History

The journal *Rendiconti dell'Istituto di Matematica dell'Università di Trieste* was founded in 1969 with the aim of publishing original research articles in all fields of mathematics. The first director of the journal was Arno Predonzan, subsequent directors were Graziano Gentili, Enzo Mitidieri and Bruno Zimmermann.

Rendiconti dell'Istituto di Matematica dell'Università di Trieste has been the first Italian mathematical journal to be published also on-line. The access to the electronic version of the journal is free. All articles are available on-line.

In 2008 the *Dipartimento di Matematica e Informatica*, the owner of the journal, decided to renew it. In particular, a new Editorial Board was formed, and a group of four Managing Editors was selected. The name of the journal however remained unchanged; just the subtitle *An International Journal of Mathematics* was added. Indeed, the opinion of the whole department was to maintain this name, not to give the impression, if changing it, that a further new journal was being launched.

Website Address: http://rendiconti.dmi.units.it

Rendiconti dell'Istituto di Matematica dell'Università di Trieste

An International Journal of Mathematics

Volume 46 (2014)
Dipartimento
di Matematica
e Geoscienze

ISSN 0049-4704

EUT EDIZIONI UNIVERSITÀ DI TRIESTE

ISSN 0049-4704

EUT – Edizioni Università di Trieste
via E. Weiss, 21 – 34128 Trieste
http://eut.units.it
https://www.facebook.com/EUTEdizioniUniversitaTrieste

Rend. Istit. Mat. Univ. Trieste
Volume 46 (2014), 1–17

On the L-infinity description of the Hitchin Map

Peter Dalakov

Communicated by Ugo Bruzzo

Abstract. *We exhibit, for a G-Higgs bundle on a compact complex manifold, a subspace of the second cohomology of the controlling dg Lie algebra, containing the obstructions to smoothness. For this we construct an L_∞-morphism, which induces the Hitchin map and whose "toy version" controls the adjoint quotient morphism. This extends recent results of E. Martinengo.*

Keywords: Hitchin map, Higgs bundles, L-infinity morphism
MS Classification 2010: 17B70, 14D15, 14D20

1. Introduction

The idea that deformation problems are controlled by differential graded Lie algebras (dgla's) has been a key guiding principle in (characteristic zero) deformation theory. This philosophy, currently subsumed in the work of J.Lurie, has been actively exploited by Deligne, Drinfeld, Gerstenhaber, Goldman–Millson, Kontsevich, Nijenhuis–Richardson, Schlessinger, Stasheff and Quillen, since the earliest days of the subject. Kontsevich argued in [14] that the formal geometry of moduli problems is governed by a richer structure: an L_∞-algebra (strongly homotopy Lie algebra), and natural transformations between deformation functors are induced by L_∞-morphisms of the controlling dgla's. For example, it is shown in [6] that Griffiths' period map is induced by an L_∞-morphism. In a certain sense, an L_∞-morphism encodes the "Taylor expansion" of a morphism of pointed formal varieties ([14, §4.1]). By general deformation-theoretic arguments its linear part is a morphism of obstruction theories.

In view of this, given a pair of deformation functors and a natural transformation between them, one is confronted with the questions of identifying controlling dgla's and a corresponding L_∞-morphism. Apart from being aesthetically pleasing, this gives additional information for the obstruction spaces of the two functors. The main result in this note is a particular example of such a setup.

Let X be a compact complex manifold, and G be a complex reductive Lie group of rank N, with Lie algebra $\mathfrak{g}$. By an *Ω^1_X-valued G-Higgs bundle (Higgs pair) on X* we shall mean a pair $(\mathbf{P}, \theta)$, consisting of a holomorphic principal G-bundle $\mathbf{P} \to X$, and a section $\theta \in H^0(X, \mathrm{ad}\mathbf{P} \otimes \Omega^1_X)$, satisfying $\theta \wedge \theta = 0$. The Hitchin map associates to $(\mathbf{P}, \theta)$ the spectral invariants of θ. After some choices, these invariants determine a point in $\mathcal{B} = \bigoplus_{d_i} H^0(X, S^{d_i}\Omega^1_X)$, where d_i are the degrees of the basic G-invariant polynomials on $\mathfrak{g}$. For example, if G is a classical group, then θ can be represented locally on X by a matrix of holomorphic 1-forms, with commuting components. Then the Hitchin map assigns to it the coefficients of its characteristic polynomial. Considering Higgs pairs on $X \times \mathrm{Spec}A$, for an Artin local ring A, allows one to define the Hitchin map as a natural transformation between suitable deformation functors, see Section 2.2.

If X is projective ([22]) or compact Kähler ([7]) there exist actual (coarse) moduli spaces of (semi-stable) Higgs pairs with fixed topological invariants. In the projective case the Hitchin map is known to be a proper morphism to $\mathcal{B}$ ([22, §6]), and if $\dim_{\mathbb{C}} X = 1$ it determines an algebraic completely integrable Hamiltonian system ([11, 12]).

The main resluts in this note are the following two theorems.

