{Z})$ is the group of 3×3 upper triangular matrices of the form $\begin{pmatrix} 1 & X & Z \\ 0 & 1 & Y \\ 0 & 0 & 1 \end{pmatrix}$, *where $X, Y, Z \in \mathbb{Z}$.*

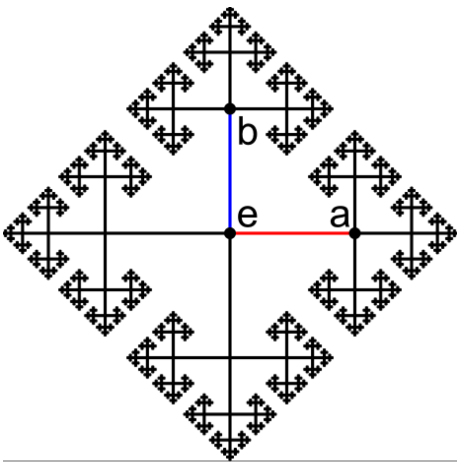
From its definition, it is easily seen that the discrete Heisenberg group is generated by $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$. The main tool to study metric properties of discrete groups finitely generated is called the Cayley graph.

[1.G] Definition ***(Cayley Graph)***

Suppose that G is a group and S is a generating set. The Cayley graph $\Gamma = \Gamma(G, S)$ is a colored directed graph constructed as follows.

- *Each element g of G is assigned a vertex : the vertex set $V(\Gamma)$ of Γ is identified with G.*
- *Each generator s of S is assigned a color c_s.*
- *For any $g \in G$, $s \in S$, the vertices corresponding to the elements g and gs are joined by a directed edge of colour c_s. Thus the edge set $E(\Gamma)$ consists of pairs of the form (g, gs), with $s \in S$ providing the color.*

To see what a Cayley graph looks like, here is the example of the Cayley graph of the free group with two generators.



A graph is always naturally equipped with a distance : the graph distance which is the distance of the shortest path between two edges. A remarkable fact about the Heisenberg group is that if the norm on the tangent bundle is the ℓ_1 norm, the distances defined by $d_{CC|H_3(\mathbb{Z})}$ and the graph distance are the same. This follows from the following observation, which is fundamental in the geometry of the Carnot-Carathéodory metric : geodesics for the Carnot-Carathéodory metric correspond to solutions of Dido's problem[1] in $\mathbb{R}^2$. Let us explain what this means : It is easily shown that for an arbitrary curve on the Heisenberg group, being tangent to the horizontal distribution is equivalent to satisfying $dz - xdy = 0$. Hence we have $z(t) = \int_0^t x(s)y'(s)ds$ which, by Green-Riemann (or Stokes)'s formula is, up to a constant, exactly the area in $\mathbb{R}^2$ between the curve and the y axis. As the solutions to Dido's problem in $\mathbb{R}^2$ for the ℓ_1 norm are well known (these are arcs of squares with sides parallel to the axes), it follows easily that the graph and CC distances are exactly the same.

2 Bilipschitz maps

2.1 Rademacher's theorem

In this section we will prove the following theorem, initially stated by Rademacher, which is the analog of Pansu's theorem in the usual $\mathbb{R}^n$ Euclidean space.

[1]Dido's problem is the following variant of the isoperimetric problem : you are given a segment $[AB]$ and a constant L, and you need to find the curve with length L between A and B which maximizes the area between the segment and the curve. If the norm is Euclidean, the solution is obviously a circle

[2.A] Theorem *(Rademacher)*

> *Let $n, m \in (\mathbb{N}^*)^2$ and $f : \mathbb{R}^n \to \mathbb{R}^m$ a C-Lipschitz map. Then f is differentiable almost everywhere.*

The proof of this theorem is split into two parts. The first one shows that one can suppose $m = 1$, and the second part is the proof of the case $n = 1$, for which there is a more general theorem, first stated by Lebesgue, which applies here because Lipschitz maps have bounded variation. There are numerous proofs for both those results, and we have tried to find the easiest ones, these come from [GT98]. However, the proof of $[2.A]$ which we present here is pretty hard to generalize to Heisenberg groups, this is why another one will be used in the third part, and for the course itself. We chose to keep this one in the notes because it is the most direct one and it gives a good grasp on how this works, while everything is kept hidden behind Lusin and Egorov theorems in the other one.

