REGULARITY RESULTS FOR SUB-RIEMANNIAN GEODESICS

ROBERTO MONTI

Abstract. We study length minimality of abnormal curves in rank 2 sub-Riemannian manifolds of polynomial type. As a corollary, we prove a $C^{1,\delta}$ regularity result for Carnot-Carathéodory geodesics in a class of rank 2 Carnot groups.

1. Introduction

Let M be a connected n-dimensional manifold and let $\mathcal{D}$ be a bracket generating distribution on M. Any fixed quadratic form on $\mathcal{D}$ induces a distance on M, known as sub-Riemannian or Carnot-Carathéodory distance. If the resulting metric space is proper, length minimizing curves between any given pair of points do exist. We call these curves "geodesics".

The a priori regularity of geodesics is the Lipschitz regularity. A natural question is whether they have more regularity (see [M2], Chapter 10, Problem 10.1). In fact, there have been several attempts to prove the C^∞-regularity of sub-Riemannian geodesics. A first wrong proof of this claim was based on an incorrect use of Pontryagin Maximum Principle, [S]. This principle provides necessary conditions for solutions of optimal control problems. According to the principle, a sub-Riemannian geodesic is either the projection of a "normal extremal" or the projection of an "abnormal (singular) extremal" (or both). Normal extremals are in fact C^∞ curves solving a system of Hamilton equations. Abnormal extremals, however, satisfy weaker conditions, that in general provide no further regularity beyond the Lipschitz regularity.

The question whether abnormal extremals can be length minimizing was answered in the affirmative by Montgomery [M1]. His example is a C^∞ curve in a three dimensional manifold with bracket generating distribution of rank 2. No example of nonsmooth length minimizing curve is known. In the case of rank 2 distributions, Sussmann and Liu [LS] discovered later a class of abnormal extremals, called "regular abnormal extremals", that are always locally length minimizing. On the other hand, Chitour, Jean, and Trélat recently showed that generically there is no length minimizing singular curve associated to distributions of rank larger or equal to 3 (see Theorem 2.8 in [CJT]).

In [LM], Leonardi and the author proved, in a class of sub-Riemannian manifolds, that curves with corners cannot be length minimizing. So far, this is the unique regularity result covering abnormal geodesics. In this paper, we pursue further the

Date: Revised: 6th November 2012.

This work was partially supported by the Fondazione CaRiPaRo Project "Nonlinear Partial Differential Equations: models, analysis, and control-theoretic problems", Padova.

question of regularity. We restrict the study to the case of rank 2 distributions that are analytic, and in fact polynomial. In a preliminary step, we classify the structure of the singularities of abnormal extremals. This is relatively easy, in analytic, rank two distributions that satisfy Assumption 2.1 below. In a second step, we study an abnormal curve near a singular point. Our working hypothesis is that length minimality for an abnormal curve gets lost when approaching its singularity locus. Roughly speaking, we can prove that this is actually the case when the singularity at some point is of the type $C^{1,\alpha}$, with $\alpha > 0$ small enough. This is the difficult part of the program. One example of such results is Theorem 10.1, that treats the case of Carnot groups. This theorem is a corollary of the construction of Sections 4–8.

The family of admissible (or horizontal) trajectories joining two given points is a manifold that may have points of nondifferentiability. Abnormal extremals are precisely the singular points of this manifold. Correspondingly, the differential of the end-point map is singular at abnormal extremals. In a deep paper [AS], Agrachev and Sarychev developed a second order analysis of the end-point map at abnormal curves. However, a second order analysis may not suffice to capture the behavior of the map in connection with length minimality. Roughly speaking, the singularity may indeed be of "higher order". Objective of this paper is to reduce the analysis of the end-point map to an algebraic problem. We do this through a cut-and-adjust technique that is a nontrivial generalization of the ideas introduced in [LM].

We take an abnormal extremal and we cut it near a singularity. In fact, we cut the horizontal projection of the curve. This produces a gain of length. The new curve is lifted to a horizontal curve. The end-point of the new lifted curve has changed. We perturb the projection using devices depending on various parameters and then we lift again. The goal is to restore the end-point adding a length not exceeding the gained one. If we succeed, we will prove that the abnormal extremal is not length minimizing near the singularity.

From a technical point of view, we have to solve a system of end-point equations with estimates on the solutions. The system can by split into subsystems by means of an equivalence relation of arithmetic type. Each subsystem is singular, meaning that its linearization in the relevant unknowns is singular. This singularity reflects in a precise and effective way the singularity of the end-point map at abnormal extremals. To solve each subsystem, we exploit some algebraic cancellations that are hidden inside the formulas providing the effect of devices on nonhorizontal coordinates. In the end, we have to solve a nonlinear system of equations of Vandermonde type. This is part of an inductive correction that decreases the errors related to all subsystems.

Finally, an iterative procedure sets to zero all the errors. Keeping track of the length employed for the correction provides the threshold of singularity that we can cut with a gain of length after the adjustment. The correction argument begins in Section 4 and finishes in Section 8. In Section 9, we collect the results proved in the previous sections and we comment on the various restrictions that are introduced along the way. In Section 10, we study the case of Carnot groups.

2. Preliminary analysis of the structure of abnormal curves

Let M be an n-dimensional analytic manifold and let $\mathcal{D}$ be an analytic, 2-dimensional distribution on M that is bracket generating. The distribution $\mathcal{D}$ is called *horizontal bundle*. Let $X_1, \dots X_n$ be an analytic frame of vector fields in M. Such a frame always exists locally. The vector fileds X_1, X_2 are such that $\mathcal{D}(x) = \operatorname{span}\{X_1(x), X_2(x)\}$ for all $x \in M$. The vector fields $X_3, \dots, X_n$ are commutators of X_1, X_2. Since our analysis is of a local nature, we may identify M with $\mathbb{R}^n$. Let us assume that we have exponential coordinates of the second type:

$$x = (x_1, \dots, x_n) = \exp(x_1 X_1) \dots \exp(x_n X_n)(0), \quad x \in \mathbb{R}^n.$$

Then we may identify $X_1, \dots, X_n$ with vector fields in $\mathbb{R}^n$ such that for all $x \in \mathbb{R}^n$ we have

$$X_1(x) = \partial_1 \quad \text{and} \quad X_i(x) = \sum_{j=1}^{n} a_{ij}(x)\partial_j, \quad i = 2, \dots, n,$$

where a_{ij} are analytic functions on $\mathbb{R}^n$ such that $a_{ij}(0) = \delta_{ij}$. Through elementary algebraic operations on the horizontal frame, we may always assume that $a_{21} = 0$ and $a_{22} = 1$ in $\mathbb{R}^n$. Eventually, we have $\mathcal{D} = \operatorname{span}\{X_1, X_2\}$ with

$$X_1 = \partial_1 \quad \text{and} \quad X_2 = \partial_2 + \sum_{j=3}^{n} f_j(x)\partial_j, \quad x \in \mathbb{R}^n, \tag{2.1}$$

where $f_3, \dots, f_n$ are analytic functions. To compute in an effective way the formulas of Section 3 for the horizontal lift of plane curves, we need the following nontrivial structural hypothesis.

Assumption 2.1. There exists a system of coordinates such that the horizontal bundle $\mathcal{D}$ is spanned by vector fields X_1, X_2 as in (2.1) such that $f_j(x) = f_j(x_1, x_2)$ for all $j = 3, \dots, n$.

Then the functions f_j also satisfy $f_j(0, x_2) = 0$, as soon as we are in exponential coordinates of the second type. We may also assume that $f_3, \dots, f_n$ are linearly independent. The linear dependence of $f_3, \dots, f_n$ would contradict the bracket generating assumption of $\mathcal{D}$.

A Lipschitz curve $\gamma : [0, 1] \to M$ is $\mathcal{D}$-horizontal if $\dot\gamma(t) \in \mathcal{D}(\gamma(t))$ for a.e. $t \in [0, 1]$. When $\mathcal{D}$ is spanned point wise by the frame X_1, X_2 as in (2.1), γ is horizontal if and only if we have

$$\dot\gamma(t) = \sum_{i=1}^{2} \dot\gamma_i(t) X_i(\gamma(t)), \quad \text{for a.e. } t \in [0, 1].$$

We call the plane curve $\kappa : [0, 1] \to \mathbb{R}^2$, $\kappa = (\gamma_1, \gamma_2)$, the *horizontal projection* of γ and we let $\kappa = \operatorname{Proj}(\gamma)$. The curve γ is determined by its horizontal projection κ. In fact, we have for any $i = 3, \dots, n$ and for all $t \in [0,1]$

$$\begin{aligned} \gamma_i(t) &= \gamma_i(0) + \int_{\kappa|_{[0,t]}} f_i(x_1, x_2) dx_2 \\ &= \gamma_i(0) + \int_0^t \dot\kappa_2(s) f_i(\kappa(s))\, ds. \end{aligned} \tag{2.2}$$

Here and hereafter, we are using Assumption 2.1. Given a Lipschitz curve κ in the plane, a horizontal curve γ defined by $(\gamma_1, \gamma_2) = \kappa$ and by (2.2) is said to be a *lift of* κ. If the starting point $\gamma(0)$ is fixed, we call γ the lift of κ and we let $\gamma = \mathrm{Lift}(\kappa)$.

Let us fix on $\mathcal{D}$ a quadratic form g. As the choice of g is not relevant for our argument, we can assume that g is the quadratic form that makes X_1, X_2 orthonormal. Then the length of γ in $(M, \mathcal{D}, g)$ is

$$L(\gamma) = \int_0^1 |\dot\kappa(t)|\, dt, \tag{2.3}$$

where $|\dot\kappa|$ is the standard length of $\dot\kappa$. We are interested in length-minimizing curves. A $\mathcal{D}$-horizontal curve γ is *length minimizing* if $L(\gamma) \le L(\widetilde\gamma)$ for any other $\mathcal{D}$-horizontal curve $\widetilde\gamma : [0,1] \to M$ such that $\widetilde\gamma(0) = \gamma(0)$ and $\widetilde\gamma(1) = \gamma(1)$. According to Pontryagin Maximum Principle, length minimizing curves may be either abnormal (singular) or normal (or both). When the frame X_1, X_2 satisfies (2.1), these facts can be summarized as follows. We need the function $K : \mathbb{R}^{n-2} \times \mathbb{R}^2 \to \mathbb{R}$ defined as

$$K(\mu, x) = \sum_{i=1}^{n-2} \mu_i \partial_1 f_{i+2}(x),$$

where $\mu = (\mu_1, \dots, \mu_{n-2}) \in \mathbb{R}^{n-2}$ and $x \in \mathbb{R}^2$. Notice that the function $x \mapsto K(\mu, x)$ does not vanish identically, if $\mu \neq 0$. If we had $K(\mu, x) = 0$ for all $x \in \mathbb{R}^2$ and for some $\mu \neq 0$, then there would hold

$$\sum_{i=1}^{n-2} \mu_i f_{i+2}(x_1, x_2) = \psi(x_2)$$

for some function ψ. From the property $f_{i+2}(0, x_2) = 0$ it would follow that $\psi = 0$ and the functions $f_3, \dots, f_n$ would be linearly dependent.

Proposition 2.2. Let $\gamma : [0,1] \to M$ be a $\mathcal{D}$-horizontal curve that is length minimizing in $(M, \mathcal{D}, g)$. Let $\kappa = \mathrm{Proj}(\gamma)$ be the horizontal projection of γ and assume that $|\dot\kappa| = 1$ almost everywhere. Then one of (or both) the following two statements holds:

1) There exists $\mu \in \mathbb{R}^{n-2}$, $\mu \neq 0$, such that

$$K(\mu, \kappa(t)) = 0, \qquad \text{for all } t \in [0,1]. \tag{2.4}$$

2) The curve γ is smooth (analytic) and there exists $\mu \in \mathbb{R}^{n-2}$ such that κ solves the the system of equations

$$\ddot\kappa = K(\mu, \kappa)\dot\kappa^\perp, \tag{2.5}$$

where $\kappa^\perp = (-\kappa_2, \kappa_1)$.

A proof of Proposition 2.2 can be found in [LM], Propositions 4.2 and 4.3.

The interesting and difficult case is 1): the curve κ is in the zero set of an analytic function that does not vanish identically. Fix $\mu \in \mathbb{R}^{n-2}$ and let

$$f(x) = \sum_{i=1}^{n-2} \mu_i \partial_1 f_{i+2}(x) \neq 0,$$

with $x = (x_1, x_2) \in \mathbb{R}^2$. Let the support of κ be contained in the zero set of f. Without loss of generality, we can assume that $\kappa(0) = 0$, i.e., we can assume that $f(0) = 0$. If 0 is a regular point of f, i.e., $\nabla f(0) \neq 0$, then κ is an analytic curve passing through 0. The zero set of f, however, may have a singularity at 0 and the curve κ may "switch" from one piece of analytic curve in the zero set to another piece of analytic curve in the zero set. When the two pieces form a corner, the curve κ passing through the corner can not be the horizontal projection of a length minimizing curve. The singularity destroys length minimality. This is proved in [LM].

Here, we study the general question whether γ looses length minimality when we approach along κ the singular part of the zero set of the analytic function f. We need a precise description of the behavior of κ near the singular set. By the Puiseux expansion theorem (see the more general uniformization theorem for real analytic sets, Theorem 5.1 in [BM]), there exists an analytic mapping $\Phi = (\varphi_1, \varphi_2) : [0,1] \to \mathbb{R}^2$ such that $\Phi(0) = 0$ and $\kappa([0,1]) = \Phi([0,1])$. For some integers $\alpha, \beta \in \mathbb{N}$, $\alpha, \beta \geq 1$, we have the convergent power series in $t \in [0, \varepsilon]$, for some $\varepsilon > 0$ say $\varepsilon = 1$,

$$\varphi_1(t) = \sum_{i=\alpha}^{\infty} c_i t^i, \quad \varphi_2(t) = \sum_{i=\beta}^{\infty} d_i t^i, \tag{2.6}$$

with real coefficients $c_i, d_i \in \mathbb{R}$ such that $c_\alpha \neq 0$ and $d_\beta \neq 0$. In the case $\varphi_1 = 0$ or $\varphi_2 = 0$ we have lines. We are ignoring this case.

By an elementary blow-up argument it can be proved that the derivative $\dot\kappa(0)$ does exist. After a rotation of the coordinates in the plane, we may assume that $\dot\kappa(0) = (1, 0)$. As the curve $t \mapsto \Phi(t)$, $t \geq 0$, is a re-parameterization of κ near 0, we deduce that in (2.6) we have $\alpha < \beta$. Throughout the paper, we denote by

$$r = \frac{\beta}{\alpha} \in \mathbb{Q}, \quad r > 1, \tag{2.7}$$

the exponent describing the behavior of the curve near 0.

After a new re-parameterization, we may assume that $\kappa : [0,1] \to \mathbb{R}^2$ is the curve $\kappa(t) = (t, \varphi(t))$ where the function $\varphi : [0,1] \to \mathbb{R}$ is given by the Puiseux' series

$$\varphi(t) = \sum_{i=\beta}^{\infty} c_i t^{\frac{i}{\alpha}}, \quad t \in [0,1]. \tag{2.8}$$

The numbers $c_i \in \mathbb{R}$ are the coefficients of the series. We assume without loss of generality that $c_\beta = 1$. Then we have $\varphi(t) = t^r + o(t^r)$ and $\varphi'(t) = rt^{r-1} + o(t^{r-1})$ as $t \to 0^+$. It is easy to check that φ is of class $C^{1,r-1}([0,1])$, when $1 < r \le 2$.

According to whether all coefficients c_i with $i > \beta$ vanish or not, we distinguish two cases:

1) *Homogeneous case:* $c_i = 0$ for all $i > \beta$. In this case we have $\varphi(t) = t^r$. The systems of algebraic equations studied in Sections 5 and 6 are singular in a precise sense that will be clear later.
2) *Nonhomogeneous case:* there exists $i > \beta$ such that $c_i \neq 0$. In this case, we obtain better estimates on the length. In a certain sense, the algebraic systems that we solve are "less singular". This will be clear in Section 8.

We feel that the nonhomogeneous case can be reduced to the homogeneous case. We tried to do this by a blow-up argument tailored to the curve $t \mapsto (t, t^r)$, i.e., using dilatations in the plane of the form $(x_1, x_2) \mapsto (\lambda x_1, \lambda^r x_2)$, $\lambda > 0$, suitably extended to $\mathbb{R}^n$. This would also have the advantage of reducing the case of analytic distributions to the case of polynomial distributions. The metric g, however, does not pass correctly to the limit. Thus we have been forced to study both cases, the homogeneous one and the nonhomogeneous one.