Theorem 1.1. *Let X be a compact complex manifold, G a complex reductive Lie group, and $\{p_i, i \in E\}$ homogeneous generators of $\mathbb{C}[\mathfrak{g}]^G$, $\deg p_i = d_i$. Let $(\mathbf{P}, \theta)$ be a G-Higgs bundle on X, and $\mathscr{C}^\bullet = \bigoplus_{p+q=\bullet} A^{0,p}(ad\mathbf{P} \otimes \Omega^q_X)$ its controlling dgla. Then the obstruction space $O_{Def_{\mathscr{C}^\bullet}} \subset H^2(\mathscr{C}^\bullet)$ is contained in the kernel of the map*

$$H^2(\mathscr{C}^\bullet) \longrightarrow \bigoplus_{i \in E} H^1(X, S^{d_i}\Omega^1_X)$$

$$\left[s^{2,0}, s^{1,1}, s^{0,2}\right] \longmapsto \bigoplus_{i \in E} (\partial p_i)(s^{1,1} \otimes \theta^{d_i - 1}).$$

Here ∂p_i denotes the differential of p_i, thought of as an element of $\mathfrak{g}^\vee \otimes S^{d_i-1}(\mathfrak{g}^\vee)$.

We make some remarks about the geometrical meaning of this theorem in Section 4.2. **Theorem 1.1** is an easy consequence of the more technical

Theorem 1.2. *Let X be a compact complex manifold, G a complex reductive Lie group, and $\{p_i, i \in E\}$ homogeneous generators of $\mathbb{C}[\mathfrak{g}]^G$, $\deg p_i = d_i$. Let $(\mathbf{P}, \theta)$ be a G-Higgs bundle on X, and $\mathscr{C}^\bullet = \bigoplus_{p+q=\bullet} A^{0,p}(ad\mathbf{P} \otimes \Omega^q_X)$ its controlling dgla. Let $\mathbf{p}_0 : \overline{S^\bullet(\mathscr{C}^\bullet)} \to \overline{S^\bullet\left(A^{0,\bullet}(ad\mathbf{P} \otimes \Omega^1_X)\right)}$ be the homomorphism induced by $\bigoplus_{p,q} s^{p,q} \mapsto s^{1,q} \in A^{0,q}(ad\mathbf{P} \otimes \Omega^1_X)$. Then the collection of maps*

$$\bigoplus_i h^{d_i}_k = \bigoplus_i (\partial^k p_i)(\_ \otimes \theta^{d_i - k}) \circ \mathbf{p}_0 : \qquad S^k(\mathscr{C}^\bullet[1]) \longrightarrow \bigoplus_i A^{0,\bullet}(S^{d_i}\Omega^1_X)$$

$$\bigoplus_{p_1,q_1} s_1^{p_1,q_1} \cdot \bigoplus_{p_2,q_2} s_2^{p_2,q_2} \cdot \ldots \cdot \bigoplus_{p_k,q_k} s_k^{p_k,q_k}$$

$$\longmapsto \bigoplus_i \sum_{q_1,\ldots,q_k} (\partial^k p_i)(s_1^{1,q_1} \otimes \ldots \otimes s_k^{1,q_k} \otimes \theta^{d_i-k})$$

for all $k \geq 1$, induces an L_∞-morphism

$$h_\infty : \mathscr{C}^\bullet = \bigoplus_{p+q=\bullet} A^{0,p}(ad\mathbf{P} \otimes \Omega_X^q) \to \mathscr{B}^\bullet = \bigoplus_{i \in E} A^{0,\bullet}(S^{d_i}\Omega_X^1)[-1].$$

The natural transformation of deformation functors, induced by h_∞ is the Hitchin map: Def$(h_\infty) = H$ under the identifications Def$_{\mathscr{B}^\bullet} \simeq$ Def$_{H(E,\theta)}$ and Def$_{\mathscr{C}^\bullet} \simeq$ Def$_{(\mathbf{P},\theta)}$.

The content of this note is organised as follows. In Section 2 we discuss dgla's and L_∞-algebras, and give some examples. In Section 3 we study a Lie-algebraic "toy model" for the Hitchin map. For that, we fix homogeneous generators of $\mathbb{C}[\mathfrak{g}]^G$, which allows us to identify the adjoint quotient morphism $\mathfrak{g} \to \mathfrak{g} /\!\!/ G$ with a polynomial map $\chi : \mathfrak{g} \to \mathbb{C}^N$. We associate to a fixed $v \in \mathfrak{g}$ a pair of (very simple) dgla's, $C^\bullet$ and $B^\bullet$ (2.3, 3.1), whose Maurer–Cartan functors satisfy $\mathrm{MC}_{C^\bullet} = \mathfrak{g}$, $\mathrm{MC}_{B^\bullet} = \mathbb{C}^N$. Motivated by [14, §4.2] we construct an L_∞-morphism $h_\infty : C^\bullet \to B^\bullet$, such that $\mathrm{MC}(h_\infty) = \chi$ (after some identifications).