[2.B] Theorem *(Lebesgue)*

> *Let $f : [a, b] \to \mathbb{R}$ be a bounded variation map. Then f is differentiable almost everywhere.*

Proof of [2.A].

Let us prove the first part of the theorem, i.e. prove it under the assumption that it is true if $n = 1$. One can suppose that $m = 1$, and that's what we do, and we write $<,>$ for the Euclidean scalar product, and $\mathcal{S}$ for the corresponding unit sphere. Two facts can be immediately deduced from this.

- $\nabla f(x)$ exists for almost all $x \in \mathbb{R}^n$.
- If $e \in \mathcal{S}$, $\frac{f(x+te)-f(x)}{t}$ has, for almost all $x \in \mathbb{R}^n$, a limit when $t \to 0$, that we will name $L_x(e)$.

We claim that for a fixed $e \in \mathcal{S}$, for almost all $x \in \mathbb{R}^n$, $L_x(e) =< \nabla f(x), e >$. Indeed, we will show that for every map φ $\mathcal{C}^\infty$ with compact support, $\int_{\mathbb{R}^n} (L_x(e)- < \nabla f(x), e >)\varphi(x)dx = 0$, which proves the announced result. The dominated convergence theorem gives easily, as φ has compact support,

$$\int_{\mathbb{R}^n} \frac{f(x+te)-f(x)}{t}\varphi(x)dx \to_{t\to 0} \int_{\mathbb{R}^n} L_x(e)\varphi(x)dx \tag{2.i}$$

We also have

$$\int_{\mathbb{R}^n} \frac{f(x+te)-f(x)}{t}\varphi(x)dx = \int_{\mathbb{R}^n} \frac{\varphi(x-te)-\varphi(x)}{t} f(x)dx \tag{2.ii}$$

The dominated convergence theorem shows that this tends to $\int_{\mathbb{R}^n} < \nabla\varphi(x), e > f(x)dx$ as $t \to 0$. By integration by parts, we also have (legitimate because the partial derivatives are defined almost everywhere).

$$\begin{aligned}
\int_{\mathbb{R}^n} < \nabla\varphi(x), e > f(x)dx &= \sum_{i=1}^n e_i \int_{\mathbb{R}^n} \frac{\partial\varphi}{\partial x_i} f(x)dx \\
&= -\sum_{i=1}^n e_i \int_{\mathbb{R}^n} \frac{\partial f}{\partial x_i}(x)\varphi(x)dx \\
&= -\int_{\mathbb{R}^n} < \nabla f(x), e > \varphi(x)dx
\end{aligned} \tag{2.iii}$$

And $2.i, 2.ii, 2.iii$ show that :

$$\int_{\mathbb{R}^n} (L_x(e)- < \nabla f(x), e >)\varphi(x)dx = 0$$

which is what we wanted.

Let $\mathcal{A} = \{x \in \mathbb{R}^n, \nabla f(x)\text{exists}\}$. We know that $\mu(\mathbb{R}^n \backslash A) = 0$, and for a fixed $e \in \mathcal{S}$, $\forall x \in \mathcal{A}$,

$$\frac{f(x+te)-f(x)}{t} \rightarrow_{t\rightarrow 0} < \nabla f(x), e > \tag{2.iv}$$

We still have to show that $\forall x \in \mathcal{A}$, $2.iv$ is true for all $e \in \mathcal{S}$, and this uniformly in e. To prove this, let $(e_i)_{i\geq 1}$ be a dense sequence in $\mathcal{S}$, and let

$$B = \{x \in A, \forall i \geq 1, \frac{f(x+te_i)-f(x)}{t} \rightarrow_{t\rightarrow 0} < \nabla f(x), e_i >\}$$

We have $\mu(\mathbb{R}^n \backslash B) = 0$. But if $x \in B$, the maps $\Phi_t : e \in \mathcal{S} \mapsto \frac{f(x+te)-f(x)}{t}$ are all C-Lipschitz. Hence the convergence $\Phi_t(e_i) \rightarrow < \nabla f(x), e_i >$ shows the uniform convergence of Φ_t to $< \nabla f(x), e >$, and thus the differentiability almost everywhere of f.