In the next section, we introduce the tools used in the proofs and we set up the algebraic framework of the correction argument.

3. Equivalence classes and correction devices

We are interested in the length minimality near $t = 0$ of the horizontal curve $\gamma = \mathrm{Lift}(\kappa)$ with $\kappa : [0,1] \to \mathbb{R}^2$, $\kappa(t) = (t, \varphi(t))$ and φ as in (2.8). We cut the curve κ through a segment near $t = 0$. For any $0 < \eta < 1$, let $T_\eta \subset \mathbb{R}^2$ be the set

$$T_\eta = \Big\{ (x_1, x_2) \in \mathbb{R}^2 : \varphi(x_1) < x_2 < \frac{\varphi(\eta)}{\eta} x_1,\ 0 < x_1 < \eta \Big\}.$$

The boundary ∂T_η is oriented counterclockwise. Let $\kappa^\eta : [0,1] \to \mathbb{R}^2$ be the curve $\kappa^\eta(t) = (t, \varphi(\eta)t/\eta)$ for $0 \le t \le \eta$ and $\kappa^\eta(t) = (t, \varphi(t))$ for $\eta \le t \le 1$, and let $\gamma^\eta = \mathrm{Lift}(\kappa^\eta)$ be the horizontal lift of κ^η. This lift can be computed using the formulas (2.2). The length $L(\gamma^\eta)$ of the curve γ^η is shorter than the length of γ. By formula (2.3), we can compute the gain of length $\Delta L(\eta)$:

$$\Delta L(\eta) = L(\gamma) - L(\gamma^\eta) = \frac{(r-1)^2}{2(2r-1)} \eta^{2r-1} + o\big(\eta^{2r-1}\big). \tag{3.1}$$

The end point $\gamma(1)$ is modified, i.e., $\gamma^\eta(1) \neq \gamma(1)$. In the next sections, we develop a technique to restore the end-point on modifying κ^η away from the cut. Formula (3.1) gives us the total amount of length that we can use for this adjustment.

To compute the error produced by the cut T_η on each coordinate $h = 3, \dots, n$, we compute the error of the cut on each monomial of the Taylor expansion of each analytic function f_h, $h = 3, \dots, n$. Assume that we have

$$f_h(x) = \sum_{i,j=0}^{\infty} b_{ijh} x_1^{i+1} x_2^j, \quad x \in \mathbb{R}^2,$$

for suitable coefficients $b_{ijh} \in \mathbb{R}$. Formula (2.2) for the lift provides the effect of the cut T_η on the hth-coordinate

$$\int_{\partial T_\eta} f_h(x) dx_2 = \sum_{i,j=0}^{\infty} b_{ijh} \int_{\partial T_\eta} x_1^{i+1} x_2^j dx_2.$$

This leads us to define the error T_η^{ij} produced by the cut T_η on each monomial $x_1^{i+1} x_2^j$, equivalently on the pair (i,j), for $i, j \in \mathbb{N}$:

$$T_\eta^{ij} = \int_{\partial T_\eta} x_1^{i+1} x_2^j dx_2. \tag{3.2}$$

By Stokes' theorem, after some computations we find the following expression

$$\begin{aligned} T_\eta^{ij} &= \frac{i+1}{j+1}\Big[\frac{1}{i+j+2} - \frac{1}{i+(j+1)r+1}\Big]\eta^{i+r(j+1)+1} + o(\eta^{i+r(j+1)+1}) \\ &= c_{ij}\eta^{i+r(j+1)+1} + o(\eta^{i+r(j+1)+1}), \end{aligned} \tag{3.3}$$

where the constants c_{ij} are defined through the last equality. Notice that the exponent $i + r(j+1) + 1$ may attain the same value for different pairs of integers (i,j). This fact reflects the singularity of the end-point map at abnormal curves.

We introduce the correction devices that will be used to correct the end-point. For fixed parameters $b > 0$, $\lambda > 0$ and $\varepsilon > 0$, let us define the curvilinear rectangles

$$R_{b,\lambda}(\varepsilon) = \big\{(x_1, x_2) \in \mathbb{R}^2 : b < x_1 < b + |\varepsilon|^\lambda,\ \varphi(x_1) < x_2 < \varphi(x_1) + \varepsilon\big\}. \tag{3.4}$$

When $\varepsilon < 0$ we let

$$R_{b,\lambda}(\varepsilon) = \big\{(x_1, x_2) \in \mathbb{R}^2 : b < x_1 < b + |\varepsilon|^\lambda,\ \varphi(x_1) + \varepsilon < x_2 < \varphi(x_1)\big\}. \tag{3.5}$$

The boundary $\partial R_{b,\lambda}(\varepsilon)$ is oriented counterclockwise if $\varepsilon > 0$, while it is oriented clockwise when $\varepsilon < 0$. The *cost of length* of the device $R_{b,\lambda}(\varepsilon)$ is

$$\Lambda\big(R_{b,\lambda}(\varepsilon)\big) = 2|\varepsilon|. \tag{3.6}$$

We denote by $R_{b,\lambda}^{ij}(\varepsilon)$ the *effect of the device* $R_{b,\lambda}(\varepsilon)$ *on the pair* (i,j). Namely, we let

$$R_{b,\lambda}^{ij}(\varepsilon) = \int_{\partial R_{b,\lambda}^{ij}(\varepsilon)} x_1^{i+1} x_2^j dx_2.$$

We disaggregate this effect into a sum of effects highlighting the leading term, the second leading term and so on. To this aim, notice that for any integer $h \in \mathbb{N}$ there are constants $c_{h\ell} \in \mathbb{R}$, $\ell \in \mathbb{N}$, such that the h-power of φ has the expansion

$$\varphi(t)^h = \sum_{\ell=0}^{\infty} c_{h\ell} t^{rh+\frac{\ell}{\alpha}}, \quad t \in [0,1]. \tag{3.7}$$

By Stokes' theorem and formula (3.7), we find after some computations the explicit formula

$$R^{ij}_{b,\lambda}(\varepsilon) = \sum_{\ell=0}^{\infty}\sum_{h=0}^{j} c_{ijh\ell} R^{ijh\ell}_{b,\lambda}(\varepsilon), \tag{3.8}$$

where we let

$$\begin{aligned} c_{ijh\ell} &= \frac{(i+1)c_{h\ell}}{(j+1)(i+rh+\frac{\ell}{\alpha}+1)}\binom{j+1}{h}, \\ R^{ijh\ell}_{b,\lambda}(\varepsilon) &= \varepsilon^{j+1-h}\Big[(b+|\varepsilon|^{\lambda})^{i+rh+\frac{\ell}{\alpha}+1} - b^{i+rh+\frac{\ell}{\alpha}+1}\Big]. \end{aligned} \tag{3.9}$$

Formula (3.8) with (3.9) holds for any positive or negative ε. Notice that when $j+1-h$ is even, we do not have control on the sign of $R^{ijh\ell}_{b,\lambda}$.

Next, let us introduce square-devices. Let $0 < b < 1$ be a position parameter. For any $\varepsilon \in (-1,1)$ let

$$Q_b(\varepsilon) = \big\{(x_1,x_2)\in\mathbb{R}^2 : b < x_1 < b+|\varepsilon|,\ \varphi(x_1) < x_2 < \varphi(x_1)+|\varepsilon|\big\}. \tag{3.10}$$

When $\varepsilon > 0$, the boundary $\partial Q_b(\varepsilon)$ of the square is oriented clockwise. When $\varepsilon < 0$ the boundary is oriented counterclockwise. We talk of "squares" because the sides of $Q_b(\varepsilon)$ are of the same size. Applying the device $Q_b(\varepsilon)$ to the curve κ^η means that at the point $(b,\varphi(b))$ the curve κ^η is deviated along the boundary of $Q_b(\varepsilon)$ in the sense determined by the sign of ε and, after one loop, we follow again the curve κ^η. The *cost of length* $\Lambda(Q_b(\varepsilon))$ of the square is the sum of the length of the four sides. For some constant $C>0$ independent of b and ε we have

$$\Lambda(Q_b(\varepsilon)) \le C|\varepsilon|. \tag{3.11}$$

We denote by $Q^{ij}_b(\varepsilon)$ the *effect of the device* $Q_b(\varepsilon)$ *on the pair* (i,j). Namely, we let

$$Q^{ij}_b(\varepsilon) = \int_{\partial Q_b(\varepsilon)} x_1^{i+1}x_2^j\,dx_2,$$

where the boundary is oriented according to the sign of ε. By Stokes' theorem, we find the formula

$$Q^{ij}_b(\varepsilon) = \sum_{\ell=0}^{\infty}\sum_{h=0}^{j} c_{ijh\ell} Q^{ijh\ell}_b(\varepsilon), \tag{3.12}$$

where the constants $c_{ijh\ell}$ are the same as in (3.9) and

$$Q^{ijh\ell}_b(\varepsilon) = \mathrm{sgn}(\varepsilon)|\varepsilon|^{j+1-h}\Big[(b+|\varepsilon|)^{i+rh+\frac{\ell}{\alpha}+1} - b^{i+rh+\frac{\ell}{\alpha}+1}\Big].$$

Here, we control the sign.

The numbers $R^{ijh\ell}_{b,\lambda}(\varepsilon)$ and $Q^{ijh\ell}_b(\varepsilon)$ satisfy the following identities. Recall that the devices are defined starting from the function φ with its parameters α and β.

Proposition 3.1. For all integers $i, j, h, p, \ell \in \mathbb{N}$ such that $h \leq j$ and $p\beta \leq i$ there holds:

$$\begin{aligned} R_{b,\lambda}^{ijh\ell}(\varepsilon) &= R_{b,\lambda}^{i-p\beta,j+p\alpha,h+p\alpha,\ell}(\varepsilon), \\ Q_b^{ijh\ell}(\varepsilon) &= Q_b^{i-p\beta,j+p\alpha,h+p\alpha,\ell}(\varepsilon). \end{aligned} \tag{3.13}$$

The proof is elementary. These identities play a central role in the quantitative solution of a nonlinear system of equations in Sections 5–6 and Section 8.

Before proceeding, we need to group the monomials $x_1^{i+1}x_2^j$ into equivalence classes that correspond to proportional effect of cut and devices. Recall that $r \in \mathbb{Q}$, $r > 1$, is the rational number $r = \beta/\alpha$, where $\alpha, \beta \in \mathbb{N}$ with $1 < \alpha < \beta$ are the integers appearing in (2.8). From now on, the integers α and β are fixed.

Let $\sim$ be the equivalence relation on $\mathbb{N} \times \mathbb{N}$:

$$(i, j) \sim (i', j') \quad \text{if and only if} \quad i\alpha + j\beta = i'\alpha + j'\beta.$$

For any $k \in \mathbb{N}$ we have the equivalence class

$$L_k = \Big\{(i, j) \in \mathbb{N} \times \mathbb{N} : i\alpha + j\beta = k\Big\}.$$

It may be $L_k = \emptyset$ for a finite set of integers k. For any $k \in \mathbb{N}$ such that $L_k \neq \emptyset$, we call the representative $(i, j) \in L_k$ such that $j = 0, 1, \dots, \alpha - 1$ the *first representative* of the class. For $k \in \mathbb{N}$, let (i, j) be the first representative of L_k. We let

$$\bar{k} = [i/\beta], \tag{3.14}$$

where $[\cdot]$ stands for the integer part. Then, L_k has exactly $\bar{k}+1$ elements, and namely:

$$L_k = \big\{(i - p\beta, j + p\alpha) \in \mathbb{N} \times \mathbb{N} : p = 0, 1, \dots, \bar{k}\big\}. \tag{3.15}$$

In the sequel, it will be useful to have a short notation for the following number depending on $k \in \mathbb{N}$:

$$\ell_k = i + jr + 1 = \frac{k}{\alpha} + 1, \tag{3.16}$$

where (i, j) is any pair such that $(i, j) \in L_k$. This number appears, e.g., in the exponent of η in (3.3).

The remaining part of the paper is organized as follows:

1) In a first step, we correct all the first representatives. This is done in Section 4. Here, we use only rectangles and not squares. The identities (3.13) are not needed. This section is relevant for the analysis of the curve κ both in the homogeneous case and in the nonhomogeneous case.
2) In a second step, we study the homogeneous case. In particular, we correct the error of all equivalence classes. We have to solve a singular system of algebraic equations and now we need squares. This is done in Section 5 when the system is two dimensional and in Section 6 for the general case. The identities (3.13) are central.

3) Then we analyze the nonhomogeneous case. For the sake of simplicity, we confine ourselves to equivalence classes with at most two elements. This is done in Section 8, that continues, also in the notation, Section 4.
4) In Section 9, we briefly estimate the cost of length of the various procedures and we collect the results of the previous sections.
5) Finally, in Section 10 we discuss the case of Carnot groups. In particular, in Theorem 10.1 we prove a $C^{1,\delta}$ regularity result for Carnot-Carathéodory geodesics.

Notation. If A and B are real functions depending on η, the notation $A \lesssim B$ means that there is a constant $C > 0$ independent of η such that $|A| \le C|B|$. The notation $A \simeq B$ means that there is a constant $C > 0$ independent of η such that $C^{-1}|B| \le |A| \le C|B|$. The notations $A \lesssim B$ and $A \simeq B$ for vector valued functions have the same meaning, but component wise.

If $A = A(\lambda, \mu)$ and $B = B(\lambda, \mu)$ are functions depending on a finite set of real parameters $\lambda = (\lambda_0, \lambda_1, \dots)$ and $\mu = (\mu_0, \mu_1, \dots)$, by $A \sim B$ we mean $A(0) = B(0)$.

4. Correction of first representatives

There is a bijection between the pairs $(i, j) \in \mathbb{N} \times \mathbb{N}$ such that $j \le \alpha - 1$ and the integer $k \in \mathbb{N}$ such that $i\alpha + j\beta = k$. The pair (i, j) is the first representative of the equivalence class L_k.

We strengthen Assumption 2.1 requiring the functions $f_3, \dots, f_n$ to be polynomials.

Assumption 4.1. There exists a system of coordinates such that the horizontal bundle $\mathcal{D}$ is spanned by vector fields X_1, X_2 as in (2.1) such that the functions $f_j(x) = f_j(x_1, x_2)$, $j = 3, \dots, n$, are polynomials.

Under assumption (4.1) we have the bound $k \le K$ for some fixed $K \in \mathbb{N}$. This assumption is needed to bound with a finite multiplicative constant the cost of length. See Remark 4.3 below.

We call *error* or *effect on* k the error or effect on the pair (i, j) corresponding to k. The error T_η^k produced by the cut T_η on k is given by formula (3.3). Namely, we have

$$T_\eta^k = c_k \eta^{\ell_k + r} + o(\eta^{\ell_k + r}), \tag{4.1}$$

where c_k are constants depending on $(i, j) \in L_k$ such that $j = 0, 1, \dots, \alpha - 1$. We also define the *initial error produced by the cut* T_η as the vector:

$$\mathrm{E}(\eta) = \big(c_0 \eta^{\ell_0 + r} + o(\eta^{\ell_0 + r}), \dots, c_K \eta^{\ell_K + r} + o(\eta^{\ell_K + r})\big). \tag{4.2}$$

In general, the space of errors is indicated by $\mathcal{E} = \mathbb{R}^{K+1}$.

A *correction* of an error $\mathrm{E} \in \mathcal{E}$ is an at most countable union R of devices R_h as in (3.4)–(3.5), $h \in \mathbb{N}$, that sets to zero the vector E. The *cost of length* of the correction R is the sum of the cost of length of the devices

$$\Lambda(R) = \sum_{h \in \mathbb{N}} \Lambda(R_h).$$

Theorem 4.2. *Let $\varepsilon > 0$. There are numbers $C > 0$ and $0 < \delta < 1$ such that for any $0 < \eta < \delta$ there is a correction R of the initial error* $\mathrm{E}(\eta)$ *in* (4.2) *with cost of length satisfying*

$$\Lambda(R) \leq C\eta^{1+r-\varepsilon}. \tag{4.3}$$

Proof. We divide the proof into several steps.

Step 1: Correction of $k = 0$. Let us fix parameters $\eta_0 = \eta < \delta$ and $\lambda_0 > 0$. The precise choice of $\delta > 0$ and λ_0 will be done at the end of our construction.