A suitable modification of h_∞ gives an L_∞-description of the Hitchin map, described in Section 4.1, where we prove **Theorem 1.2**. That in turn gives information about obstructions to smoothness for the functor $\mathrm{Def}_{\mathscr{C}^\bullet}$. These are considered in Section 4.2, together with the proof of **Theorem 1.1**. For details about obstruction calculus we refer to [5] and [18].

Our results are the natural generalisation of [19, §7], where the case of $G = GL(n,\mathbb{C})$ is treated by ingenuous use of powers and traces of matrices.

2. Preliminaries

2.1. Notation and Conventions

The ground field is $\mathbb{C}$. We denote by $\mathrm{Art}_\mathbb{C}$ the category of local Artin $\mathbb{C}$-algebras with residue field $\mathbb{C}$, and denote by $\mathfrak{m}_A$ the maximal ideal of $A \in \mathrm{Art}_\mathbb{C}$. We denote by $\mathrm{Fun}(\mathrm{Art}_\mathbb{C}, \mathrm{Sets})$ the category of functors from $\mathrm{Art}_\mathbb{C}$ to Sets, and use "morphism of functors" and "natural transformations" interchangeably. We use the standard acronym "dgla" for a "differential graded Lie algebra". If $V^\bullet$ is a graded vector space, we denote by $V[n]$ its shift by n, i.e., $V[n]^i = V^{n+i}$. We denote by $T(V)$ the tensor algebra and by $S(V) = \bigoplus_{k \geq 0} S^k(V)$ the symmetric algebra of a (graded) vector space V. The same notation is used for the underlying vector spaces of the corresponding coalgebras, but we use $S_c(V)$ and $T_c(V)$ when we want to emphasise the coalgebra structure. The *reduced* symmetric, resp. tensor (co)algebra is denoted $\overline{S(V)} = \bigoplus_{k\geq 1} S^k(V)$, resp. $\overline{T(V)}$. We denote by $\cdot$ the multiplication in $S(V)$. By $S(k, n-k)$ we denote the $(k, n-k)$ unshuffles: the permutations $\sigma \in \Sigma_n$, satisfying $\sigma(i) < \sigma(i+1)$ for all $i \neq k$.

Next, G is a complex reductive Lie group of rank N and $\mathfrak{g} = \mathrm{Lie}(G)$. We use fixed homogeneous generators, $\{p_i, i \in E\}$, of the ring of G-invariant polynomials on $\mathfrak{g}$. The degrees of the invariant polynomials are $d_i = \deg p_i$, so the exponents of $\mathfrak{g}$ are $d_i - 1$. The adjoint quotient map will always be given in terms of this basis, i.e., $\chi : \mathfrak{g} \to \mathbb{C}^N \simeq \mathfrak{g} /\!\!/ G$.

The base manifold X is assumed to be compact and complex. For a holomorphic principal bundle, $\mathbf{P}$, we denote by $\mathrm{ad}\mathbf{P}$ its associated bundle of Lie algebras, $\mathrm{ad}\mathbf{P} = \mathbf{P} \times_{\mathrm{ad}} \mathfrak{g}$. We denote by Ω^p_X the sheaf of holomorphic p-forms on X, and by $A^{p,q}$ the global sections of the sheaf $\mathscr{A}^{p,q}$ of complex differential forms of type (p, q).

We use $\mathcal{B}$ for the Hitchin base, $\mathscr{B}^\bullet$ for the abelian dgla governing the deformations of an element of $\mathcal{B}$, and $B^\bullet := \mathbb{C}^N[-1]$ for the "toy model" of $\mathscr{B}^\bullet$, see Section 2.3. We use $\mathscr{C}^\bullet$ for the dgla controlling the deformations of a Higgs pair $(\mathbf{P}, \theta)$ (see Section 2.2) and $C^\bullet$ for its "toy version" $\mathfrak{g} \otimes \mathbb{C}[\varepsilon]/\varepsilon^2$.

2.2. Differential Graded Lie Algebras

Since there exist numerous introductory references for this material ([8, 9, 16, 17, 18]), we present here only the basic definitions, without attempting to motivate them in any way. A *differential graded Lie algebra* (dgla) is a triple $(\mathscr{C}^\bullet, d, [\ ,\ ])$. Here $\mathscr{C}^\bullet = \bigoplus_{i\in\mathbb{N}} \mathscr{C}^i[-i]$ is a graded vector space, endowed with a bracket $[\ ,\ ] : \mathscr{C}^i \times \mathscr{C}^j \to \mathscr{C}^{i+j}$. The bracket is graded skew-symmetric and satisfies a graded Jacobi identity. Finally, $d : \mathscr{C} \to \mathscr{C}[1]$ is a differential ($d^2 = 0$), which is a degree 1 derivation of the bracket. To a dgla $\mathscr{C}^\bullet$ we associate a Maurer–Cartan functor $\mathrm{MC}_{\mathscr{C}^\bullet} : \mathrm{Art}_{\mathbb{C}} \to \mathrm{Sets}$,

$$\mathrm{MC}_{\mathscr{C}^\bullet}(A) = \left\{ u \in \mathscr{C}^1 \otimes \mathfrak{m}_A \,\middle|\, du + \frac{1}{2}[u, u] = 0 \right\}$$

and a deformation functor $\mathrm{Def}_{\mathscr{C}^\bullet} : \mathrm{Art}_{\mathbb{C}} \to \mathrm{Sets}$,

$$\mathrm{Def}_{\mathscr{C}^\bullet}(A) = \mathrm{MC}_{\mathscr{C}^\bullet}(A) / \exp(\mathscr{C}^0 \otimes \mathfrak{m}_A).$$