Before proving Lebesgue's theorem, we first have to recall some properties of bounded variation functions.

[2.C] Definition ***(Bounded variation functions)***

A function $f : I \rightarrow \mathbb{R}$ is said to have bounded variation if it satisfies one of the following equivalent conditions :

1. *The total variation of f defined by*

$$V_{a,b}(f) = \sup\{\sum_{i=1}^{n-1} |f(x_{i+1}) - f(x_i)|, a \leq x_1 < x_2 < .. < x_n \leq b\}$$

   is finite.

2. *The graph of f is rectifiable.*

3. *f is the the difference of two nondecreasing functions.*

The equivalence between 1. and 2. is trivial, 3. $\Rightarrow$ 1. is clear, as a nondecreasing function has bounded variation, 1. $\Rightarrow$ 3. is easy if we notice that $x \mapsto V_{a,x}(f)$ is nondecreasing on $[a, b]$, and $x \mapsto V_{a,x}(f) - f(x)$ too. Condition 3. shows that such a function is continuous except on a countable set.

We will first need the following lemma, which is a weak version of Vitali's lemma.

[2.D] Lemma

Let $(I_n =]x_n - r_n, x_n + r_n[)_{1\leq n\leq N}$ be N intervals of $\mathbb{R}$. Then there exists a subset $J \subseteq [|1, N|]$ such that the $(I_j)_{j\in J}$ are pairwise disjoint and

$$\bigcup_{n=1}^{N} I_n \subseteq \bigcup_{j\in J}]x_j - 3r_j, x_j + 3r_j[$$

which implies

$$\mu(\bigcup_{n=1}^{N} I_n) \leq 3 \sum_{j\in J} \mu(I_j)$$

Proof of [2.D].

Let us sort the r_n in a decreasing sequence $R_1..R_k$. Take then $J_1 \subseteq [|1, N|]$ maximal such that the $(I_j)_{j\in J_1}$ don't intersect and have radius R_1. Then $J_2 \subseteq [|1, N|]$ maximal such that the $(I_j)_{j\in J_1\cap J_2}$ don't intersect and such that the $(I_j)_{j\in J_2}$ have radius R_2. Go on until J_k is defined, and then define $J = \bigcup_{i=1}^{k} J_i$. Then J verifies the given properties : the I_j are pairwise disjoint by definition, and the other property follows immediately from the following easy fact : If $]x-r, x+r[\cap]y-s, y+s[\neq \emptyset$ with $r \leq s$, then $]x-r, x+r[\subseteq]y-3s, y+3s[$.

We now have all the tools needed to prove Lebesgue's theorem.

Proof of [2.B].

We will first need some notations. Write D for the the set (at most countable) of discontinuity points of f. For $x \in]a,b[\backslash D$, write

$$f^+(x) = \limsup_{y \to x} \frac{f(y)-f(x)}{y-x}, f^-(x) = \liminf_{y \to x} \frac{f(y)-f(x)}{y-x}$$

and define the following sets :

$$A^+ = \{x \in]a,b[\backslash D, f^+(x) = +\infty\}, A^- = \{x \in]a,b[\backslash D, f^-(x) = -\infty\}$$

$$B = \{x \in]a,b[\backslash D, f^+(x) > f^-(x)\}$$

We will show that $\mu(A^+) = \mu(A^-) = \mu(B) = 0$, which will prove Lebesgue's theorem.