We correct $k = 0$ with the device $R_{\eta_0,\lambda_0}(\varepsilon)$, for some $\varepsilon \in \mathbb{R}$. From the formula (3.8), we obtain the effect $R^0_{\eta_0,\lambda_0}(\varepsilon)$ produced by $R_{\eta_0,\lambda_0}(\varepsilon)$ on $k = 0$ (i.e., with $i = j = 0$):

$$R^0_{\eta_0,\lambda_0}(\varepsilon) = \varepsilon|\varepsilon|^{\lambda_0}. \tag{4.4}$$

The effect does not depend on η_0. By (4.1) with $k = 0$ (and thus $\ell_k = 1$), the equation $R^0_{\eta_0,\lambda_0}(\varepsilon) + T^k_\eta = 0$ is then $\varepsilon|\varepsilon|^{\lambda_0} + c_0\eta^{1+r} = 0$ and its solution $\varepsilon = \varepsilon_0$ is

$$\varepsilon_0 = -c_0^{\frac{1}{1+\lambda_0}}\eta^{\frac{1+r}{1+\lambda_0}}. \tag{4.5}$$

Here, c_0 is the constant appearing in (4.2). The cost of length of $R_{\eta_0,\lambda_0}(\varepsilon)$ is

$$\Lambda(R_{\eta_0,\lambda_0}(\varepsilon_0)) = 2|\varepsilon_0| = 2c_0^{\frac{1}{1+\lambda_0}}\eta^{\frac{1+r}{1+\lambda_0}}. \tag{4.6}$$

The device $R_{\eta_0,\lambda_0}(\varepsilon_0)$ produces an additional error $R^k_{\eta_0,\lambda_0}(\varepsilon_0)$ on $k \neq 0$. This error can be obtained starting from formula (3.8):

$$\begin{aligned} R^k_{\eta_0,\lambda_0}(\varepsilon_0) = {} & \frac{i+1}{(j+1)\ell_k}\varepsilon_0\Big[(\eta_0 + |\varepsilon_0|^{\lambda_0})^{\ell_k} - \eta_0^{\ell_k}\Big] + \\ & + \frac{i+1}{j+1}\sum_{h=0}^{j-1}\binom{j+1}{h}\frac{\varepsilon_0^{j+1-h}}{i+rh+1}\Big[(\eta_0 + |\varepsilon_0|^{\lambda_0})^{i+rh+1} - \eta_0^{i+rh+1}\Big]. \end{aligned} \tag{4.7}$$

We highlighted the jth summand. In order to determine the leading term, we compare $\eta_0 = \eta$ and $|\varepsilon_0|^{\lambda_0}$. We have $\eta_0 < \frac{1}{2}|\varepsilon_0|^{\lambda_0}$ as soon as η is small enough and

$$1 > \frac{\lambda_0(1+r)}{1+\lambda_0}, \quad \text{that is} \quad \lambda_0 < \frac{1}{r}.$$

The condition $\lambda_0 < 1/r$ is our first condition on λ_0. Then for any fixed $h = 0, 1, \dots, j$, we have

$$\varepsilon_0^{j+1-h}\Big[(\eta_0 + |\varepsilon_0|^{\lambda_0})^{i+rh+1} - \eta_0^{i+rh+1}\Big] \simeq \varepsilon_0^{j+1-h+\lambda_0(i+rh+1)}.$$

The exponent $e(h) = j + 1 - h + \lambda_0(i + rh + 1)$ satisfies $e'(h) = r\lambda_0 - 1 < 0$ and thus it achieves the minimum value when h is maximum. Thus formula (4.7) may be written in the following way:

$$R^k_{\eta_0,\lambda_0}(\varepsilon_0) = \frac{i+1}{(j+1)\ell_k}\varepsilon_0\Big[(\eta_0 + |\varepsilon_0|^{\lambda_0})^{\ell_k} - \eta_0^{\ell_k}\Big] + c_{ij}|\varepsilon_0|^{2+\lambda_0(\ell_k-r)} + \mathrm{Error}, \tag{4.8}$$

where Error is a negligible quantity with respect to the preceding term and c_{ij} are constants that can be computed. Similar formulas with the precise second leading term will be needed in Sections 5, 6, and 8. A rougher expression for the effect is

$$R^k_{\eta_0,\lambda_0}(\varepsilon_0) \simeq |\varepsilon_0|^{1+\ell_k\lambda_0} \simeq \eta^{\frac{(1+\ell_k\lambda_0)(1+r)}{1+\lambda_0}}.$$

This error dominates the error T^k_η produced by the cut T_η. In fact, we have

$$\frac{(1+\ell_k\lambda_0)(1+r)}{1+\lambda_0} < \ell_k + r \quad \Leftrightarrow \quad \lambda_0 < \frac{1}{r}.$$

After correcting $k = 0$ we have a new vector of errors E_0 such that:

$$\mathrm{E}_0 \lesssim \big(0, \dots, \eta^{\frac{(1+\ell_k\lambda_0)(1+r)}{1+\lambda_0}}, \dots\big). \tag{4.9}$$

In the vector above, k ranges from 1 to K.

Step 2: Correction of $k = \alpha$. For the sake of clearness, we provide details for the correction of $k = \alpha$. This is the minimum $k \geq 1$ such that $L_k \neq \emptyset$. The integers $k = 1, 2, \dots, \alpha - 1$ are not related to any pair (i, j). In the next step, we shall set up the inductive construction.

When $k = \alpha$ we have $i = 1$ and $j = 0$. We correct the error with the device $R_{\eta_\alpha,\lambda_\alpha}(\varepsilon)$ where the parameters η_α and λ_α are chosen according to the following rules.

i) The parameter λ_α is such that

$$0 < \lambda_\alpha < \lambda_0. \tag{4.10}$$

This choice will ensure a general decrease of errors.

ii) The position parameter η_α should be as small as possible in order to produce the smallest effect on $k > \alpha$. The position must be compatible with the devices already present along the curve. We may then choose $\eta_\alpha \geq \eta_0 + |\varepsilon_0|^{\lambda_0}$. In fact, as the second term is larger, we choose

$$\eta_\alpha = 2|\varepsilon_0|^{\lambda_0} \simeq \eta^{\frac{\lambda_0(1+r)}{1+\lambda_0}}.$$

We denote by E^α_0 the error on $k = \alpha$ after the cut and the correction of $k = 0$. This error is $E^\alpha_0 = T^\alpha_\eta + R^\alpha_{\eta_0,\lambda_0}(\varepsilon_0) \simeq R^\alpha_{\eta_0,\lambda_0}(\varepsilon_0)$ where

$$R^\alpha_{\eta_0,\lambda_0}(\varepsilon_0) = \varepsilon_0\Big[(\eta_0 + |\varepsilon_0|^{\lambda_0})^{\ell_\alpha} - \eta_0^{\ell_\alpha}\Big] = -c_0^{\frac{1+\ell_\alpha\lambda_0}{1+\lambda_0}}\eta^{\frac{(1+\ell_\alpha\lambda_0)(1+r)}{1+\lambda_0}} + \text{Error}, \tag{4.11}$$

and Error is a negligible quantity depending on η. The equation $R^\alpha_{\eta_\alpha,\lambda_\alpha}(\varepsilon) + E^\alpha_0 = 0$ in the unknown ε is then

$$\varepsilon\Big[(\eta_\alpha + |\varepsilon|^{\lambda_\alpha})^{\ell_\alpha} - \eta_\alpha^{\ell_\alpha}\Big] = c_0^{\frac{1+\ell_\alpha\lambda_0}{1+\lambda_0}}\eta^{\frac{(1+\ell_\alpha\lambda_0)(1+r)}{1+\lambda_0}} + \text{Error}. \tag{4.12}$$

We look for a solution ε satisfying the condition

$$\frac{1}{2}|\varepsilon|^{\lambda_\alpha} > \eta_\alpha. \tag{4.13}$$

In this case, the equation $R^\alpha_{\eta_\alpha,\lambda_\alpha}(\varepsilon)+E^3_0=0$ can be approximated by the following equation:

$$\varepsilon|\varepsilon|^{\ell_\alpha\lambda_\alpha}=c_0^{\frac{1+\ell_\alpha\lambda_0}{1+\lambda_0}}\eta^{\frac{(1+\ell_\alpha\lambda_0)(1+r)}{1+\lambda_0}}+\text{Error}.$$

This equation has a solution $\varepsilon=\varepsilon_\alpha$ that satisfies:

$$\varepsilon_\alpha=c_0^{\frac{1+\ell_\alpha\lambda_0}{(1+\lambda_0)(1+\ell_\alpha\lambda_\alpha)}}\eta^{\frac{(1+\ell_\alpha\lambda_0)(1+r)}{(1+\lambda_0)(1+\ell_\alpha\lambda_\alpha)}}+\text{Error}. \tag{4.14}$$

This solution satisfies condition (4.13). In fact, we have

$$\frac{\lambda_\alpha(1+r)(1+\ell_\alpha\lambda_0)}{(1+\lambda_0)(1+\ell_\alpha\lambda_\alpha)}<\frac{\lambda_0(1+r)}{1+\lambda_0}\quad\Leftrightarrow\quad\frac{1+\ell_\alpha\lambda_0}{\lambda_0}<\frac{1+\ell_\alpha\lambda_\alpha}{\lambda_\alpha}\quad\Leftrightarrow\quad\lambda_\alpha<\lambda_0.$$

Notice that $\ell_\alpha=2$. This shows that (4.13) holds.

The argument starting from (4.13) can now be made rigorous in the following way. Let ε_α be given by formula (4.14). By the intermediate value theorem we can show that equation (4.12) has a solution ε such that $\varepsilon\in[\varepsilon_\alpha/M,\varepsilon_\alpha M]$ where $M>1$ is a suitable constant independent of η. The solution ε satisfies (4.13). We shall use this argument freely in the next step.

The cost of length of $R_{\eta_\alpha,\lambda_\alpha}(\varepsilon_\alpha)$ is

$$\Lambda(R_{\eta_\alpha,\lambda_\alpha}(\varepsilon_\alpha))=2|\varepsilon_\alpha|.$$

Now we compute the effect $R^k_{\eta_\alpha,\lambda_\alpha}(\varepsilon_\alpha)$ of $R_{\eta_\alpha,\lambda_\alpha}(\varepsilon_\alpha)$ on $k\neq\alpha$. This effect is given by formula (4.11), replacing 0 with α:

$$\begin{aligned}R^k_{\eta_\alpha,\lambda_\alpha}(\varepsilon_\alpha)&=\frac{i+1}{(j+1)\ell_k}\varepsilon_\alpha\Big[(\eta_\alpha+|\varepsilon_\alpha|^{\lambda_\alpha})^{\ell_k}-\eta_\alpha^{\ell_k}\Big]+c_{ij}|\varepsilon_\alpha|^{2+\lambda_\alpha(\ell_k-r)}+\text{Error}\\&\simeq\eta^{\frac{(1+\ell_\alpha\lambda_0)(1+\ell_k\lambda_\alpha)(1+r)}{(1+\lambda_0)(1+\ell_\alpha\lambda_\alpha)}}.\end{aligned} \tag{4.15}$$

We compare this effect with the error on $k\geq1$ produced by the correction of $k=0$ (see (4.9)). Notice that for $k>\alpha$ we have $\ell_k>\ell_\alpha$ and thus

$$\begin{aligned}\frac{(1+\ell_\alpha\lambda_0)(1+\ell_k\lambda_\alpha)(1+r)}{(1+\lambda_0)(1+\ell_\alpha\lambda_\alpha)}<\frac{(1+\ell_k\lambda_0)(1+r)}{1+\lambda_0}&\Leftrightarrow\frac{1+\ell_\alpha\lambda_0}{1+\ell_k\lambda_0}<\frac{1+\ell_\alpha\lambda_\alpha}{1+\ell_k\lambda_\alpha}\\&\Leftrightarrow\lambda_\alpha<\lambda_0.\end{aligned}$$

The new error dominates the old one. Therefore the new vector of errors E_α satisfies

$$\mathrm{E}_\alpha\lesssim\big(\eta^{\frac{(1+\ell_\alpha\lambda_0)(1+\ell_0\lambda_\alpha)(1+r)}{(1+\lambda_0)(1+\ell_\alpha\lambda_\alpha)}},0,\dots,\eta^{\frac{(1+\ell_\alpha\lambda_0)(1+\ell_k\lambda_\alpha)(1+r)}{(1+\lambda_0)(1+\ell_\alpha\lambda_\alpha)}},\dots\big).$$

We compare the new error on $k=0$ with the initial error on $k=0$ (see (4.2)). Notice that $\ell_0=1$. We have

$$\begin{aligned}\ell_0+r=1+r<\frac{(1+\ell_\alpha\lambda_0)(1+\lambda_\alpha)(1+r)}{(1+\lambda_0)(1+\ell_\alpha\lambda_\alpha)}\quad&\Leftrightarrow\quad\frac{1+\ell_\alpha\lambda_0}{1+\lambda_0}>\frac{1+\ell_\alpha\lambda_\alpha}{1+\lambda_\alpha}\\&\Leftrightarrow\quad\lambda_0>\lambda_\alpha.\end{aligned}$$

The new error is thus infinitesimal of higher order with respect to the old one.

Step 3: Inductive correction. Let $k\in\mathbb{N}$ be a fixed number. Assume that numbers $\lambda_0>\ldots>\lambda_k>0$ and numbers $\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_k>0$ are already chosen. To avoid pointless complications, we ignore the fact that for a finite set of k we have $L_k = \emptyset$. We define by induction

$$\varepsilon_{k+1} = d_k |\varepsilon_k|^{\frac{1+\ell_{k+1}\lambda_k}{1+\ell_{k+1}\lambda_{k+1}}},$$

where $d_k \neq 0$ are constants that are not of interest for us. In the following inequalities, we assume that $d_k = 1$. This is without loss of generality. The number ε_{k+1} will be the solution of our correcting equation. By our base construction for $k = 0$, we have ε_0 as in (4.5) and we find in closed form

$$\varepsilon_{k+1} \simeq \eta^{\frac{1+r}{1+\lambda_0} \prod_{h=1}^{k+1} \frac{1+\ell_h\lambda_{h-1}}{1+\ell_h\lambda_h}}. \tag{4.16}$$

Assume that position parameters $0 < \eta_0 < \ldots < \eta_k$ are given such that $\eta_k \leq \frac{1}{2}|\varepsilon_k|^{\lambda_k}$. We define the position parameter

$$\eta_{k+1} = 2|\varepsilon_k|^{\lambda_k} \simeq |\varepsilon_k|^{\lambda_k}.$$

In closed form, we have

$$\eta_{k+1} \simeq \eta^{\lambda_k \frac{1+r}{1+\lambda_0} \prod_{h=1}^{k} \frac{1+\ell_h\lambda_{h-1}}{1+\ell_h\lambda_h}}. \tag{4.17}$$

Finally, assume that, after correcting k, we have a vector of errors E_k such that:

$$\mathrm{E}_k \lesssim \big(\underbrace{\ldots, |\varepsilon_{h+1}|^{1+\ell_h\lambda_{h+1}}, \ldots}_{h=0,\ldots,k-1}, 0, \underbrace{\ldots, |\varepsilon_k|^{1+\ell_h\lambda_k}, \ldots}_{h=k+1,\ldots,K}\big). \tag{4.18}$$

Notice that the structure of the error is different when $h < k$ and when $h > k$. The index h refers to the position in the vector. The 0 is at position k.

We correct the error E_k^{k+1}, the $(k+1)$-th component of E_k, using the device $R_{\eta_{k+1},\lambda_{k+1}}(\varepsilon)$, where $0 < \lambda_{k+1} < \lambda_k$ and $\varepsilon \in \mathbb{R}$ has to be computed. We look for a solution $\varepsilon = \varepsilon_{k+1}$ such that

$$\frac{1}{2}|\varepsilon_{k+1}|^{\lambda_{k+1}} \geq \eta_{k+1}. \tag{4.19}$$

In this case, the equation $R^{k+1}_{\eta_{k+1},\lambda_{k+1}}(\varepsilon_{k+1}) + E_k^{k+1} = 0$ has a solution $\varepsilon = \varepsilon_{k+1}$ such that

$$|\varepsilon_{k+1}| \simeq |\varepsilon_k|^{\frac{1+\ell_{k+1}\lambda_k}{1+\ell_{k+1}\lambda_{k+1}}} \tag{4.20}$$

This solution satisfies (4.19).