The (gauge) action of $\exp(\mathscr{C}^0 \otimes \mathfrak{m}_A)$ on $\mathscr{C}^1 \otimes \mathfrak{m}_A$ is given by

$$\exp(\lambda) : u \mapsto \exp(\mathrm{ad}\lambda)(u) + \frac{I - \exp(\mathrm{ad}\lambda)}{\mathrm{ad}\lambda}(d\lambda). \tag{1}$$

Often $\mathrm{MC}_{\mathscr{C}^\bullet}(A)$ is considered as the set of objects of a groupoid (*the Deligne groupoid*), which is the action groupoid for the gauge action on $\mathrm{MC}_{\mathscr{C}^\bullet}(A)$ ([8, §2.2]).

2.3. Examples

Deformation problems are described by deformation functors $\mathrm{Def} : \mathrm{Art}_{\mathbb{C}} \to \mathrm{Sets}$, and we say that a problem is governed (controlled) by a dgla $\mathscr{C}^\bullet$, if there exists an isomorphism $\mathrm{Def}_{\mathscr{C}^\bullet} \simeq \mathrm{Def}$. A compendium of examples can be found in [16, §1], or in [20]. The controlling dgla is by no means unique, but quasi-isomorphic dgla's have isomorphic deformation functors ([16, Corollary 3.2]). We give now a minimalistic (abelian) example, which will be used later.

Let V be a finite-dimensional vector space, and $\xi \in V$. We consider the functor $\mathrm{Def}_{\xi,V} : \mathrm{Art}_{\mathbb{C}} \to \mathrm{Sets}$ of embedded deformations of $\xi \in V$. That is, for any $A \in \mathrm{Art}_{\mathbb{C}}$,

$$\mathrm{Def}_{\xi,V}(A) = \{\sigma \in V \otimes A \,|\, \sigma = \xi \mod \mathfrak{m}_A\} = \{\xi\} + V \otimes \mathfrak{m}_A \subset V \otimes A,$$

with the obvious map on morphisms. Then $V[-1]$, a dgla with trivial bracket and trivial differentials, concentrated in degree 1, controls the deformation problem. Indeed, $\mathrm{MC}_{V[-1]}(A) \equiv \mathrm{Def}_{V[-1]}(A) = V \otimes \mathfrak{m}_A$, which we write as $\mathrm{MC}_{V[-1]} = V = \mathrm{Def}_{V[-1]}$. The bijection $\mathrm{Def}_{V[-1]}(A) \simeq \mathrm{Def}_{\xi,V}(A)$, $s \mapsto \xi + s$, induces an isomorphism of functors $\mathrm{Def}_{V[-1]} \simeq \mathrm{Def}_{\xi,V}$.

Suppose now X is Kähler, and $V = H^0(X, F)$, for a holomorphic vector bundle $F \to X$, with $h^i(F) = 0$ for $i \geq 1$. It is then easy to see that $\mathrm{Def}_{\xi,H^0(F)}$ is isomorphic to the deformation functor of the abelian dgla $\left(A^{0,\bullet}(F)[-1], \overline{\partial}_F\right)$, where $\overline{\partial}_F$ is the Dolbeault operator of F. The isomorphism is induced by the canonical inclusion $H^0(X, F) \subset A^{0,0}(X, F) \subset A^{0,\bullet}(F)[-1]^1$. The existence of such an inclusion relies on Hodge theory (see [10, Chapter 0 §6, Chapter 1 § 2]), and this is where we use the Kähler condition. To prove that the two dglas $H^0(X, F)[-1]$ and $\left(A^{0,\bullet}(F)[-1], \overline{\partial}_F\right)$ are quasi-isomorphic, one can use the Hodge decomposition and follow the general setup from [9, §2] or [13].

We denote by Def_ξ the functor of deformations of a section $\xi \in H^0(X, F)$. It is isomorphic to the deformation functor of $\left(A^{0,\bullet}(F)[-1], \overline{\partial}_F\right)$ on an arbitrary X and without the vanishing condition. There is a natural morphism $\mathrm{Def}_{\xi,H^0(F)} \to \mathrm{Def}_\xi$.

We reserve special notation for two instances of this example, namely $B^\bullet := \mathbb{C}^N[-1]$ and $\mathscr{B}^\bullet := \left(\oplus_i A^{0,\bullet}(S^{d_i}\Omega^1_X)[-1], \overline{\partial}_X\right)$. We also use for the Hitchin base $\mathcal{B} = \bigoplus_i H^0(X, S^{d_i}\Omega^1_X)$.