Let us begin with A^+ and A^-. Suppose ad absurdum that $\mu(A^+) > 0$ and let K be an arbitrary big constant. For $x \in A^+$, there exists $I_x =]a_x, b_x[$ included in $]a,b[$ and containing x such that $f(b_x) - f(a_x) \geq K(b_x - a_x)$. $\mathbb{R}$ being separable, one can extract a countable covering $(I_{x_n})_{n \in \mathbb{N}}$ out of I_x, and there exist $N \geq 0$ such that

$$\mu(\bigcup_{n=0}^{N} I_{x_n}) \geq \frac{1}{2}\mu(\bigcup_{n \in \mathbb{N}} I_{x_n})$$

By applying the lemma to this family, we now have a finite set F such that :

1. The $(I_x)_{x \in F}$ are pairwise disjoint.
2. $\sum_{x \in F} \mu(I_x) \geq \frac{1}{6}\mu(\bigcup_{x \in]a,b[})$.

Hence

$$\sum_{x \in F} (f(b_x) - f(a_x)) \geq \frac{K}{6}\mu(A^+)$$

and $V(f) \geq \frac{K}{6}\mu(A^+)$. K being arbitrary, this contradicts the fact that f has bounded variation.

Suppose now that $\mu(B) > 0$. Then there exist $(\alpha, \beta) \in \mathbb{Q}^{+*} \times \mathbb{Q}$ such that the set

$$C = \{x \in B, f^+(x) > \beta + \alpha \text{and} f^-(x) < \beta - \alpha\}$$

has a positive measure (because $\mathbb{Q}$ is countable, dense in $\mathbb{R}$, and a measure is countably additive).

By adding an affine function, one can suppose that $\beta = 0$

Here begins the technical part of the proof. For a bounded variation map g, write $l(g)$ for the length of its graph. Let F be an arbitrary finite subset in $[a,b]$ containing a and b and a_F the function interpolating affinely f at the points of F. We will build another finite subset $G \subseteq [a,b]$ such that

$$l(a_G) \geq l(a_F) + \frac{\mu(C)}{6}(\sqrt{1+\alpha^2} - 1)$$

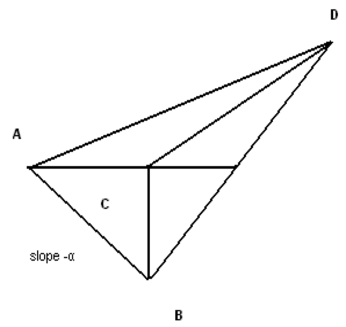
This will be absurd because by iterating the same process with G instead of F and so on, this shows that the graph of f is not rectifiable. Let us build this subset G.

For $x \in C\backslash F$, choose $I_x =]a_x, b_x[$ included in $]a,b[$ and containing x such that $I_x \cap F = \emptyset$ and

$$\begin{cases} f(b_x) - f(a_x) \leq -\alpha(b_x - a_x) & \text{if } a_F \text{ is nondecreasing on } I_x. \\ f(b_x) - f(a_x) \geq \alpha(b_x - a_x) & \text{otherwise.} \end{cases}$$

The same technique as the one used for A^+ gives us a finite subset E of $C\backslash F$ such that the $(I_x)_{x \in E}$ are pairwise disjoint and $\displaystyle\sum_{x \in E} \mu(I_x) \geq \frac{\mu(C)}{6}$. We now just have to show that $G = E \cup F$ verifies the announced inequality.

Instead of a complicated rigorous proof, we will use the following picture, corresponding to the case where a_F is always nondecreasing, the reader will have no trouble considering the general case which works the same way.



With the notations of the picture, AD is the graph of a_F, ABD is the graph of a_G, and we have $l(a_G) \geq AB + BD$, thus

$$\begin{aligned} l(a_F) = AD &\leq AC + CD = (AC - AB) + (AB + CD) \\ &\leq (AC - AB) + (AB + BD) \leq AC - AB + l(a_G) \end{aligned}$$

But

$$AC - AB = (1 - \sqrt{1+\alpha^2})AC = (1 - \sqrt{1+\alpha^2})\mu(C)$$

This gives exactly the result we announced.