The effect of $R_{\eta_{k+1},\lambda_{k+1}}(\varepsilon_{k+1})$ on $h \neq k+1$ is

$$R^h_{\eta_{k+1},\lambda_{k+1}}(\varepsilon_{k+1}) = |\varepsilon_{k+1}|^{1+\ell_h\lambda_{k+1}}.$$

We compare this effect with the errors in the vector (4.18). We deal first with the component $h > k+1$. In this case, the new error dominates the old one. In fact, by (4.20) we have

$$\begin{aligned}
|\varepsilon_{k+1}|^{1+\ell_h\lambda_{k+1}} > |\varepsilon_k|^{1+\ell_h\lambda_k} \quad &\Leftrightarrow \quad |\varepsilon_k|^{\frac{(1+\ell_{k+1}\lambda_k)(1+\ell_h\lambda_{k+1})}{1+\ell_{k+1}\lambda_{k+1}}} > |\varepsilon_k|^{1+\ell_h\lambda_k} \\
&\Leftrightarrow \quad \frac{1+\ell_h\lambda_{k+1}}{1+\ell_{k+1}\lambda_{k+1}} < \frac{1+\ell_h\lambda_k}{1+\ell_{k+1}\lambda_k} \\
&\Leftrightarrow \quad \lambda_{k+1} < \lambda_k.
\end{aligned}$$

The last equivalence holds because $\ell_h > \ell_{k+1}$ if $h > k+1$.

We claim that in the case $h < k$ the old error dominates the new one. Namely, for $h = 0, \dots, k-1$ we have

$$|\varepsilon_{k+1}|^{1+\ell_h\lambda_{k+1}} < |\varepsilon_{h+1}|^{1+\ell_h\lambda_{h+1}}.$$

This inequality follows from the fact that for $j = h+1, \dots, k$ we have

$$\begin{aligned}
|\varepsilon_{j+1}|^{1+\ell_h\lambda_{j+1}} < |\varepsilon_j|^{1+\ell_h\lambda_j} \quad &\Leftrightarrow \quad |\varepsilon_j|^{\frac{(1+\ell_h\lambda_{j+1})(1+\ell_{j+1}\lambda_j)}{1+\ell_{j+1}\lambda_{j+1}}} < |\varepsilon_j|^{1+\ell_h\lambda_j} \\
&\Leftrightarrow \quad \frac{1+\ell_h\lambda_{j+1}}{1+\ell_{j+1}\lambda_{j+1}} > \frac{1+\ell_h\lambda_j}{1+\ell_{j+1}\lambda_j} \\
&\Leftrightarrow \quad \lambda_{j+1} < \lambda_j.
\end{aligned}$$

In fact, $\ell_{j+1} > \ell_h$.

After the correction of $k+1$, we have a new vector of errors E_{k+1} that satisfies

$$\mathrm{E}_{k+1} \lesssim \big(\underbrace{\dots, |\varepsilon_{h+1}|^{1+\ell_h\lambda_{h+1}}, \dots}_{h=0,\dots,k}, 0, \underbrace{\dots, |\varepsilon_k|^{1+\ell_h\lambda_k}, \dots}_{h=k+2,\dots,K}\big). \tag{4.21}$$

This is the estimate in (4.18) for the step $k+1$. This ends the inductive construction.

Step 4: General decrease of errors. We started from the error $\mathrm{E}(\eta)$. In the first step, we corrected the error of $k = 0$, producing a new error on each $k \geq 1$. The new error dominates the error of the cut T_η. Then we corrected inductively the error on each $k = 1, \dots, K$, component by component, as described in Step 3. During the procedure, the error on a fixed h increases till k reaches h. When correcting $k > h$ the new error added on h is however negligible. These facts are clear from formula (4.18).

We deduce that the real error on k that has to be corrected is

$$R^k_{\eta_{k-1},\lambda_{k-1}}(\varepsilon_{k-1}) \simeq |\varepsilon_{k-1}|^{1+\ell_k\lambda_{k-1}}.$$

Based on this assessment, we define the *theoretical initial vector of errors* $\bar{\mathrm{E}}(\eta)$ produced by the cut T_η as

$$\bar{\mathrm{E}}(\eta) = \big(T^0_\eta, \underbrace{\dots, |\varepsilon_{k-1}|^{1+\ell_k\lambda_{k-1}}, \dots}_{k=1,\dots,K}\big).$$

We compare this error with the vector of errors after the correction of $k = K$, i.e., with

$$\mathrm{E}_K(\eta) = \mathrm{E}_K \simeq \big(\underbrace{\dots, |\varepsilon_{k+1}|^{1+\ell_k\lambda_{k+1}}, \dots}_{k=0,\dots,K-1}, 0\big). \tag{4.22}$$

We compare first the component $k = 0$. Notice that the first relevant k after $k = 0$ is $k = \alpha$. We have $T^0_\eta \simeq \eta^{1+r}$ and

$$E^0_K(\eta) \simeq |\varepsilon_\alpha|^{1+\lambda_\alpha} \simeq \eta^{\varrho_0(1+r)} \simeq \bar{E}^0(\eta)^{\beta_0}, \quad \text{where } \varrho_0 = \frac{(1+\ell_\alpha\lambda_0)(1+\lambda_\alpha)}{(1+\lambda_0)(1+\ell_\alpha\lambda_\alpha)} > 1.$$

Analogously, for any relevant $0 < k < K$ we have

$$E_K^k(\eta) \simeq \Big[|\varepsilon_{k-1}|^{1+\ell_k\lambda_{k-1}}\Big]^{\frac{(1+\ell_{k+1}\lambda_k)(1+\ell_k\lambda_{k+1})}{(1+\ell_k\lambda_k)(1+\ell_{k+1}\lambda_{k+1})}} \simeq \bar{E}^k(\eta)^{\varrho_k},$$

where the exponent ϱ_k is

$$\varrho_k = \frac{(1+\ell_{k+1}\lambda_k)(1+\ell_k\lambda_{k+1})}{(1+\ell_k\lambda_k)(1+\ell_{k+1}\lambda_{k+1})} > 1.$$

With the choice

$$\varrho = \min\big\{\varrho_0, \dots, \varrho_{K-1}\big\} > 1,$$

we have the following quantitative general decrease of errors

$$\mathrm{E}_K(\eta) \lesssim \bar{\mathrm{E}}(\eta^\varrho), \quad 0 < \eta < \delta, \tag{4.23}$$

for a suitable constant $0 < \delta < 1$.

In order to achieve this general decrease of errors, we paid a certain cost of length that is the sum of the cost of length of all the devices that have been used. As $|\varepsilon_K| \leq \ldots \leq |\varepsilon_0|$, (see (4.16)), the total cost is

$$\sum_{k=0}^{K} 2|\varepsilon_k| \simeq |\varepsilon_0| \simeq \eta^{\frac{1+r}{1+\lambda_0}}. \tag{4.24}$$

Remark 4.3. The sequence $(\varepsilon_k)_{k\in\mathbb{N}}$ is converging to 0 rather fastly, see (4.16). We could not understand whether there exists a fine choice of the parameters $\lambda_k > 0$ such that

$$\sum_{k=0}^{\infty} 2|\varepsilon_k| \simeq |\varepsilon_0|$$

This would permit us to drop Assumption 4.1.

Step 5: Iteration argument. We can iterate countably many times the construction described in the steps 1–3. We get in this way a correction R of the error $\mathrm{E}(\eta)$, i.e., an at most countable choice of devices that sets to zero all the coordinates of $\mathrm{E}(\eta)$. The cost of length of R is

$$\Lambda(R) \leq \sum_{h=0}^{\infty} \eta^{\frac{1+r}{1+\lambda_0}\varrho^h} \simeq \eta^{\frac{1+r}{1+\lambda_0}}.$$

Now the proof of the theorem can be concluded on choosing $\lambda_0 > 0$ small enough.

□

5. Equivalence classes with two elements. Homogeneous case

In this section, we correct an equivalence class with two elements. In the homogeneous case (i.e., $\varphi(t) = t^r$), the coefficients $c_{h\ell}$ in (3.7) satisfy $c_{h\ell} = 0$ for all $\ell > 0$ and thus $c_{ijh\ell} = 0$ for all $\ell > 0$. We shall adjust the notation introduced between (3.8) and (3.13) on letting

$$R^{ijh}_{\eta,\lambda}(\varepsilon) = R^{ijh0}_{\eta,\lambda}(\varepsilon), \quad Q^{ijh}_{b}(\varepsilon) = Q^{ijh0}_{b}(\varepsilon), \quad c_{ijh} = c_{ijh0}. \tag{5.1}$$

When no confusion arises, we shall drop η, λ, and b in our notation.

Assume that during a certain iterative correction, at a certain step we want to correct the error of the equivalence class L_k, for some k. Let us denote by $\mathcal{R}_k$ the set of all rectangles R of the type (3.4)–(3.5) with their parameters (that are omitted in our notation), which have been used in the correction of the first representatives of the equivalence classes $L_0, L_1, \dots, L_k$. Let us denote by $\mathcal{Q}_k$ the set of all squares Q of the type (3.10), with their parameters, which have been used in the correction of the equivalence classes $L_0, L_1, \dots, L_{k-1}$.

Assume that the error of the first representative $(i, j) \in L_k$, $j < \alpha$, was set to zero at the previous step. The total error on a pair $(i, j) \in L_k$ with, $j \geq \alpha$ is

$$E^{ij} = T^{ij}_\eta + \sum_{R\in\mathcal{R}_k} R^{ij} + \sum_{Q\in\mathcal{Q}_k} Q^{ij}. \tag{5.2}$$

This error is a function of the cut parameter η. According to the notation introduced in (3.8) and (3.12), we have

$$E^{ij} = T^{ij}_\eta + \sum_{h=0}^{j} c_{ijh} E^{ijh}, \quad \text{with} \quad E^{ijh} = \sum_{R\in\mathcal{R}_k} R^{ijh} + \sum_{Q\in\mathcal{Q}_k} Q^{ijh}. \tag{5.3}$$

From now on, we assume that L_k contains two elements, $L_k = \{(i_0, j_0), (i_1, j_1)\}$, with the notation (3.15). We let $i = i_0$ and $j = j_0$. In this section, we describe the procedure to correct the error of L_k. In the next section, we treat the case of an arbitrary equivalence class. We use two squares of the type (3.10). Let $\eta > 0$ be the cut parameter and let $b_q = a_q \eta^\mu$, $q = 0, 1$, be the position parameters of the squares, where $\mu > 0$ is a parameter that will be needed in the next sections to control the propagation of errors, and $\frac{1}{2} \leq a_q \leq 1$. The squares are

$$Q_{b_q}(\sigma_q) = \big\{(x_1, x_2) \in \mathbb{R}^2 : b_q < x_1 < b_q + |\sigma_q|,\, x_1^r < x_2 < x_1^r + |\sigma_q|\big\}, \quad q = 0, 1.$$

We have the system of two equations in the unknowns σ_q

$$\begin{cases} \displaystyle\sum_{q\in\{0,1\}} \sum_{h=0}^{j} c_{ijh} Q^{ijh}_{b_q}(\sigma_q) = 0 \\ \displaystyle\sum_{q\in\{0,1\}} \sum_{h=0}^{j_1} c_{i_1 j_1 h} Q^{i_1 j_1 h}_{b_q}(\sigma_q) + E^{i_1 j_1} = 0. \end{cases}$$

We multiply the first equation by $c_{i_1j_1j_1}$, we multiply the second equation by c_{ijj}, and we subtract the first equation from the second one. Using the identities (3.13), the system transforms into the equivalent system

$$\begin{cases} \displaystyle\sum_{q\in\{0,1\}} \sum_{h=0}^{j} c_{ijh} Q_{b_q}^{ijh}(\sigma_q) = 0 \\ \displaystyle\sum_{q\in\{0,1\}} \sum_{h=0}^{j_1-1} c'_{ijh} Q_{b_q}^{i_1j_1h}(\sigma_q) + c_{ijj} E^{i_1j_1} = 0, \end{cases} \tag{5.4}$$

where c'_{ijh} are explicit constants. Assume that the error in the left hand side of the second equation satisfies for some $s > 0$

$$E^{i_1j_1} \simeq \eta^s. \tag{5.5}$$

We call s the *structural exponent of* L_k. The exponent s for the first equivalence class with two elements will be computed later. The further exponents s are defined inductively in Section 7. We look for solutions σ_0, σ_1 to the system (5.4) satisfying the condition

$$\frac{1}{C}\eta^{\frac{1}{3}\{s-\mu(\ell_k-r-1)\}} \le |\sigma_q| \le C\eta^{\frac{1}{3}\{s-\mu(\ell_k-r-1)\}}, \tag{5.6}$$

where $C > 1$ is a suitably large constant. The reason for this restriction will be clear in (5.12). Recall that $\ell_k = i + rj + 1$. Now assume that the parameter μ satisfies

$$0 < \mu < \frac{1}{3}\big\{s - \mu(\ell_k - r - 1)\big\}. \tag{5.7}$$

Any $\mu > 0$ small enough satisfies this condition. If σ_q satisfies the constraint (5.6) and (5.7) holds, then we have

$$|\sigma_q| < \frac{1}{4}\eta^{\mu}, \tag{5.8}$$

as soon as $\eta > 0$ is small enough. Then we have the Taylor development

$$\begin{aligned} Q_{b_q}^{ijh}(\sigma_q) &= \operatorname{sgn}(\sigma_q)|\sigma_q|^{j+1-h}\Big[\big(a_q\eta^{\mu} + |\sigma_q|\big)^{i+hr+1} - \big(a_q\eta^{\mu}\big)^{i+hr+1}\Big] \\ &= \operatorname{sgn}(\sigma_q)(i+hr+1)a_q^{i+hr}\eta^{\mu(i+hr)}\big\{|\sigma_q|^{j+2-h} + o(|\sigma_q|^{j+2-h})\big\}. \end{aligned}$$

On the other hand, when σ_q is subject to the constraint (5.6), we have

$$\eta^{\mu(i+hr)}|\sigma_q|^{j+2-h} \simeq \eta^{\mu(i+hr)+\frac{1}{3}\{s-\mu(\ell_k-r-1)\}(j+2-h)}. \tag{5.9}$$

If we assume the following condition on μ

$$\mu r - \frac{1}{3}\{s - \mu(\ell_k - r - 1)\} < 0, \tag{5.10}$$

then the exponent of η in the right hand side of (5.9) is minimum when h is maximum. Condition (5.10) holds if μ is sufficiently close to 0 and implies condition (5.7). Thus the leading terms in the sums in h appearing in the two equations of system (5.4) are obtained on choosing $h = j$ and $h = j_1 - 1$, respectively. The system can therefore be approximated in the following way (we also assume $a_0 = 1$):

$$\begin{cases} \eta^{\mu(\ell_k-1)}\big\{\sigma_0|\sigma_0| + o(\sigma_0^2)\big\} + a_1^{\ell_k-1}\eta^{\mu(\ell_k-1)}\big\{\sigma_1|\sigma_1| + o(\sigma_1^2)\big\} = 0 \\ \eta^{\mu(\ell_k-r-1)}\big\{\sigma_0^3 + o(\sigma_0^3)\big\} + a_1^{\ell_k-r-1}\eta^{\mu(\ell_k-r-1)}\big\{\sigma_1^3 + o(\sigma_1^3)\big\} + c_1 E^{i_1 j_1} = 0, \end{cases} \tag{5.11}$$

where $c_1 \neq 0$ is a constant. From the first equation, we deduce that

$$\sigma_0 = -a_1^{(\ell_k-1)/2}\sigma_1 + o(\sigma_1).$$

Replacing this relation into the second equation, we obtain

$$\eta^{\mu(\ell_k-r-1)}\sigma_1^3 + o(\sigma_1^3) + c_2 E^{i_1 j_1} = 0,$$

where $c_2 \neq 0$ is a new constant. The error on the left hand side satisfies (5.5). Then, by the intermediate value theorem, this equation has a solution σ_1 subject to the constraint (5.6), provided that $C > 1$ is large enough (independently from η) and $\eta > 0$ is small enough. The solution σ_1 satisfies

$$\sigma_1 \simeq \eta^{\frac{1}{3}\{s-\mu(\ell_k-r-1)\}}. \tag{5.12}$$

The solution σ_0 satisfies the same estimate.

We specialize to the first equivalence class with two elements. The equivalence class L_k with $k = \alpha \cdot \beta$ contains exactly two elements, $L_k = \big\{(\beta, 0), (0, \alpha)\big\}$. In particular, we have $\bar{k} = 1$. We call L_k for $k = \alpha \cdot \beta$ the *leading equivalence class with two elements*. We compute the exponent s appearing in (5.5) in this case. The number s governs the inductive construction of Section 7.