2.4. L-infinity algebras: Motivation

The notion of L_∞-algebra (strongly homotopy Lie algebra, Sugawara algebra) generalises the notion of a dgla by relaxing the Jacoby identity, and allowing it be satisfied only "up to homotopy" ("BRST-exact term"), determined by "higher brackets". For a detailed motivation to this (somewhat technical) subject and its applications to geometry and physics we refer to [14, 15, 20] and the references therein. Here we make some non-rigorous remarks along the lines of [14, §4] and give the precise definitions (following [17, Chapter VIII]) in the next subsection.

Suppose that we want to study (algebraically) a formal neighbourhood of $0 \in V$, where V is a (possibly infinite-dimensional) vector space. One way to do this is to consider the reduced cofree cocommutative coassociative coalgebra, cogenerated by V, that is, $\overline{C(V)} = \bigoplus_{n\geq 1} (V^{\otimes n})^{\Sigma_n} \subset \overline{T_c(V)}$. Indeed, if V is *finite-dimensional*, then $\overline{C(V)}^{\vee}$ is the maximal ideal of the algebra of formal power series. Next, a morphism $\overline{C(V)} \to \overline{C(W)}$ is determined, by the universal property of cofree coalgebras, by a linear map $h : \overline{C(V)} \to W$, with homogeneous components $h^{(n)} : (V^{\otimes n})^{\Sigma_n} \to W$, which are closely related to the Taylor coefficients of h. Indeed, the Taylor coefficients of h are symmetric multilinear maps $h_n = \partial^n h : V^{\otimes n} \to W$. They factor through the quotient $S^n(V)$, and are carried to $h^{(n)}$ under the identification $S^n(V) \simeq (V^{\otimes n})^{\Sigma_n}$. All of this can be done with graded vector spaces as well.

An L_∞-structure on a graded vector space $V^\bullet$ is the data of a degree $+1$ coderivation Q of the coalgebra $\overline{C_c(V[1])}$, satisfying $Q^2 = 0$, i.e., a codifferential. This is thought of as an odd vector field on the formal graded manifold $(V[1], 0)$. Its Taylor coefficients $q_n = \partial^n Q : \overline{S^n_c(V[1])} \to V[1]$ can be considered, by the décalage isomorphism $S^n(V[1]) \simeq \Lambda^n(V)[n]$, as maps $\mu_n \in \mathrm{Hom}^{2-n}(\Lambda^n V^\bullet, V^\bullet)$.

2.5. L-infinity algebras: Definitions

An L_∞*-algebra structure* $(V^\bullet, q)$ on a graded vector space $V^\bullet$ is a collection of linear maps $q_k \in \mathrm{Hom}^1(S^k_c(V[1]), V[1])$, $k \geq 1$, such that the natural extension of $q = \sum_k q_k$ to a degree $+1$ coderivation Q on $\overline{S_c(V[1])}$ is a codifferential, i.e., $Q^2 = 0$. We recall ([17, Corollary VIII.34]) that

$$Q(v_1 \cdot \ldots \cdot v_n) = \sum_{k=1}^{n} \sum_{\sigma \in S(k,n-k)} \epsilon(\sigma) q_k(v_{\sigma(1)} \cdot \ldots \cdot v_{\sigma(k)}) \cdot v_{\sigma(k+1)} \cdot \ldots \cdot v_{\sigma(n)}.$$

A dgla is an L_∞-algebra with $q_1(a) = -da$, $q_2(a \cdot b) = (-1)^{\deg a}[a, b]$, and $q_k = 0$ for $k \geq 3$. To an L_∞-algebra $(V^\bullet, q)$ one associates a Maurer–Cartan functor $\mathrm{MC}_{V^\bullet} : \mathrm{Art}_{\mathbb{C}} \to \mathrm{Sets}$

$$\mathrm{MC}_V(A) = \left\{ u \in V^1 \otimes \mathfrak{m}_A \,\middle|\, \sum_{k\geq 1} \frac{q_k(u^k)}{k!} = 0 \right\}$$

and a deformation functor $\mathrm{Def}_{V^\bullet}$, $\mathrm{Def}_V(A) = \mathrm{MC}_V(A)/\sim_{\text{homotopy}}$. We refer to [17, IX] and [16, §5] for the definition of homotopy equivalence between two Maurer–Cartan elements. We do not give it here, since we shall work only with dgla's considered as L_∞-algebras, and for these gauge equivalence coincides with homotopy equivalence, see [16, Theorem 5.5].

A *morphism* $h_\infty : (V, q) \to (W, \hat{q})$ between two L_∞-algebras is a sequence of linear maps $h_k \in \mathrm{Hom}^0(S^k_c V[1], W[1])$, $k \geq 1$, for which the induced coalgebra morphism $H : \overline{S_c V[1]} \to \overline{S_c W[1]}$ is a chain map, i.e., satisfies $H \circ Q = \widehat{Q} \circ H$. If we denote the components of Q and H by $Q^n_k : S^k_c(V) \to S^n_c(V)$ and H^n_k, respectively, then the morphism condition reads

$$\sum_{a=1}^{\infty} h_a \circ Q^a_k = \sum_{a=1}^{\infty} \hat{q}_a \circ H^a_k,$$

for all $k \in \mathbb{N}$.