2.2 Pansu's theorem

Let us state Pansu's theorem [Pan89], which is our main result. It has a more general setting with the Carnot groups, but for the sake of simplicity we will only use see its special case with the Heisenberg group.

[2.E] Theorem *(Pansu's theorem)*

Let f be a Lipschitz map from the Heisenberg group to another one or the Euclidean $\mathbb{R}^n$ space. Then f is differentiable almost everywhere.

The proof of this theorem will be delayed until the 3rd part, but we can already state a relatively easy, but essential, corollary, initially stated in [Sem96].

[2.F] Corollary *(Semmes)*

A map $f : (H_3, d_{CC}) \to (\mathbb{R}^n, d_{eucl})$ is never bilipschitz.

Proof of [2.F].

If f was bilipschitz, it would be differentiable almost everywhere, and as the definition of the derivative scales in the natural way, the differential is bilipschitz too. This gives a contradiction, because any homomorphism from the Heisenberg group to the Euclidean $\mathbb{R}^n$ must have a kernel which is at least 1-dimensional (all commutators must be mapped to 0) and hence cannot be bilipschitz.

Such a corollary shows that there is a fundamental interplay between differentiation theorems and bilipschitz embeddings, which can be used in both directions. Indeed, to see if we can find differentation theorems about maps between a starting space X and an arbitrary space Y, a good idea is to look for bilipschitz maps : their existence will prove that differentiation fails for such metric spaces. A good example of this technique is the case $Y = L^\infty(X)$. In this case, there is the classical Kuratowsky embedding : $x \mapsto d(x, \bullet) - d(x_0, \bullet)$ where x_0 is a basepoint. It is easily seen that it is an isometry, and that shows that there cannot be any relevant differentiation theorem from an arbitrary space (e.g. H_3) and $L^\infty(H)$. Another classical example is the case $X = [0;1]$ and $Y = L^1(\mathbb{R})$: one can use the 'moving indicator' embedding : $f : t \mapsto \chi_{[0;t]}$ where $\chi_{[0;t]}$ is the indicator function of $[0;t]$. It is easily shown that this embedding is an isometry, and one can even show directly that this embedding is not differentiable almost everywhere, because the difference quotient tends to a Dirac δ_t, which is not in L^1.

This continuous group provides therefore a good example of a metric space which is hard to embed, but it is a very abstract one which doesn't apply to the theoretical computer science problematics we are interested in : embedding algorithms such as the one from Linial, London and Rabinovich are used on graphs, which are discrete and finite metric spaces, and not on continuous Lie groups. However, the following corollary shows that the same conclusion holds for the discrete Heisenberg group, giving a much handier and visual example of hard-to-embed metric spaces.

[2.G] Corollary ***(Discrete version)***

A map $f : (H_3(\mathbb{Z}), d_{graph}) \to (\mathbb{R}^n, d_{eucl})$ is never bilipschitz.

Proof of [2.G].

Let $f : H_3(\mathbb{Z}) \to R^n$ be a L-bilipschitz map and write δ_t for the dilatation on H_3 and the homothetic dilatation on $\mathbb{R}^n$. Define $f_t = \delta_t \circ f \circ \delta_{\frac{1}{t}} : \delta_t(H_3(\mathbb{Z})) \to \mathbb{R}^n$, it is clearly a family of L-bilipschitz maps, and the starting space can be trivially embedded in $H_3(\mathbb{R})$. As the L constant is uniform, the family is equicontinuous, and a variant of Ascoli's theorem (the usual theorem cannot be applied here as the domains change with the maps, but the same demonstration can be applied to our setting) shows that there exists a subsequence of this family which converges to $f_0 : H_3(\mathbb{R}) \to R^n$. f_0 being L-bilipschitz too. As we have seen, this is absurd, and therefore the first map f cannot exist.