Assume that we already corrected the errors of all first representatives, according to the procedure described in Section 4. By formula (5.2) with $\mathcal{Q}_K = \emptyset$, we have

$$E^{\beta 0} = T_\eta^{\beta 0} + \sum_{h=0}^{K} R_{\eta_h,\lambda_h}^{\beta 0}(\varepsilon_h) = T_\eta^{\beta 0} + \sum_{h=0}^{K} \varepsilon_h \Big[(\eta_h + |\varepsilon_h|^{\lambda_h})^{\beta+1} - \eta_h^{\beta+1}\Big] = 0. \tag{5.13}$$

After this correction, the total error on the pair $(0, \alpha)$ is precisely (recall the notation introduced in (3.8) and (5.1))

$$E^{0\alpha} = T_\eta^{0\alpha} + \sum_{h=0}^{K}\sum_{p=0}^{\alpha} c_{0\alpha p} R_{\eta_h,\lambda_h}^{0\alpha p}(\varepsilon_h). \tag{5.14}$$

By the formulas (3.13), $R_{\eta_h,\lambda_h}^{0\alpha\alpha}(\varepsilon_h)$ is a constant multiple of $R_{\eta_h,\lambda_h}^{\beta 0}(\varepsilon_h)$, by a universal proportionality constant. Also $T_\eta^{\beta 0}$ and $T_\eta^{0\alpha}$ are proportional, independently from η. Then we can subtract identity (5.13) from (5.14), to obtain

$$E^{0\alpha} = cT_\eta^{0\alpha} + \sum_{h=0}^{K}\sum_{p=0}^{\alpha-1} c_{0\alpha p}\varepsilon_h^{\alpha+1-p}\Big[(\eta_h + |\varepsilon_h|^{\lambda_h})^{pr+1} - \eta_h^{pr+1}\Big], \tag{5.15}$$

where $c \in \mathbb{R}$ is a constant (in fact, $c \neq 0$). By the results in Section 4, $T_\eta^{0\alpha}$ is dominated by the sum in (5.15). Moreover, by (4.19) we have

$$\varepsilon_h^{\alpha+1-p}\Big[(\eta_h + |\varepsilon_h|^{\lambda_h})^{pr+1} - \eta_h^{pr+1}\Big] \simeq \varepsilon_h^{\alpha+1-p}|\varepsilon_h|^{\lambda_h(pr+1)},$$

and the quantity is maximum when p is maximum, i.e., $p = \alpha - 1$. Thus we obtain the estimate for the error

$$E^{0\alpha} \simeq \sum_{h=0}^{K} |\varepsilon_h|^{2+(\ell_k - r)\lambda_h}.$$

We compute the largest term in the previous sum. We have the elementary equivalences (when η is small enough)

$$\begin{aligned} |\varepsilon_{h+1}|^{2+(\ell_k-r)\lambda_{h+1}} < |\varepsilon_h|^{2+(\ell_k-r)\lambda_h} &\Leftrightarrow \frac{1+\ell_{h+1}\lambda_h}{2+(\ell_k-r)\lambda_h} > \frac{1+\ell_{h+1}\lambda_{h+1}}{2+(\ell_k-r)\lambda_{h+1}} \\ &\Leftrightarrow \ell_k - r < 2\ell_{h+1}. \end{aligned}$$

Let $\widehat{k} = \min\{h \in \mathbb{N} : h \le K-1,\ \ell_k - r < 2\ell_{h+1}\}$. Then we have $E^{0\alpha} \simeq |\varepsilon_{\widehat{k}}|^{2+(\ell_k-r)\lambda_{\widehat{k}}}$ and, by (4.16), we obtain $E^{0\alpha} \simeq \eta^s$ where

$$s = (1+r)\frac{2+(\ell_k-r)\lambda_{\widehat{k}}}{1+\ell_{\widehat{k}}\lambda_{\widehat{k}}}\prod_{h=1}^{\widehat{k}-1}\frac{1+\ell_{h+1}\lambda_h}{1+\ell_h\lambda_h} > (1+r)\frac{2+(\ell_k-r)\lambda_{\bar{k}}}{1+\ell_{\bar{k}}\lambda_{\bar{k}}}. \tag{5.16}$$

We call the number s the *leading structural exponent.*

We estimate the cost of length of the correction. Let μ_k be a number satisfying (5.7), i.e.,

$$0 < \mu_k < \frac{1}{3}\big\{s - \mu_k(\ell_k - r - 1)\big\},$$

and let $b_q = a_q\eta^{\mu_k}$, $q = 0, 1$, be position parameters. We correct the error of the class L_k, i.e., the pair $E^{\beta 0} = 0$ and $E^{0\alpha}$, with the pair of squares $Q_{b_0}(\sigma_0), Q_{b_1}(\sigma_1)$. The procedure is explained above. The solutions σ_0, σ_1 satisfy (5.12), and namely,

$$\sigma_0 \simeq \sigma_1 \simeq \eta^{\frac{1}{3}\{s-\mu_k(\ell_k-r-1)\}}.$$

The cost of length of the correction is

$$\Lambda\big(Q_{b_0}(\sigma_0) \cup Q_{b_1}(\sigma_1)\big) = 4|\sigma_0| + 4|\sigma_1| \simeq \eta^{\frac{1}{3}\{s-\mu_k(\ell_k-r-1)\}},$$

where s is the leading structural exponent in (5.16). Setting all the parameters λ_k and μ_k equal to zero, we realize from (5.16) that

$$\Lambda\big(Q_{b_0}(\sigma_0) \cup Q_{b_1}(\sigma_1)\big) \sim \eta^{\frac{2}{3}(1+r)}. \tag{5.17}$$

This is the cost of length for the correction of the first equivalence class with two elements.

6. Generic equivalence class. Homogeneous case

In this section, we generalize the construction of Section 5 to the case of system of any dimension. We correct the error of the equivalence class

$$L_k = \big\{(i - p\beta, j + p\alpha) \in \mathbb{N} \times \mathbb{N} : p = 0, 1, \dots, \bar{k}\big\}$$

by means of $\bar{k} + 1$ squares of the type (3.10)

$$Q_{b_q}(\sigma_q) = \big\{(x_1, x_2) \in \mathbb{R}^2 : b_q < x_1 < b_q + |\sigma_q|,\ x_1^r < x_2 < x_1^r + |\sigma_q|\big\},$$

where $q = 0, 1, \dots, \bar{k}$. Here, $\sigma_q \in \mathbb{R}$ are size parameters and b_q are position parameters of the form $b_q = a_q \eta^\mu$, $q = 0, 1, \dots, \bar{k}$, where $0 < \mu < 1$ is to be fixed later in the iterative argument and $\frac{1}{2} \leq a_0, a_1, \dots, a_{\bar{k}} \leq 1$ are parameters that will be fixed in a suitable way.

In this section, we solve the following nonlinear system of equations

$$\sum_{q=0}^{\bar{k}} \sum_{h=0}^{j_p} c_{i_p j_p h} Q_{b_q}^{i_p j_p h}(\sigma_q) + E^{i_p j_p} = 0, \quad p = 0, 1, \dots, \bar{k}. \tag{6.1}$$

The unknowns are $\sigma_0 \dots, \sigma_{\bar{k}}$. The errors $E^{i_p j_p}$ have the structure (5.2). Recall our notation (3.12). As above, we let $i_p = i - p\beta$ and $j_p = j + p\alpha$, so that $i = i_0$ and $j = j_0$. We may assume $E^{ij} = 0$.

Now we transform the system (6.1) into a new equivalent system exploiting the cancellations (3.13). The procedure produces errors with a new structure. We shall call them *reduced errors*. We perform the following operations:

(1) We multiply by $c_{i_p j_p j_p}$ the first equation (i.e., the equation with $p = 0$), we multiply the p-equation by c_{ijj}, and we subtract the resulting equations. By the identities (3.13), the p-equation with $p \geq 1$ transforms into the following:

$$\sum_{q=0}^{\bar{k}} \sum_{h=0}^{j_p - 1} c_{i_p j_p h} Q_{b_q}^{i_p j_p h}(\sigma_q) + c_{ijj} E^{i_p j_p} - c_{i_p j_p j_p} E^{ij} = 0, \quad p = 1, 2, \dots, \bar{k}, \tag{6.2}$$

where $c_{i_p j_p h}$ are new constants.

(2) We multiply by $c_{i_p j_p j_p}$, $p \geq 2$, the second equation in (6.2) (i.e., the equation with $p = 1$), we multiply by $c_{i_1 j_1 j_1}$ the p-equation in (6.2), and we subtract the resulting equations. We use again the identities (3.13).

(3) We repeat $\bar{k}$ times the procedure described in (1) and (2). Eventually, the system of equations (6.1) transforms into the equivalent system of equations

$$\sum_{q=0}^{\bar{k}} \sum_{h=0}^{j_p - p} c'_{ijph} Q_{b_q}^{i_p j_p h}(\sigma_q) + \sum_{q=0}^{p} d_{ijq} E^{i_q j_q} = 0, \quad p = 0, 1, \dots, \bar{k}, \tag{6.3}$$

where c'_{ijph} and d_{ijq} are suitable constants that can be computed.

We call the new errors

$$F^{i_p j_p} = \sum_{q=0}^{p} d_{ijq} E^{i_q j_q} \tag{6.4}$$

the *reduced errors of the equivalence class* L_k. The sums defining $F^{i_p j_p}$ enjoy the same cancellation properties described above, by (3.13) and (5.3).

Now let us assume that the *leading reduced error* $F^{i_{\bar{k}} j_{\bar{k}}}$ satisfies the following estimate for some exponent $s > 0$:

$$F^{i_{\bar{k}} j_{\bar{k}}} \simeq \eta^s. \tag{6.5}$$

We look for solutions σ_q, $q=0,1,\dots,\bar k$, to the system (6.3) satisfying for a suitable constant $C>1$ independent of η the condition

$$C^{-1}\eta^{\frac{1}{\bar k+2}\{s-\mu(\ell_k-\bar kr-1)\}}\le|\sigma_q|\le C\eta^{\frac{1}{\bar k+2}\{s-\mu(\ell_k-\bar kr-1)\}},\quad q=0,1,\dots,\bar k. \tag{6.6}$$

The reason for this restriction will be clear later. The parameter $\mu>0$ is chosen such that

$$0<\mu<\frac{1}{\bar k+2}\{s-\mu(\ell_k-\bar kr-1)\}. \tag{6.7}$$

Any small enough $\mu>0$ satisfies (6.7). If σ_q satisfies (6.6) and (6.7) holds, then we have

$$|\sigma_q|\le\frac14\eta^\mu, \tag{6.8}$$

as soon as $\eta>0$ is small enough. Then we have the Taylor development (as $\eta\to0$)

$$Q^{ijh}_{b_q}(\sigma_q)=(i+hr+1)a_q^{i+hr}\eta^{\mu(i+hr)}\mathrm{sgn}(\sigma_q)|\sigma_q|^{j-h+2}+o(\eta^{\mu(i+hr)}\sigma_q^{j-h+2}).$$

On the other hand, when σ_q is subject to the constraint (6.6) we have

$$\eta^{\mu(i+hr)}|\sigma_q|^{j-h+2}\simeq\eta^{\mu(i+hr)+\frac{1}{\bar k+2}\{s-\mu(\ell_k-\bar kr-1)\}(j-h+2)}. \tag{6.9}$$

Under the following condition on μ

$$\mu r-\frac{1}{\bar k+2}\big\{s-\mu(\ell_k-\bar kr-1)\big\}<0,$$

that holds for all $\mu>0$ sufficiently close to 0 and implies (6.7), the exponent of η in the right hand side of (6.9) is minimum when h is maximum. Thus the leading term in the sum in h appearing in (6.3) is obtained on chosing $h=j_p-p$. The system may then be approximated in the following way:

$$\eta^{\mu(\ell_k1+pr)}\Big\{\sum_{q=0}^{\bar k}a_q^{\ell_k-1-pr}\mathrm{sgn}(\sigma_q)|\sigma_q|^{p+2}+o(\sigma_q^{p+2})\Big\}+c_{ijp}F^{i_pj_p}=0,\quad p=0,1,\dots,\bar k. \tag{6.10}$$

Above, c_{ijp} are explicit constants.

We prove existence of solutions $\sigma_0,\sigma_1,\dots,\sigma_{\bar k}$ of system (6.10) by an inductive argument on $\bar k\ge1$. The errors $F^{i_pj_p}$ are assumed to satisfy the following structural assumptions: there exist $h\in\mathbb N$, $\widetilde\mu>0$, and $\widetilde s>0$ such that

$$F^{i_pj_p}\simeq\eta^{(\ell_k-pr-1)\widetilde\mu+\frac{p+2}{\bar h+2}\{\widetilde s-\widetilde\mu(\ell_h-\bar hr-1)\}},\quad p=0,1,\dots,\bar k. \tag{6.11}$$

Above, $\bar h=[h/\beta]$ as in (3.14). Assumption (6.11) is motivated by the structure (5.3) of the errors. In particular, by (6.5), s and $\widetilde s$ are related through the relation

$$s=(\ell_k-\bar kr-1)\widetilde\mu+\frac{\bar k+2}{\bar h+2}\big\{\widetilde s-\widetilde\mu(\ell_h-\bar hr-1)\big\}. \tag{6.12}$$

This relation describes the transformation of structural exponents, and, eventually, governs the propagation of errors between equivalence classes. This will be the motivation for the introduction of the function Δ in (7.10).

Proposition 6.1. Assume the reduced errors $F^{i_p j_p}$ satisfy the structural assumptions (6.11). Then there exist position parameters $a_0, a_1, \dots, a_{\bar{k}}$ such that the system (6.10) has solutions $\sigma_0, \sigma_1, \dots, \sigma_{\bar{k}}$ satisfying (6.6).

Proof. When $\bar{k} = 1$, i.e., when we have a system of two equations, the claim is proved in Section 5. Assume the claim holds for all $\bar{h} = 1, 2, \dots, \bar{k} - 1$.

We solve system (6.10) on considering $\sigma_{\bar{k}}$ as a fixed parameter satisfying (6.6) with $q = \bar{k}$. We may rewrite the first $\bar{k}$ equations in (6.10) as

$$\eta^{\mu(\ell_k - pr - 1)} \Big\{ \sum_{q=0}^{\bar{k}-1} a_q^{\ell_k - pr - 1} \mathrm{sgn}(\sigma_q) |\sigma_q|^{p+2} + o(\sigma_q^{p+2}) \Big\} + \widetilde{F}^{i_p j_p} = 0, \quad p = 0, 1, \dots, \bar{k} - 1, \tag{6.13}$$

where, up to constants,

$$\widetilde{F}^{i_p j_p} = F^{i_p j_p} + \eta^{\mu(\ell_k - pr - 1)} \Big\{ a_{\bar{k}}^{\ell_k - \bar{k}r - 1} \mathrm{sgn}(\sigma_{\bar{k}}) |\sigma_{\bar{k}}|^{p+2} + o(\sigma_{\bar{k}}^{\bar{k}+2}) \Big\}, \quad p = 0, 1, \dots, \bar{k} - 1.$$

We claim that we have

$$\widetilde{F}^{i_p j_p} \simeq \eta^{\mu(\ell_k - pr - 1) + \frac{p+2}{\bar{k}+2}\{s - \mu(\ell_k - \bar{k}r - 1)\}}, \quad p = 0, 1, \dots \bar{k} - 1. \tag{6.14}$$

Statement (6.14) follows from (6.11) and from the inequality

$$\mu(\ell_k - pr - 1) + \frac{p+2}{\bar{k}+2} \big\{ s - \mu(\ell_k - \bar{k}r - 1) \big\} < \widetilde{\mu}(\ell_k - pr - 1) + \frac{p+2}{\bar{h}+2} \big\{ \widetilde{s} - \widetilde{\mu}(\ell_h - \bar{h}r - 1) \big\},$$

that, by (6.12), is equivalent to the inequality

$$\frac{\ell_k - pr - 1}{p+2} > \frac{\ell_k - \bar{k}r - 1}{\bar{k}+2},$$

that holds for all $p = 0, 1, \dots, \bar{k} - 1$.