We emphasise that the category of dgla's is a subcategory of the category of L_∞-algebras, but it is *not* full. For a dgla the only possibly non-zero components of Q are Q^k_k, $k \geq 1$ and Q^{k-1}_k, $k > 1$. For an *abelian* dgla $q_2 = 0 = Q^{k-1}_k$. We spell out the condition for an L_∞-morphism $h_\infty : (V, q) \to (W, \hat{q})$ from a dgla to an abelian dgla. The $\{h_k\}$ determine an L_∞-morphism if

$$h_1 \circ q_1 = \hat{q}_1 \circ h_1, \tag{2}$$

which says that h_1 is a morphism of complexes, and

$$h_k \circ Q^k_k + h_{k-1} \circ Q^{k-1}_k = \hat{q}_1 \circ h_k,\ k \geq 2. \tag{3}$$

The last condition, when evaluated on homogeneous elements $s_1, \ldots, s_k$ reads

$$\begin{gathered} h_k \left(- \sum_{\sigma \in S(1,k-1)} \varepsilon(\sigma) d(s_{\sigma_1}) \cdot s_{\sigma_2} \cdot \ldots \cdot s_{\sigma_k} \right) + \qquad (4) \\ h_{k-1} \left(\sum_{\sigma \in S(2,k-2)} \varepsilon(\sigma)(-1)^{\deg s_{\sigma_1}} [s_{\sigma_1}, s_{\sigma_2}] \cdot s_{\sigma_3} \cdot \ldots \cdot s_{\sigma_k} \right) = \\ -dh_k(s_1 \cdot \ldots \cdot s_k). \end{gathered}$$

It expresses the failure of h_{k-1} to preserve the bracket in terms of a homotopy given by h_k. Finally, $(V, q) \mapsto \mathrm{MC}_V$ determines a functor $\mathrm{MC} : L_\infty \to Fun(\mathrm{Art}_{\mathbb{C}}, \mathrm{Sets})$, whose action on morphisms is given by sending an L_∞-morphism $h_\infty \in \mathrm{Hom}_{L_\infty}(V, W)$ to a natural transformation $\mathrm{MC}(h_\infty) : \mathrm{MC}_V \to \mathrm{MC}_W$, and, for each $A \in \mathrm{Art}_{\mathbb{C}}$,

$$\mathrm{MC}(h_\infty)(A) : MC_V(A) \ni x \longmapsto \sum_{k=1}^{\infty} \frac{1}{k!} h_k(x^k) \in \mathrm{MC}_W(A). \tag{5}$$

This descends to a natural transformation $\mathrm{Def}(h_\infty) : \mathrm{Def}_V \to \mathrm{Def}_W$. For more details, see, e.g., [17].

2.6. Deformation functors for Higgs bundles

As already stated, for us a Higgs bundle (Higgs pair) is a pair $(\mathbf{P}, \theta)$, $\theta \in H^0(X, \mathrm{ad}\mathbf{P} \otimes \Omega^1_X)$, $\theta \wedge \theta = 0$. We use the term *L-valued Higgs bundle* if instead $\theta \in H^0(X, \mathrm{ad}\mathbf{P} \otimes L)$, for some vector bundle $L \to X$ (as in [4, §17]).

Infinitesimal deformations of Higgs bundles have been studied extensively. Biswas and Ramanan ([2]) discussed the functor of deformations $\mathrm{Def}_{(\mathbf{P},\theta)}$ of a Higgs pair $(\mathbf{P}, \theta)$ for $\dim X = 1$, and identified a deformation complex, while in [1] a deformation complex is given for $G = GL(n, \mathbb{C})$ and a higher-dimensional (varying) base X. For arbitrary (fixed) compact Kähler X and arbitrary reductive G, the dgla controlling the deformations of $(\mathbf{P}, \theta)$ is

$$\mathscr{C}^\bullet = \bigoplus_{p+r=\bullet} A^{0,p}(X, \mathrm{ad}\mathbf{P} \otimes \Omega^r_X), \tag{6}$$

with differential $\overline{\partial}_{\mathbf{P}} + \mathrm{ad}\theta$, see [23, §9], [21, §2], [22, §10]. The complex $\mathscr{C}^\bullet$ is the Dolbeault resolution of the complex from [1, 2]. For the case of $G = GL(n, \mathbb{C})$ and L-valued pairs, one replaces Ω^q_X with $\Lambda^q(L)$, see [19]. We note that the isomorphism $\mathrm{Def}_{\mathscr{C}^\bullet}(A) \simeq \mathrm{Def}_{(P,\theta)}(A)$ is obtained by mapping $[(s^{1,0}, s^{0,1})]$ to $(\ker(\overline{\partial} + s^{0,1}), \theta + s^{1,0})$, see [1, 19, 22].