The last proof is relatively easy but its main drawback is that it can't be applied when the end space is not locally compact (because Ascoli's theorem can not be applied in such spaces), which will be the case in the other courses (for example L^p and especially L^1 spaces). However, there is another construction, based on ultralimits and asymptotic cones, which yields the same result : one can associate a map on the real group to a map on the discrete group, with the same Lipschitz constant. Therefore, any result on the real Heisenberg group, which is easier to manipulate due to the numerous tools we have for Lie groups, can be extended to the discrete Heisenberg group.

3 Proof of Pansu's theorem

As in the proof of Rademacher's theorem, the proof is split in two parts. The first step is to reduce the problem to a one dimensional problem, the second one is to prove the differentiability almost everywhere of rectifiable curves. However, the second part is not needed in this course. Indeed, using only the first one and Lebesgue's theorem, we shows that every Lipschitz map $f : H_3(\mathbb{R}) \to \mathbb{R}^n$ is differentiable almost everywhere, which is the result we need. The second part is useful to consider the case where the range is another Heisenberg group or a Carnot group (i.e. a group on which one can define a Carnot-Carathéodory metric the same way that we did for the Heisenberg group), but it is of no use for computer science problems, and we will therefore omit this part.

[3.A] Proposition

Let $f : O \to \mathbb{R}^n$ be a Lipschitz map where O is an open subset of H. Let $a, b \in H$ be such that for almost all $x \in H$, the limits

$$dF(x)(a) = \lim_{t\to 0} \delta_t^{-1}(F(x))^{-1}F(x\delta_t a)$$

$$dF(x)(b) = \lim_{t\to 0} \delta_t^{-1}(F(x))^{-1}F(x\delta_t b)$$

exist. Then for almost all c of the form $x = \delta_u(a)\delta_{u'}(b)$ and for almost all $x \in H$, the limit

$$dF(x)(c) = \lim_{t\to 0} (\delta_t^{-1})(F(x))^{-1}F(x\delta_t c)$$

exists and is $\delta_u dF(x)(a)\delta_{u'} dF(x)(b)$.

Proof of [3.A].

Let us write M for the Haar mesure on H (which exists because H is a locally compact group), it coincides with the Lebesgue measure in the case of the Heisenberg group.

We first note that if $dF(x)(a)$ exists, then $dF(x)(\delta_t a)$ exists for all t so we can suppose $c = ab$ with $d(1, a) = 1$. The following facts result from the classical Egorov and Lusin theorems :

There exists , for all $\tau > 0$ a closed subspace $F \subseteq H$ such that $M(N \backslash F) < \tau$ and

1. $dF(x)(a), dF(x)(b)$ exist for all $x \in F$.
2. $x \mapsto dF(x)(b)$ is continuous on F.
3. $(\delta_t^{-1})(F(x))^{-1}F(x\delta_t b)$ tends to $dF(x)(b)$, uniformly for $x \in F$.

It would suffice to show that $x\delta_t a \in F$. Indeed, we have

$$\delta_t^{-1}(F(x))^{-1}F(x\delta_t c) = (1)(2)(3)$$

where

$$(1) = \delta_t^{-1}(F(x))^{-1}F(x\delta_t^{-1}a)$$

tends to $dF(x)(a)$ because of 1.

$$(2) = \delta_t^{-1}(F(x\delta_t^{-1}a))^{-1}F(x\delta_t^{-1}a\delta_t b)(dF(x\delta_t a)(b))$$

tends to 1 because of 3. and

$$(3) = dF(x(\delta_t)^{-1}a)(b)$$

tends to $dF(x)(b)$ because of 2.

However, this is not always true. But if x is a density point of F, i.e. if

$$\frac{M(B(x,r)\backslash F)}{M(B(x,r))}$$

tends to 0 with r, then there is a point of F very close to $x\delta_t b$. Set $\lambda = d(x{\delta_t}^{-1}a, F)$ and $x\delta_t^{-1}a'$ a point where this distance is reached. As x is a density point, we see easily that $d(x{\delta_t}^{-1}a, x\delta_t^{-1}a')$ tends to 0 as t tends to 0, and thus $a' \to a$.