By induction, there exist position parameters $a_0, a_1, \dots, a_{\bar{k}-1}$ such that the system (6.13) has solutions $\sigma_0, \sigma_1, \dots, \sigma_{\bar{k}-1}$ (depending on $\sigma_{\bar{k}}$). Moreover, since we have

$$F^{i_{\bar{k}-1} j_{\bar{k}-1}} \simeq \eta^{\bar{s}} \quad \text{with} \quad \bar{s} = \mu(\ell_k - (\bar{k} - 1)r - 1) + \frac{\bar{k}+1}{\bar{k}+2} \big\{ s - \mu(\ell_k - \bar{k}r - 1) \big\},$$

the solutions $\sigma_0, \sigma_1, \dots, \sigma_{\bar{k}-1}$ satisfy (6.6) with $\bar{k} - 1$ replacing $\bar{k}$ and with $\bar{s}$ replacing s, i.e.,

$$|\sigma_q| \simeq \eta^{\frac{1}{\bar{k}+1}\{\bar{s} - \mu(\ell_k(1 - (\bar{k}-1)r)\}} = \eta^{\frac{1}{\bar{k}+2}\{s - \mu(\ell_k - \bar{k}r - 1)\}}, \quad q = 0, 1, \dots, \bar{k} - 1.$$

Equivalently, σ_q satisfies (6.6) itself. In particular, we have $\sigma_q \simeq \sigma_{\bar{k}}$ for all $q = 0, 1, \dots, \bar{k} - 1$. The solutions $\sigma_0, \sigma_1, \dots, \sigma_{\bar{k}-1}$ depend on $\sigma_{\bar{k}}$. We insert these functions of $\sigma_{\bar{k}}$ into the last equation of system (6.10). The leading term in $\sigma_{\bar{k}}$ is the power $\sigma_{\bar{k}}^{\bar{k}+2}$. With a suitable choice of the position parameter $a_{\bar{k}}$, the coefficient of $\sigma_{\bar{k}}^{\bar{k}+2}$ is nonzero. Then the $(\bar{k}+1)$-th equation in (6.10) is of the form

$$\eta^{\mu(\ell_k - \bar{k}r - 1)} \Big\{ \mathrm{sgn}(\sigma_{\bar{k}}) |\sigma_{\bar{k}}|^{\bar{k}+2} + o(\sigma_{\bar{k}}^{\bar{k}+2}) \Big\} + c F^{i_{\bar{k}} j_{\bar{k}}} = 0, \tag{6.15}$$

where $c \neq 0$ is a constant. By our assumptions (6.11)-(6.12) with $p = \bar{k}$, equation (6.15) has a solution $\sigma_{\bar{k}}$ satisfying (6.6). This follows by an easy continuity argument. This ends the proof.

Let us notice here for future reference that the cost of lenght of the correction of the equivalence class L_k is

$$\Lambda\Big(\bigcup_{q=0}^{\bar{k}} Q_{b_q}(\sigma_q)\Big) = 4\sum_{q=0}^{\bar{k}} |\sigma_q| \simeq \eta^{\frac{1}{\bar{k}+2}\{s-\mu(\ell_k-\bar{k}r-1)\}}, \tag{6.16}$$

where s is the structural exponent of the equivalence class L_k given by (6.12).

□

7. Iterative correction of all equivalence classes. Homogeneous case

In this section, we iterate the construction of Section 6. At the same time, we take into account the procedure of Section 4.

For $k \in \mathbb{N}$, let $\bar{k}$ be the number in (3.14). The equivalence class L_k has $\bar{k}+1$ distinct elements, as in (3.15). An *error of* L_k is a $\bar{k}+1$-dimensional vector $\mathrm{v} = (v_0, v_1, \dots, v_{\bar{k}})$ of real numbers. The number v_0 is the error of the first representative of L_k, etc.. The index k ranges from 0 to a fixed integer $K \in \mathbb{N}$. The *total space of errors* is denoted by

$$\mathcal{E} = \mathbb{R}^{\bar{0}+1} \times \mathbb{R}^{\bar{1}+1} \times \ldots \times \mathbb{R}^{\bar{K}+1}.$$

Let us denote a generic element of $\mathcal{E}$ by $\mathrm{v} = (\mathrm{v}^0, \mathrm{v}^1, \dots, \mathrm{v}^K)$, with $\mathrm{v}^k \in \mathbb{R}^{\bar{k}+1}$. By (3.3), the *initial error produced by the cut* T_η is

$$E(\eta) = \mathrm{v}, \quad \text{with } \mathrm{v}^k = (c_0\eta^{\ell_k+r}, c_1\eta^{\ell_k+r}, \dots, c_{\bar{k}}\eta^{\ell_k+r}), \tag{7.1}$$

where $c_0, c_1, \dots, c_{\bar{k}}$ are constants depending on k. A *correction of* $E(\eta)$ is a countable union $R = \bigcup_{h\in\mathbb{N}} R_h$ of rectangles of the type (3.4)-(3.5) and a countable union $Q = \bigcup_{h\in\mathbb{N}} Q_h$ of squares of the type (3.10) that set to zero all the components of $E(\eta)$. The *cost of length* $\Lambda(R \cup Q)$ of the correction is

$$\Lambda(R\cup Q) = \sum_{h\in\mathbb{N}} \Lambda(R_h) + \Lambda(Q_h).$$

In this section, we make the following structural assumption. Let $k_0 = \min\{k \in \mathbb{N} : \bar{k} = 1\} = \alpha \cdot \beta$.

Assumption 7.1. The function $\varphi : \{k_0, k_0+1, \dots, K\} \to \mathbb{Q}$

$$\varphi(k) = \frac{\ell_k - \bar{k}r + 1}{\bar{k}+2} \text{ is injective.} \tag{7.2}$$

Remark 7.2. Notice that φ is injective when restricted to indices k with the same $\bar{k}$. Namely, by the monotonicity of ℓ_k, if $\bar{k} = \bar{h}$ then there holds $\varphi(k) < \varphi(h)$ if and only if $k < h$.

In general, it may happen that $\varphi(h) = \varphi(k)$ for some $h \neq k$ (and $\bar{h} \neq \bar{k}$). The construction of Theorem 7.3 does not work in this case, because there is a loop of errors between the equivalence classes L_h and L_k. Dropping Assumption 7.1 requires new more refined ideas.

Theorem 7.3. *Assume* (7.2). *For any* $\varepsilon > 0$ *there are numbers* $C > 0$ *and* $0 < \delta < 1$ *such that for any* $0 < \eta < \delta$ *there is a correction* $R \cup Q$ *of the initial error* $\mathrm{E}(\eta)$ *in* (7.1) *with cost of length satisfying*

$$\Lambda(R \cup Q) \leq C\eta^{\frac{2}{3}(1+r)-\varepsilon}. \tag{7.3}$$

Proof. For $k = k_0, k_0+1, \ldots, K$ we will fix parameters $\mu_k > 0$ satisfying various smallness conditions and strictly ordered in the following way:

$$\mu_k > \mu_h \quad \Leftrightarrow \quad \varphi(k) < \varphi(h). \tag{7.4}$$

This is possible by our Assumption 7.1.

Now we divide the argument in several steps.

Step 1. We correct the first representatives as described in Section 4. In this step, we only use rectangles of the type (3.4)-(3.5) depending on parameters $\lambda_0, \lambda_1, \ldots, \lambda_K$.

Step 2. We correct the leading equivalence class with two elements, as described in Section 5. In this case, we have $k = k_0 = \alpha \cdot \beta$. The structural exponent s appearing in (5.5) is computed in (5.16). Notice that we have

$$s \sim 2(1+r), \tag{7.5}$$

where $\sim$ means that we set to 0 the parameters λ_h appearing in (5.5). The parameter $\mu_k > 0$ is suitably small. This choice is made after we fixed the parameters $\lambda_0, \lambda_1, \ldots, \lambda_K$.

Step 3. We prove that the additional errors on first representatives produced in Step 2 are negligible for the procedure used in Step 1.

Let $h > k = \alpha \cdot \beta$, $(i,j) \in L_h$, let $b = \eta^{\mu_k}$, and $\sigma = \eta^{\frac{1}{3}\{s-\mu_k(\ell_k-r-1)\}}$. The effect $Q_b^{ij}(\sigma)$ of the square $Q_b(\sigma)$ on the pair (i,j) is

$$Q_b^{ij}(\sigma) \simeq \sigma^2 \eta^{\mu_k(i+jr)} = \eta^{\frac{2}{3}\{s-\mu_k(\ell_k-r-1)\}+\mu_k(\ell_h-1)}.$$

This formula can be obtained by the argument contained between (5.6) and (5.9).

Assume that (i,j) is the first representative of the class L_h. The error produced on (i,j) by the correction of first representatives can be found in line (4.22) and is

$$|\varepsilon_{h+1}|^{1+\ell_h\lambda_{h+1}} = \eta^d, \quad d = (1+r)\frac{1+\ell_h\lambda_{h+1}}{1+\lambda_0}\prod_{p=1}^{h+1}\frac{1+\ell_p\lambda_{p-1}}{1+\ell_p\lambda_p} \sim 1+r.$$

On the other hand, by (7.5) we have

$$\frac{2}{3}\{s-\mu_k(\ell_k-r-1)\}+\mu_k(\ell_h-1) \sim \frac{4}{3}(1+r)+\mu_k\big\{\ell_h-1-\frac{2}{3}(\ell_k-r-1)\big\}.$$

It follows that if μ_k is small enough, then the error produced on (i,j) by the procedure of Step 2 is negligible.

Step 4. We claim that the reduced error produced by the squares used in Step 2 dominates the reduced error produced by the rectangles used for the correction of first representatives in Step 1.

Let us first develop some general formulas. For fixed $k,h\in\mathbb{N}$, $\mu>0$, and $s>0$, let $b=\eta^{\mu}$ and $\sigma=\eta^{\frac{1}{\bar{k}+2}\{s-\mu(\ell_k-\bar{k}r-1)\}}$. We know from (6.6) and Proposition 6.1 that this is the size of the parameters σ for the correction of L_k starting from the parameters s and μ. By (3.12), the effect of the square $Q_b(\sigma)$ on the pair $(i_p,j_p)=(i-p\beta,j+p\alpha)\in L_h$, $p\in\{1,\dots,\bar{h}\}$, is

$$Q_b^{i_pj_p}(\sigma)=\sum_{q=0}^{j_p}c_{i_pj_pq}Q_b^{i_pj_pq}(\sigma).$$

Let us determine the reduced error. When we correct the equivalence class L_h, we exploit several cancellations in the sum for $Q_b^{i_pj_p}(\sigma)$, as described in Section 6. The leading term after the cancellations is obtained on taking the index $q=j_p-p$, and namely

$$\begin{aligned}Q_b^{i_pj_p,j_p-p}(\sigma)&=\operatorname{sgn}(\sigma)|\sigma|^{p+1}\big[(b+|\sigma|)^{\ell_h-pr}-b^{\ell_h-pr}\big]\\&\simeq\eta^{\frac{p+2}{\bar{k}+2}\{s-\mu(\ell_k-\bar{k}r-1)\}+\mu(\ell_h-pr-1)}.\end{aligned}\tag{7.6}$$

By (3.8), the effect onto the pair (i_p,j_p) of the generic rectangle $R_{\eta_q,\lambda_q}(\varepsilon_q)$ used in the correction of first representatives is

$$R_{\eta_q\lambda_q}^{i_pj_p}(\varepsilon_q)=\sum_{u=0}^{j_p}c_{i_pj_pu}R_{\eta_q\lambda_q}^{i_pj_pu}(\varepsilon_q).$$

The parameters $\eta_q,\lambda_q,\varepsilon_q$ are defined in Section 4.

The leading term after the cancellations is obtained on taking $u=j_p-p$, and, namely, by (4.16) we have

$$\begin{aligned}R_{\eta_q\lambda_q}^{i_pj_p,j_p-p}(\varepsilon_q)&=\varepsilon_q^{p+1}\big[(\eta_q+|\varepsilon_q|^{\lambda_q})^{\ell_h-pr}-\eta_q^{\ell_h-pr}\big]\\&\simeq\eta^{\frac{1+r}{1+\lambda_0}[p+1+\lambda_q(\ell_h-pr)]\prod_{u=1}^{q}\frac{1+\ell_u\lambda_{u-1}}{1+\ell_u\lambda_u}}.\end{aligned}\tag{7.7}$$

We compare (7.6) and (7.7) when $k=\alpha\cdot\beta$, so that $\bar{k}=1$, $h>k$, and the exponent s is given by (5.16):

$$s=(1+r)\frac{(2+(\ell_k-r)\lambda_{\widehat{k}})}{1+\ell_{\widehat{k}}\lambda_{\widehat{k}}}\prod_{u=1}^{\widehat{k}-1}\frac{1+\ell_{u+1}\lambda_u}{1+\ell_u\lambda_u},$$

where $\widehat{k}$ is an integer such that $0\le\widehat{k}\le k$. We claim that for a suitable choice of the parameters $\lambda_0,\lambda_1,\dots,\lambda_k$, and $\mu>0$, there holds

$$Q_b^{i_pj_p,j_p-p}(\sigma)\ge R_{\eta_q\lambda_q}^{i_pj_p,j_p-p}(\varepsilon_q).$$

For small $\eta > 0$, this is equivalent to the inequality $A < B$, where

$$\begin{aligned} A &:= \frac{p+2}{3}\big\{s - \mu(\ell_k - r - 1)\big\} + \mu(\ell_h - pr - 1) \\ B &:= \frac{1+r}{1+\lambda_0}\{p + 1 + \lambda_q(\ell_h - pr)\}\prod_{u=1}^{q}\frac{1+\ell_u\lambda_{u-1}}{1+\ell_u\lambda_u}. \end{aligned} \tag{7.8}$$

For $\lambda_0 = \ldots = \lambda_k = 0$ and $\mu = 0$, we have

$$A \sim \frac{2}{3}(p+2)(r+1) \quad \text{and} \quad B \sim (p+1)(r+1).$$

Thus, when $p > 1$, we have $A < B$ as soon as $\lambda_0, \ldots, \lambda_k, \mu > 0$ are small enough. When $p = 1$ we have $A \sim B$ and we argue as follows. In this case, inequality $A < B$ reads as

$$s + \mu(\ell_h - \ell_k) < \frac{1+r}{1+\lambda_0}\{2 + \lambda_q(\ell_h - r)\}\prod_{u=1}^{q}\frac{1+\ell_u\lambda_{u-1}}{1+\ell_u\lambda_u}. \tag{7.9}$$

In the case $\mu = 0$ and $\lambda_0 = \lambda_1 = \ldots = \lambda_k = \lambda > 0$, inequality (7.9) holds. In fact, in this case we have

$$\frac{(1+r)(2+\lambda(\ell_k - r))}{1+\ell_1\lambda} < \frac{(1+r)(2+\lambda(\ell_h - r))}{1+\lambda},$$

because $\ell_k < \ell_h$ and $\ell_1 \geq 1$. By continuity, (7.9) holds for some $\lambda_k < \lambda_{k-1} < \ldots < \lambda_0$ and $\mu > 0$.

Step 5. Let us define the function $\Delta : \mathbb{N} \times \mathbb{N} \times \mathbb{Q} \to \mathbb{Q}$

$$\Delta(k, h; s) = (\bar{h} + 2)\Big\{\frac{1}{\bar{k}+2}s + \mu_k\big(\varphi(h) - \varphi(k)\big)\Big\}. \tag{7.10}$$

This function has the following meaning. If the reduced error of the equivalence class L_k has size η^s, then the correction of the equivalence class L_k produces a reduced error of the size $\eta^{\Delta(k,h;s)}$ on the equivalence class L_h. See (7.6) with $p = \bar{h}$ and $\mu = \mu_k$. We claim that for all $k, h, u \in \mathbb{N}$ such that $k \neq h$ there holds

$$\Delta\big(h, u; \Delta(k, h; s)\big) > \Delta(k, u; s). \tag{7.11}$$

The proof is an elementary computation. In fact, we have

$$\Delta\big(h, u; \Delta(k, h; s)\big) = (\bar{u} + 2)\Big\{\frac{1}{\bar{k}+2}s + \mu_k\big(\varphi(h) - \varphi(k)\big) + \mu_h\big(\varphi(u) - \varphi(h)\big)\Big\},$$

and inequality (7.11) is equivalent with

$$\big(\varphi(h) - \varphi(k)\big)(\mu_k - \mu_h) > 0,$$

that holds by Assumption 7.1 and by the choice (7.4) of the parameters μ_k, μ_h.