We set $H(P, \theta) := \chi(\theta) \equiv \oplus_i p_i(\theta) \in \mathcal{B}$. Using the notation from §2.3, define the (infinitesimal) Hitchin map as a morphism (natural transformation) of deformation functors

$$H : \mathrm{Def}_{(\mathbf{P},\theta)} \to \mathrm{Def}_{H(\mathbf{P},\theta)}$$

by $H(A)(P_A, \theta_A) = \chi(\theta_A)$, $A \in \mathrm{Art}_{\mathbb{C}}$, and the obvious map on morphisms. See also [2, Remark 2.8 (iv)] or [4, §17.7]. While the two deformation functors at hand are controlled by dgla's

$$\mathrm{Def}_{\mathscr{B}^\bullet} \simeq \mathrm{Def}_{H(E,\theta)} \text{ and } \mathrm{Def}_{\mathscr{C}^\bullet} \simeq \mathrm{Def}_{(\mathbf{P},\theta)}, \tag{7}$$

H is not a dgla morphism, unless $G = (\mathbb{C}^\times)^N$, since it is not even linear. It is, however, induced by an L_∞-morphism, as we intend to show.

In this note we are concerned with infinitesimal considerations only, but we remark that the coarse moduli spaces of semi-stable Higgs bundles (whenever they exist) carry an amazingly rich geometry. We refer to [4, 11, 12, 21, 22, 23] for insight and discussion of global questions.

3. The Adjoint Quotient in L-infinity terms

3.1. Toy Model

If one sees the Hitchin map H as a "global analogue" of the adjoint quotient $\chi : \mathfrak{g} \to \mathbb{C}^N \simeq \mathfrak{g} /\!\!/ G$, then the Higgs field θ should be regarded as a "global analogue" of an element $v \in \mathfrak{g}$. In the present section we describe the morphism χ in L_∞ terms, and in Section 4 we modify suitably this "toy model" to obtain an L_∞-description of H.

Consider first the dgla $C^\bullet := \mathfrak{g} \otimes \mathbb{C}[\varepsilon]/\varepsilon^2 = \mathfrak{g} \oplus \mathfrak{g}[-1]$, with differential $d_0 = \varepsilon \mathrm{ad} v$. Since $d_1 = 0$ and $[C^1, C^1] = 0$, we have $\mathrm{MC}_{C^\bullet} = \mathfrak{g}$, i.e., $\mathrm{MC}_{C^\bullet}(A) = \mathfrak{g} \otimes \mathfrak{m}_A$, for all $A \in \mathrm{Art}_\mathbb{C}$. Moreover, the formula (1) for the gauge action reduces to $(\lambda, a) \mapsto e^{\mathrm{ad}\lambda}(v + a) - v$. We also recall from §2.3 the dgla $B^\bullet = \mathbb{C}^N[-1]$, with $\mathrm{MC}_{B^\bullet} = \mathbb{C}^N$. To see why is it appropriate to consider $C^\bullet$, we introduce the functor $\mathrm{Def}_{v,\mathfrak{g},G} : \mathrm{Art}_\mathbb{C} \to \mathrm{Sets}$,

$$\mathrm{Def}_{v,\mathfrak{g},G}(A) = \mathrm{Def}_{v,\mathfrak{g}}(A)/\exp(\mathfrak{g} \otimes \mathfrak{m}_A),$$

with the obvious transformation under morphisms of the coefficient ring. That is, $\mathrm{Def}_{v,\mathfrak{g},G}(A)$ is the quotient of the affine subspace $\{v\} + \mathfrak{g} \otimes \mathfrak{m}_A \subset \mathfrak{g} \otimes A$ under the natural affine action of $\exp(\mathfrak{g} \otimes \mathfrak{m}_A)$ ([8, §4.2]), which we briefly recall. There is a natural Lie bracket on $\mathfrak{g} \otimes A$, obtained by extending the bracket on $\mathfrak{g}$. The adjoint action of G on $\mathfrak{g}$ extends to an action on $\exp(\mathfrak{g} \otimes \mathfrak{m}_A)$, and we denote by G_A the semidirect product $\exp(\mathfrak{g} \otimes \mathfrak{m}_A) \rtimes G$. More intrinsically, if we consider G as the group of $\mathbb{C}$-points of a $\mathbb{C}$-algebraic group $\mathbf{G}$, then $G_A = \mathbf{G}(A)$. The subgroup $\exp(\mathfrak{g} \otimes \mathfrak{m}_A) \subset G_A$ acts, via the adjoint representation, on $\mathfrak{g} \otimes A$, and preserves the affine subspace $\{v\} + \mathfrak{g} \otimes \mathfrak{m}_A$. The affine action on $\mathfrak{g} \otimes \mathfrak{m}_A$ is $(\lambda, a) \mapsto e^{\mathrm{ad}\lambda}(v + a) - v$.