We have then

$$\delta_t^{-1}(F(x)^{-1}F(x\delta_t c)) = (1)(2)(3)(4)(5)$$

where

$$(1) = \delta_t^{-1}(F(x))^{-1}F(x\delta_t a)$$

tends to $dF(x)(a)$ because of 1.

$$(2) = \delta_t^{-1}(F(x\delta_t a))^{-1}F(x\delta_t a'),$$

$$(3) = \delta_t^{-1}(F(x\delta_t a')^{-1}F(x\delta_t a'\delta_t b))(dF(x\delta_t a')(b))^{-1}$$

tends to 1 because of 3.

$$(4) = dF(x\delta_t a')(b)$$

tends to $dF(x)(b)$ because of 2. and

$$(5) = \delta_t^{-1}(F(x\delta_t a'\delta_t b)^{-1}F(x\delta_t a\delta_t b))$$

If f is M-Lipschitz, we have

$$\begin{aligned} d((2),1) &= \frac{1}{t}d(F(x\delta_t a), F(x\delta_t a')) \\ &\leq M\frac{1}{t}d(\delta_t a, \delta_t a') \\ &= Md(a,a') \end{aligned}$$

which tends to 0.
And

$$\begin{aligned} d((5),1) &= \frac{1}{t}d(F(x\delta_t a'\delta_t b, F(x\delta_t a\delta_t b))) \\ &\leq M\frac{1}{t}d(\delta_t(ab), \delta_t(a'b)) \\ &= Md(ab, a'b) \end{aligned}$$

which tends to 0 as t tends to 0. This shows the existence of the limit of $(\delta_t^{-1})((F(x))^{-1}F(x\delta_t c))$ for x a density point of F, i.e. almost everywhere in F.

[3.B] Corollary ***(Reduction to dimension 1)***

Let f be a bilipschitz map from O open subset of H to $\mathbb{R}^n$. If every curve in H $s \mapsto f(x\exp(sv))$ (where v is a horizontal vector in the Lie algebra) is differentiable almost everywhere, then f is differentiable almost everywhere and $a \to dF(x)(a)$ is a group homomorphism.

Proof of [3.B].

We just have to verify the uniform convergence for a fixed x of

$$f_t(a) = (\delta_t^{-1})((F(x))^{-1}F(x\delta_t a))$$

to $dF(x)(a)$. One easily shows that H as a group is generated by the subsets $(\delta_s \exp(\xi))_{s\in\mathbb{R}_+}$ and $(\delta_s \exp(\eta))_{s\in\mathbb{R}_+}$ (this follows from the fact that the Lie algebra of H is generated by the two horizontal vectors ξ and η). As the speed of convergence of f_t to $dF(x)$ only depends on $\sup ||a_j||$ and the speed of $f_t(v_j)$ to $d_F(x)(v_j)$, this is the uniform convergence we want.

With the results mentioned in the second part, this theorem is a powerful tool to show the absence of bilipschitz embeddings from H_3 to $\mathbb{R}^n$. As the finite dimensional case is not the only case of interest, we might wonder if the same proof can be generalized. In fact, the same proof can easily be extended to arbitrary Hilbert spaces, or more generally to spaces Y with the Radon-Nikodym property, which states that every Lipschitz map $f : \mathbb{R} \to Y$ is differentiable almost everywhere. However, as L^1 doesn't have the Radon-Nikodym property the non-embeddability result does not follow from Pansu's theorem.

References

[GT98] S. Gonnord and N. Tosel. *Calcul différentiel pour l'agrégation*. Ellipses, 1998.

[Mon02] R. Montgomery. *A Tour of Subriemannian Geometries, Their Geodesics, and Applications*, volume 91 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2002.

[Pan89] P. Pansu. Métriques de Carnot-Carathéodory et quasiisométries des espaces symétriques de rang un. *Ann. of Math.*, 1989.

[Sem96] S. Semmes. On the nonexistence of bilipschitz parametrizations and geometric problems about $a_\infty - weights$. *Revista Matemática Iberoamericana*, 1996.