Step 6. We set up the inductive step of the iterative argument. After the correction of a number of equivalence classes and after the cancellations described in Section 6, the structure of the error of a generic equivalence class L_k is determined by the size of the reduced error of the element of L_k in (3.15) with maximal p, i.e., with $p = \bar{k}$. This error is a power of η with structural exponent $s > 0$ as in (6.5). The number

s determines the entire vector of reduced errors of L_k, as in (6.11)–(6.12). For each equivalence class, one real number suffices to describe the full reduced error. Thus let us define the *total space of reduced errors* as $\mathcal{F} = \mathbb{R}^{K-k_0+1}$. We denote the components of a vector $\mathrm{F} \in \mathcal{F}$ of reduced errors as $\mathrm{F} = (\mathrm{F}^{k_0}, \mathrm{F}^{k_0+1}, \dots, \mathrm{F}^K)$, i.e., the enumeration starts from k_0. We need not consider equivalence classes with one element.

After the correction of all first representatives and of the leading equivalence class with two elements L_{k_0}, the total reduced error F_{k_0} satisfies

$$\mathrm{F}_{k_0} \simeq \big(0, \underbrace{\dots, \eta^{\Delta(k_0,h;s)}, \dots}_{h=k_0+1,\dots,K}\big). \tag{7.12}$$

This follows from Step 4. From now on, s is the leading structural exponent defined in (5.16).

Now we proceed iteratively as follows: we correct the error of all first representatives and the error of the equivalence class L_k for each $k \geq k_0 + 1$. A new error appears on the equivalence classes L_h with $h \neq k$.

We make this procedure quantitative. We claim that after $k - k_0 + 1$ steps, i.e., after the correction of L_k, the total reduced error F_k satisfies

$$\mathrm{F}_k \lesssim \Big(\underbrace{\dots, \sum_{u=h+1}^{k} \eta^{\Delta(u,h;\Delta(k_0,u;s))}, \dots}_{h=k_0,\dots,k-1}, 0, \underbrace{\dots, \eta^{\Delta(k_0,h;s)}, \dots}_{h=k+1,\dots,K}\Big). \tag{7.13}$$

We argue by induction. When $k = k_0$, the claim holds as noted in (7.12). Assume that (7.13) holds for k. We check it for $k + 1$.

We correct the error of the equivalence class L_{k+1} with the procedure of Section 6. The structural exponent in (6.5) is $\widetilde{s} = \Delta(k_0, k + 1; s)$. The new reduced error produced on L_h with $h \neq k + 1$ has the size $\eta^{\Delta(k+1,h;\widetilde{s})}$. By the argument of Step 5, see (7.11), we have

$$\Delta(k + 1, h; \Delta(k_0, k + 1; s)) > \Delta(k_0, h; s).$$

Therefore, for $h > k + 1$ there is no change in the structural exponent of L_h. After the correction of L_{k+1}, the new vector F_{k+1} of reduced errors satisfies

$$\mathrm{F}_{k+1} \lesssim \Big(\underbrace{\dots, \sum_{u=h+1}^{k+1} \eta^{\Delta(u,h;\Delta(k_0,u;s))}, \dots}_{h=k_0,\dots,k}, 0, \underbrace{\dots, \eta^{\Delta(k_0,h;s)}, \dots}_{h=k+2,\dots,K}\Big).$$

This is (7.13) for $k + 1$.

Step 7. (General decrease of errors) After the correction of the last equivalence class with $k = K$, the total reduced error $\mathrm{F}_K(\eta) = \mathrm{F}_K$ satisfies

$$\mathrm{F}_K(\eta) \lesssim \Big(\underbrace{\dots, \sum_{u=h+1}^{K} \eta^{\Delta(u,h;\Delta(k_0,u;s))}, \dots, 0}_{h=k_0,\dots,K-1}\Big). \tag{7.14}$$

Let $\mathcal{Q}$ denote the set of all squares Q used for the correction of equivalence classes up to the end of Step 6. By formula (6.16) for the cost of the correction of one equivalence class, by the recursive relation (6.12) linking structural exponents, and by the formula (5.17) for the basic cost, we deduce that the total cost of length for getting to (7.14) is

$$\Lambda\Big(\bigcup_{Q\in\mathcal{Q}} Q\Big) \sim \eta^{\frac{2(1+r)}{3}}. \tag{7.15}$$

Let us define the *theoretical initial vector of reduced errors* as

$$\overline{\mathrm{F}}(\eta) = (\underbrace{\dots, \eta^{\Delta(k_0,h;s)}, \dots}_{h=k_0,\dots,K}).$$

This is obtained from F_{k_0} in (7.12), inserting in the first coordinate $\eta^s = \eta^{\Delta(k_0,k_0;s)}$ in place of 0. By Step 3, we can without loss of generality assume that this is the total reduced error before the correction of L_{k_0}.

By the argument of Step 5, see (7.11), we have for any $h = k_0, \dots, K-1$ and for all $u = h+1, \dots, K$

$$\Delta(u, h; \Delta(k_0, u; s)) > \Delta(k_0, h; s).$$

It follows that

$$\varrho := \min\Big\{ \frac{\Delta(u, h; \Delta(k_0, u; s))}{\Delta(k_0, h; s)} : h = k_0, \dots, K-1,\ u = h+1, \dots, K \Big\} > 1.$$

Then we have the following quantitative general decrease of errors

$$\mathrm{F}_K(\eta) \le \overline{\mathrm{F}}(\eta^\varrho), \quad 0 < \eta < \delta. \tag{7.16}$$

Step 8. We can now iterate the procedure of Steps 1–7. We get in this way a correction Q of the total error with cost of length satisfying

$$\Lambda(Q) \sim \eta^{\frac{2(1+r)}{3}}.$$

This finishes the proof.

□

8. Equivalence classes with two elements. Nonhomogeneous case

In this section, we discuss the nonhomogeneous case, i.e., $c_i \neq 0$ for some $i > \beta$ in (2.8). We restrict the analysis to the case when equivalence classes have at most two elements.

We need the notions of *leading effect* and *second leading effect.* Consider some $k \in \mathbb{N}$ with $\bar{k} = 1$. As in (3.15), we have $L_k = \{(i_0, j_0), (i_1, j_1)\}$ and we let $i = i_0$ and $j = j_0$. Assume that at some previous stage we used a rectangle $R_{\eta,\lambda}(\varepsilon)$ with parameters $\eta, \lambda, \varepsilon$ satisfying (4.19), that is,

$$\frac{1}{2}|\varepsilon|^\lambda > \eta. \tag{8.1}$$

We compute the effect of the rectangle on the pairs of L_k. By formula (3.8), we have

$$R^{ij}_{\eta,\lambda}(\varepsilon) = \sum_{h=0}^{j} c_{ijh0} R^{ijh0}_{\eta,\lambda}(\varepsilon) + \sum_{h=0}^{j} \sum_{\ell=\widehat{\ell}}^{\infty} c_{ijh\ell} R^{ijh\ell}_{\eta,\lambda}(\varepsilon).$$

Above, we let

$$\widehat{\ell} = \min\big\{\ell \geq 1 : c_{h\ell} \neq 0\big\},$$

where the constants $c_{h\ell}$ are from (3.7). By elementary properties of binomial powers, one can check that $\widehat{\ell}$ does not depend on h. Under assumption (8.1), for fixed ℓ the leading $R^{ijh\ell}_{\eta,\lambda}(\varepsilon)$ is the one with maximum h, i.e., with $h = j$. The proof of this fact is in Section 4, between formulas (4.7) and (4.8).

On varying ℓ, the leading effect is obtained for $\ell = 0$. Then we have $R^{ij}_{\eta,\lambda}(\varepsilon) = c_{ijj0} R^{ijj0}_{\eta,\lambda}(\varepsilon) + \text{Error}$ where the Error is negligible. We call

$$R^{ijj0}_{\eta,\lambda}(\varepsilon) \simeq \varepsilon |\varepsilon|^{\lambda \ell_k}$$

the *leading effect*. When the effect is an error, we speak of *leading error*. By the formulas (3.13), we have $R^{ijj0}_{\eta,\lambda}(\varepsilon) = R^{i_1 j_1 j_1 0}_{\eta,\lambda}(\varepsilon)$. Later, we shall use this identity to cancel the leading error. It is then important, to identify the leading term in the difference

$$\begin{aligned} R^{ij}_{\eta,\lambda}(\varepsilon) - c_{ijj0} R^{ijj0}_{\eta,\lambda}(\varepsilon) &= \sum_{h=0}^{j-1} c_{ijh0} R^{ijh0}_{\eta,\lambda}(\varepsilon) + \sum_{h=0}^{j} \sum_{\ell=\widehat{\ell}}^{\infty} c_{ijh\ell} R^{ijh\ell}_{\eta,\lambda}(\varepsilon) \\ &\simeq R^{ij,j-1,0}_{\eta,\lambda}(\varepsilon) + R^{ijj\widehat{\ell}}_{\eta,\lambda}(\varepsilon) \\ &\simeq \varepsilon^2 |\varepsilon|^{\lambda(\ell_k - r)} + \varepsilon |\varepsilon|^{\lambda(\ell_k + \frac{\widehat{\ell}}{\alpha})}. \end{aligned}$$

For $\lambda > 0$ small enough, the leading term is the latter. We call

$$R^{ijj\widehat{\ell}}_{\eta,\lambda}(\varepsilon) \simeq \varepsilon |\varepsilon|^{\lambda(\ell_k + \ell')}, \quad \text{where } \ell' = \frac{\widehat{\ell}}{\alpha}, \tag{8.2}$$

the *second leading effect*, and when the effect is an error, we speak of *second leading error*. Notice our new notation $\ell' = \frac{\widehat{\ell}}{\alpha}$.

Let us recall that by $E(\eta)$ we denote the initial error (7.1) produced by the cut. A correction R of the error is a countable choice of rectangles that sets to zero all the components of $E(\eta)$. Under Assumption 4.1 we have a bound $K \in \mathbb{N}$ on k.

Theorem 8.1. *Let $\kappa(t) = (t, \varphi(t))$, $t \in [0,1]$, be curve as in (2.8) with $c_i \neq 0$ for some $i > \beta$. Assume that $\bar{k} \leq 1$ for all $k \leq K$. For any $\varepsilon > 0$ there are numbers $C > 0$ and $0 < \delta < 1$ such that for any $0 < \eta < \delta$ there is a correction R of the initial error* $\mathrm{E}(\eta)$ *in* (7.1) *with cost of length satisfying*

$$\Lambda(R) \leq C \eta^{1+r-\varepsilon}. \tag{8.3}$$

Proof. In the proof, we use freely the arguments and the observations made in the proof of Theorem 4.2. To construct the correction, we proceed as follows:

1) We correct all classes with one element (i.e., first representatives) with the recursive method described in Section 4.
2) Then we correct recursively all equivalence classes with two elements.
3) We iterate the construction and we estimate the costs.

The error on $k = 0$ after the cut is comparable to η^s with $s = 1+r$, see the vector of errors (4.2) at position $k = 0$. We correct recursively the error of first representatives for $h = 0, 1, \dots, k$ up to a certain $k \in \mathbb{N}$ included. This is explained in the proof of Theorem 4.2. The vector of leading errors after this correction, is the vector E_k in (4.18). It is useful to express this error in the following way. For $k, h \in \mathbb{N}$, we let

$$\Delta(k,h) = \frac{1+\ell_h\lambda_k}{1+\ell_k\lambda_k} \quad \text{and} \quad \Pi(k) = \Delta(0,1)\Delta(1,2)\dots\Delta(k-1,k).$$

The parameters $\lambda_0 > \lambda_1 > \dots > \lambda_k > 0$ are the ones fixed in Section 4. Then we have

$$\mathrm{E}_k \simeq \big(\underbrace{\dots, \eta^{s\Pi(h+1)\Delta(h+1,h)}, \dots}_{h=0,\dots,k-1}, 0, \underbrace{\dots \eta^{s\Pi(k)\Delta(k,h)}, \dots}_{h\geq k+1}\big).$$

This is formula (4.2) with our new notation. Let us introduce the vector of second leading errors. Motivated by formula (8.2), let us set for $k, h \in \mathbb{N}$

$$\Delta'(k,h) = \frac{1+(\ell_h+\ell')\lambda_k}{1+\ell_k\lambda_k}.$$

By (8.2), the total second leading error accumulated on h after correcting the first representatives $0, 1, \dots, k$ is comparable to η^{sA} with

$$A = \Delta'(0,h) + \Pi(1)\Delta'(1,h) + \dots + \Pi(k)\Delta'(k,h).$$

We search the sum for its leading term. Notice that

$$\begin{aligned} \Pi(k-1)\Delta'(k-1,h) \leq \Pi(k)\Delta'(k,h) \quad &\Leftrightarrow \quad \frac{1+(\ell_h+\ell')\lambda_{k-1}}{1+\ell_k\lambda_{k-1}} \leq \frac{1+(\ell_h+\ell')\lambda_k}{1+\ell_k\lambda_k} \\ &\Leftrightarrow \quad \ell_h + \ell' \leq \ell_k. \end{aligned} \tag{8.4}$$

For any $h \in \mathbb{N}$, let $h^* = \min\{k \in \mathbb{N} : \ell_h + \ell' \leq \ell_k\}$. Clearly, we have

$$h^* > h. \tag{8.5}$$

Then, for any $h \in \mathbb{N}$, the map $k \mapsto \Pi(k)\Delta'(k,h)$ attains the minimum when $k = h^*$.

Now let $k \in \mathbb{N}$ be the minimum $k \in \mathbb{N}$ such that $\bar{k} = 1$. In our previous notation, we have $k = k_0$. Assume we already applied the recursive procedure for correcting the first representatives from 0 to k included. The error on the first representative (i, j) of L_k is 0. Let E^{i_1,j_1} denote the total error on (i_1, j_1), the second element of L_k, accumulated during the procedure. This is made up by the sum of the errors produced by the cut and by the rectangles used in previous stages. However, the correction of (i, j) corrects automatically the leading error on (i_1, j_1), by the identities (3.13).

Therefore, E^{i_1,j_1} is the accumulated second leading error. By (8.4) and (8.5), we thus have

$$E^{i_1j_1} \simeq \eta^{s\Pi(k)\Delta'(k,k)}. \tag{8.6}$$

We correct the error of the equivalence class L_k with two rectangles $R_{\eta_{k,q},\lambda_k}(\varepsilon_{k,q})$, $q \in \{0,1\}$, where $\eta_{k,q}, \lambda_k, \varepsilon_{k,q}$ satisfy (8.2). We have the system of equations

$$\begin{cases} \displaystyle\sum_{q\in\{0,1\}} \sum_{\ell=0}^{\infty} \sum_{h=0}^{j} c_{ijh\ell} R^{ijh\ell}_{\eta_{k,q},\lambda_k}(\varepsilon_{k,q}) = 0 \\ \displaystyle\sum_{q\in\{0,1\}} \sum_{\ell=0}^{\infty} \sum_{h=0}^{j_1} c_{i_1j_1h\ell} R^{i_1j_1h\ell}_{\eta_{k,q},\lambda_k}(\varepsilon_{k,q}) + E^{i_1j_1} = 0. \end{cases} \tag{8.7}$$

We multiply the first equation by $c_{i_1j_1j_10}$, we multiply the second equation by c_{ijj0}, and we subtract the first equation from the second one. Using the identities (3.13), the second equation transforms into the equation

$$\sum_{q\in\{0,1\}} \sum_{h=0}^{j_1-1} c'_{ijh0} R^{i_1j_1h0}_{\eta_{k,q}}(\varepsilon_{k,q}) + \sum_{q\in\{0,1\}} \sum_{\ell=\widehat{\ell}}^{\infty} \sum_{h=0}^{j_1} c'_{ijh\ell} R^{i_1j_1h\ell}_{\eta_{k,q},\lambda_k}(\varepsilon_{k,q}) + c_{ijj}E^{i_1j_1} = 0, \tag{8.8}$$

where $c'_{ijh\ell}$ are explicit constants.

The leading effect of the rectangles in (8.8) is obtained for $\ell = \widehat{\ell}$ and $h = j_1$. The first equation of the system (8.7) provides $\varepsilon_{k,0} \simeq -\varepsilon_{k,1}$. The second equation in (8.8) can be then approximated as follows

$$\varepsilon_{k,1}|\varepsilon_{k,1}|^{\lambda_k(\ell_k+\ell')} \simeq \eta^{s\Pi(k)\Delta'(k,k)}. \tag{8.9}$$

This determines $\varepsilon_{k,1}$. The cost of length of the correction is

$$\Lambda\Big(\bigcup_{q\in\{0,1\}} R_{\eta_k,q}(\varepsilon_{k,q})\Big) \simeq |\varepsilon_{k,0}| \simeq \eta^{s\frac{\Pi(k)\Delta'(k,k)}{1+\lambda_k(\ell_k+\ell')}}. \tag{8.10}$$

This correction produces new leading and second leading errors. The leading effect on $h \neq k$ is comparable to

$$\eta^{s\Pi(k)\Delta'(k,k)\frac{1+\ell_h\lambda_k}{1+(\ell_k+\ell')\lambda_k}} = \eta^{s\Pi(k)\Delta(k,h)}.$$

This is the same effect produced by the correction of the first representative of L_k. The second leading effect on $h \neq k$ is comparable to

$$\eta^{s\Pi(k)\Delta'(k,k)\frac{1+(\ell_h+\ell')\lambda_k}{1+(\ell_k+\ell')\lambda_k}} = \eta^{s\Pi(k)\Delta'(k,h)}.$$

This error is the same second leading effect produced by the correction of the first representative of L_k.

Now we correct the equivalence class L_{k+1}. The error on the first representative (the leading error) is comparable to $\eta^{s\Pi(k)\Delta(k,k+1)}$. The second leading error is comparable to $\eta^{s\Pi(k)\Delta'(k,k+1)}$. The procedure described in the previous step produces the following new errors. The new leading error and second leading error produced on $h \neq k+1$ are comparable to, respectively,

$$\eta^{s\Pi(k+1)\Delta(k+1,h)}, \quad \eta^{s\Pi(k+1)\Delta'(k+1,h)}.$$

The dynamics of leading errors is the same as in Section 4. We analyze the dynamics of second leading errors. After che correction of the first representatives $0, 1, \dots, k$, the vector E'_k of second leading errors (the vector starts from the coordinate k) is

$$\mathrm{E}'_k \simeq \big(\dots, \eta^{s\Pi(k)\Delta'(k,h)}, \dots \big).$$

After the correction of the entire equivalence class L_k the error is

$$\mathrm{E}'_k \simeq \big(0, \underbrace{\dots, \eta^{s\Pi(k)\Delta'(k,h)}, \dots}_{h \geq k+1}\big).$$

After che correction of the entire equivalence class L_{k+1} the error is

$$\begin{aligned} \mathrm{E}'_{k+1} &\simeq \big(\eta^{s\Pi(k+1)\Delta'(k+1,k)}, 0, \underbrace{\dots, \eta^{s\Pi(k)\Delta'(k,h)} + \eta^{s\Pi(k+1)\Delta'(k+1,h)}, \dots}_{h \geq k+2}\big) \\ &\simeq \big(\eta^{s\Pi(k+1)\Delta'(k+1,k)}, 0, \underbrace{\dots, \eta^{s\Pi(k+1)\Delta'(k+1,h)}, \dots}_{h \geq k+2}\big). \end{aligned}$$

Inductively, for any $p > k$ we find the vector of second leading errors

$$\begin{aligned} \mathrm{E}'_p &\simeq \big(\underbrace{\dots, \sum_{q=k}^{p} \eta^{s\Pi(q)\Delta'(q,h)}, \dots}_{h<p}, 0, \underbrace{\dots, \eta^{s\Pi(p)\Delta'(p,h)}, \dots}_{h>p}\big) \\ &\simeq \big(\underbrace{\dots, \eta^{s\Pi(h_p^*)\Delta'(h_p^*,h)}, \dots}_{h<p}, 0, \underbrace{\dots, \eta^{s\Pi(p)\Delta'(p,h)}, \dots}_{h>p}\big), \end{aligned}$$

where $h_p^* = \min\{h^*, p\}$. The 0 is at position p.

For the sake of simplicity, we assume that $K = \infty$, so that we can let p go to infinity. This is without loss of generality. In the length estimate, however, we assume $k \leq K$.

After letting $p \to \infty$, the final vector E of leading errors and the final vector E' of second leading errors are, respectively,

$$\mathrm{E} \simeq \big(\dots, \eta^{s\Pi(h+1)\Delta(h+1,h)}, \dots \big), \quad \mathrm{E}' \simeq \big(\dots, \eta^{s\Pi(h^*)\Delta'(h^*,h)}, \dots \big).$$

By (8.10), the total cost of length of the recursive construction is

$$\Lambda\Big(\bigcup_{k=k_0}^{K} \bigcup_{q \in \{0,1\}} R_{\eta_k,q}(\varepsilon_{k,q}) \Big) \simeq \sum_{k=k_0}^{K} \eta^{s \frac{\Pi(k)\Delta'(k,k)}{1+\lambda_k(\ell_k+\ell')}} \sim \eta^s. \tag{8.11}$$

The new structural exponent is $s_1 = s\Pi(1)\Delta(1,0) = s\Delta(0,1)\Delta(1,0) > s$. Letting $\varrho = \Delta(0,1)\Delta(1,0)$, we have $s_1 = \varrho s$ with $\varrho > 1$. Repeating countably many times the procedure started at the beginning of the proof, we get a correction R of the error $E(\eta)$ in (7.1) with cost of lenght

$$\Lambda(R) \simeq \sum_{h=1}^{\infty} \sum_{k=k_0}^{K} \eta^{s^h \frac{\Pi(k)\Delta'(k,k)}{1+\lambda_k(\ell_k+\ell')}} \sim \eta^s.$$

The claim (8.3) follows from this estimate and from Theorem 4.2. □

9. Main results

In this section, we collect the results proved so far. In the following, $(M, \mathcal{D}, g)$ is a sub-Riemannian manifold satisfying Assumption 4.1 and $\kappa : [0,1] \to \mathbb{R}^2$, with $\kappa(0) = 0$, is the horizontal projection of an abnormal curve $\gamma : [0,1] \to M$. The curve κ may be either homogeneous or nonhomogeneous, according to the classification introduced at the end of Section 2. The curve κ has its parameters $\alpha, \beta \in \mathbb{N}$ and we assume that $1 \leq \alpha < \beta$, so that $r = \beta/\alpha > 1$.

Theorem 9.1. *Let Assumption 7.1 hold and let κ be homogeneous. If $r < 5/4$, then the curve $\gamma = \mathrm{Lift}(\kappa)$ is not length minimizing.*

Proof. We can cut κ by a triangle T_η, where $\eta > 0$ is a small parameter. Let κ^η be the new curve and let $\gamma^\eta = \mathrm{Lift}(\kappa^\eta)$ be its horizontal lift. The gain of length $\Delta L(\eta)$ is given by (3.1):

$$\Delta L(\eta) = \frac{(r-1)^2}{2(2r-1)} \eta^{2r-1} + o\big(\eta^{2r-1}\big).$$

By Theorem 7.3 we can adjust the end-point $\gamma^\eta(1)$, i.e., we can move it back to $\gamma(1)$. Namely, for any $\varepsilon > 0$ there are a constant $C > 0$ and a correction $R \cup Q$ of the end-point with cost of length

$$\Lambda(R \cup Q) \leq C \eta^{\frac{2}{3}(1+r)-\varepsilon}.$$

This holds for all $0 < \eta < \delta$ for some $\delta > 0$. On choosing $0 < \varepsilon < 5 - 4r$, we have for small enough $\eta > 0$

$$\Lambda(R \cup Q) < \Delta L(\eta).$$

Therefore, the curve γ is not length minimizing.

□

When κ is in the nonhomogeneous case, we have the following result. Recall that $\bar{k} + 1$ is the cardinality of L_k and that K is the bound on k given by Assumption 4.1.

Theorem 9.2. *Let κ be nonhomogeneous and assume that $\bar{k} \leq 1$ for all $k \leq K$. If $r < 2$, then the curve $\gamma = \mathrm{Lift}(\kappa)$ is not length minimizing.*

Proof. Now the cost of length to restore the end-point is given by Theorem 8.1. Namely, for any $\varepsilon > 0$ there exist a constant $C > 0$ and a correction R of the end-point with cost of length

$$\Lambda(R) \leq C \eta^{1+r-\varepsilon}.$$

This holds for all $0 < \eta < \delta$ for some $\delta > 0$. On choosing $0 < \varepsilon < 2 - r$, we have for small enough $\eta > 0$

$$\Lambda(R) < \Delta L(\eta).$$

Therefore, the curve γ is not length minimizing. □

Let us comment on the assumptions made in Theorems 9.1 and 9.2.

1) Both theorems hold under Assumption 4.1, that provides the bound $k \leq K$. This is needed to control the multiplicative constant in the length estimate. See Remark 4.3.

2) In Theorem 9.1 we have Assumption 7.1. We could not get rid of this assumption. Another natural question, of course, is whether the threshold 5/4 has any meaning. The construction made in Sections 4–7 seems precise. In some cases, however, the estimate can be improved. Let us consider in $\mathbb{R}^5$ the horizontal distribution spanned by the frame of vector fields

$$X_1 = \partial_1, \qquad X_2 = \partial_2 + x_1\partial_3 + x_1^5\partial_4 + x_1 x_2^3 \partial_5.$$

By Proposition 2.2, the plane curve $x_2 = x_1^{4/3}$ is the horizontal projection of an abnormal extremal. The curve is in fact of the homogeneous type. A delicate construction (all details are omitted, here) shows that the horizontal lift of this curve is not length minimizing near $x_1 = 0$ for the natural Carnot-Carathéodory metric. The singularity with exponent $4/3 > 5/4$ can be cut and adjusted with gain of length.

3) In Theorem 9.2 we have the assumption $\bar{k} \leq 1$. This restriction can be probably dropped. It is likely that one has to consider all nonzero higher order terms in the expansion of φ in (2.8) and then mix the techniques of Sections 4–7 with the ones of Section 8.

4) Our very initial starting point is a structure theorem of algebraic type for abnormal curves. This is the case 1) in Proposition 2.2, which relies upon the restrictive Assumption 2.1. In the setting of Carnot groups there is a similar explicit algebraic classification of abnormal extremals, with no restrictions on rank and step, see [LLMV].

10. $C^{1,\delta}$ regularity of geodesics in a class of Carnot groups

Let $\mathfrak{g}$ be a stratified nilpotent n-dimensional real Lie algebra with $\mathfrak{g} = \mathfrak{g}_1 \oplus \ldots \oplus \mathfrak{g}_S$, $S \geq 2$, and $\mathfrak{g}_{i+1} = [\mathfrak{g}_1, \mathfrak{g}_i]$ for $i \leq S-1$ and $\mathfrak{g}_i = \{0\}$ for $i > S$. We assume that the Lie algebra has rank 2, i.e., that the horizontal layer $\mathfrak{g}_1$ is two-dimensional, $\dim(\mathfrak{g}_1) = 2$. We also assume that the following commutativity relations on higher layers hold:

$$[\mathfrak{g}_i, \mathfrak{g}_j] = 0 \quad \text{for all } i, j \geq 2 \text{ such that } i + j > 4. \tag{10.1}$$

We can realize $\mathfrak{g}$ as a Lie algebra of left invariant vector fields in $\mathbb{R}^n$. Then we may assume that $\mathfrak{g}_1$ is spanned by a pair of vector fields X_1 and X_2 as in (2.1), where f_j, $j = 3, \ldots, n$, are polynomials with degree $\deg(f_j) \leq S-1$. By (10.1), the polynomials f_j only depends on the variables x_1 and x_2 of $\mathbb{R}^n$ (see Proposition 2.6 in [LM]). The

Lie algebra $\mathfrak{g}$ is naturally associated and may in fact be identified with a Lie group. A nilpotent and stratified Lie group is known as Carnot group. If $S \le 4$, the assumption in (10.1) is automatically satisfied. It is known that if $S \le 4$ (and the rank is 2) all sub-Riemannian geodesics are C^∞ smooth (see Example 4.6 in [LM]). The following theorem is of interest in the case $S > 4$.

Theorem 10.1. *Any length minimizing curve for the Carnot-Carathèodory distance in a Carnot group of rank 2 and step $S > 4$ satisfying* (10.1) *is of class $C^{1,\delta}$ for any*

$$0 \le \delta < \min\Big\{\frac{2}{S-4}, \frac{1}{4}\Big\}. \tag{10.2}$$

Proof. Let $\kappa : [0,1] \to \mathbb{R}^2$ be the horizontal projection of a singular curve. The support of the curve lies in the zero set of a polynomial $f = \sum_{j=3}^n \mu_j \partial_1 f_j \neq 0$, where $\mu_3, \dots, \mu_n \in \mathbb{R}$. After a left translation in the group and after a rotation of the coordinates in the plane, we may assume that $\kappa(0) = 0$ and $\kappa(t) = (t, \varphi(t))$ where φ is the function in (2.8) with its parameters $\alpha, \beta \in \mathbb{N}$, with $1 < \alpha < \beta$.

Our argument in Sections 4–8 proves that $\gamma = \mathrm{Lift}(\kappa)$ is not length minimizing, provided that $\beta/\alpha < 5/4$ and Assumption 7.1 holds. On the other hand, Assumption 7.1 does hold if equivalence classes L_k contain at most two elements, as noted in Remark 7.2.

The correction procedures of Sections 4–8 are restricted to monomials of the form $x_1^{i+1} x_2^j$ with the bound $i + j \le S - 2$. We introduced in Section 3 the equivalence classes L_k, with $k \in \mathbb{N}$. Assume that $(i, j) \in L_k$, with $j = 0, \dots, \alpha - 1$, is the first representative of L_k. Then we have $i \le S-2$, because monomials $x_1^{i+1} x_2^j$ with $i > S-2$ do not apperar in the polynomials $f_3, \dots, f_n$. Then, if β is large, the cardinality of L_k is small, and namely we have the implication

$$\beta > \frac{S}{2} - 1 \quad \Rightarrow \quad \bar{k} = \Big[\frac{i}{\beta}\Big] \le \frac{i}{\beta} < 2 \quad \Rightarrow \quad \mathrm{Card}(L_k) \le 2. \tag{10.3}$$

We estimate the minimum β/α which is not covered by the cut-and-adjust argument. By (10.3), this minimum is

$$\begin{aligned} m &= \min\Big\{\frac{\beta}{\alpha} \in \mathbb{Q} : \alpha, \beta \in \mathbb{N},\, 1 < \alpha < \beta \le \frac{S}{2} - 1\Big\} \\ &= \min\Big\{\frac{\beta}{\beta - 1} \in \mathbb{Q} : \beta \in \mathbb{N},\, \beta \le \frac{S}{2} - 1\Big\} \\ &\ge \frac{S-2}{S-4}. \end{aligned}$$

If $\beta/\alpha < \min\{m, 5/4\}$ the curve κ is not length minimizing. On the other hand, if $\beta/\alpha \ge \min\{m, 5/4\}$, then the curve κ is $C^{1,\delta}$ for any δ as in (10.2). This ends the proof. □

References

[AS] A. Agrachev, A. Sarychev, Abnormal sub-Riemannian geodesics: Morse index and rigidity, Ann. Inst. Henri Poincaré, **13** n. 16 (1996), 635–690.

[BM] E. Bierstone, P. D. Milman, Semianalytic and subanalytic sets. Inst. Hautes Études Sci. Publ. Math. No. 67 (1988), 5–42.

[CJT] Y. Chitour, F. Jean, E. Trélat, Genericity results for singular curves. J. Differential Geom. 73 (2006), no. 1, 45–73.

[LLMV] E. Le Donne, G. P. Leonardi, R. Monti, D. Vittone, Extremal curves in nilpotent Lie groups. Geom. Funct. Anal. 2012 (to appear).

[LM] G. P. Leonardi, R. Monti, End-point equations and regularity of sub-Riemannian geodesics. Geom. Funct. Anal. 18 (2008), no. 2, 552–582.

[LS] W. Liu, H. Sussmann, Shortest paths for sub-Riemannian metrics on rank-two distributions, Mem. Amer. Math. Soc. **118** (1995), no. 564.

[M1] R. Montgomery, Abnormal minimizers, SIAM J. Control Optim., **32** (1994), 1605–1620.

[M2] R. Montgomery, A tour of subriemannian geometries, their geodesics and applications, Mathematical Surveys and Monographs, 91. American Mathematical Society, Providence, RI, 2002.

[S] R. S. Strichartz, Sub-Riemannian geometry, J. Differential Geom., **24** (1986), 221–263. [Corrections to "Sub-Riemannian geometry", J. Differential Geom., **30** (1989), 595–596].

Roberto Monti: Dipartimento di Matematica, Università di Padova, Via Trieste, 63, 35121 Padova, Italy

E-mail address: monti@math.unipd.it