Thus we have a bijection $\mathrm{Def}_{C^\bullet}(A) \simeq \mathrm{Def}_{v,\mathfrak{g},G}(A)$, $a \mapsto v + a$ which induces an isomorphism $\mathrm{Def}_{C^\bullet} \simeq \mathrm{Def}_{v,\mathfrak{g},G}$, as all constructions are natural in the coefficient ring. Notice that $H^0(C^\bullet)$ is the centraliser of $v \in \mathfrak{g}$, so the functor $\mathrm{Def}_{C^\bullet}$ need not be representable. However, we have the following:

Proposition 3.1. *Let $\mathcal{K} \subset \mathfrak{g}$ be a linear complement to $Im(ad_v) \subset \mathfrak{g}$, and let $\widehat{\mathcal{O}}_{(\mathcal{K},0)}$ be its completed local ring at the origin. Then the functor $Hom_{alg}(\widehat{\mathcal{O}}_{K,0},\ )$ is a hull for $Def_{C^\bullet}$.*

Proof. By [9, Theorem 1.1], if a dgla $C^\bullet$ is equipped with a splitting and has finite-dimensional $H^k(C^\bullet)$, $k = 0, 1$, then it admits a hull $Kur \to \mathrm{Def}_{C^\bullet}$ by formal Kuranishi theory. In [9, Theorem 2.3] it is shown that under certain topological conditions $Kur = Hom_{alg}(\widehat{\mathcal{O}}_{(\mathcal{K},0)},\ )$, where $(\mathcal{K}, 0)$ is the germ of a complex-analytic space (Kuranishi space) and $\widehat{\mathcal{O}}$ is its completed local ring. In our case, $C^1 = \mathfrak{g}$, $d_1 = 0$ and $[C^1, C^1] = 0$, so by [9, Theorems 2.6 and 1.1] $\mathcal{K}$ exists and can be taken to be any linear complement to the coboundaries, i.e., any linear complement $\mathrm{Imad}_v \subset \mathfrak{g}$. □

Our next step is to construct an L_∞-morphism $h_\infty : C^\bullet \to B^\bullet = \mathbb{C}^N[-1]$, such that $\mathrm{MC}(h_\infty) : \mathrm{MC}_{C^\bullet} = \mathfrak{g} \to \mathrm{MC}_{B^\bullet} = \mathbb{C}^N$ gives the adjoint quotient. This involves two ingredients. First, as χ is given by homogeneous polynomials, Taylor's formula can be expressed conveniently by polarisation. Second, the derivatives of G-invariant polynomials satisfy extra relations. We discuss these technical properties in Section 3.2, and construct the promised L_∞-morphism in Section 3.3.

3.2. Polarisation and Invariant Polynomials

Let V be a finite-dimensional vector space. We have, for each $d, k \in \mathbb{N}$, a linear map

$$\mathscr{P}_d^{k,d-k} = \partial^k : S^d(V^\vee) \longrightarrow T^k(V^\vee)^{\Sigma_k} \otimes S^{d-k}(V^\vee),\ p \mapsto \partial^k p.$$

That is,

$$\mathscr{P}_d^{k,d-k}(p)(X_1 \otimes \ldots \otimes X_k \otimes v_1 \cdot \ldots \cdot v_{d-k}) = \mathcal{L}_{X_1} \ldots \mathcal{L}_{X_k}(p)(v_1 \cdot \ldots \cdot v_{d-k}),$$

where $\mathcal{L}_X$ denotes Lie derivative. Differently put, $\mathscr{P}_d^{k,d-k}(p)(X_1 \otimes \ldots \otimes X_k \otimes v^{d-k})$ is the coefficient in front of $t_1 \ldots t_k$ in the Taylor expansion of $p(v + \sum t_i X_i)$. For example, if $V = \mathfrak{gl}(r, \mathbb{C})$ and $p(A) = \mathrm{tr} A^d$, then

$$\mathscr{P}_d^{k,d-k}(p)(X_1 \otimes \ldots \otimes X_k \otimes A^{d-k}) = \frac{d!}{(d-k)!} \mathrm{tr}(X_1 \ldots X_k A^{d-k}).$$

In particular, $\mathscr{P}_d^{d,0} = \partial^d : S^d(V^\vee) \simeq \left(V^{\vee \otimes d}\right)^{\Sigma_d}$ is the usual polarisation map, identifying Σ_d invariants and coinvariants, and $p(X) = \frac{1}{d!}\partial^d p(X^{\otimes d})$. More generally,

$$(\partial^k p)(X_1 \otimes \ldots \otimes X_k \otimes v^{d-k}) = \frac{1}{(d-k)!}(\partial^d p)(X_1 \otimes \ldots \otimes X_k \otimes v^{\otimes d-k})$$

and by Taylor's formula

$$p(v + X) - p(v) = \sum_{k=1}^{\infty} \frac{1}{k!}(\partial^k p)(X^{\otimes k} \otimes v^{d-k}). \tag{8}$$

We prove two technical lemmas.

Lemma 3.2. *Let $p \in \mathbb{C}[\mathfrak{g}]^G$ be a homogeneous G-invariant polynomial of degree d. Then $(\partial p)(ad_X(v) \otimes v^{d-1}) = 0$, for all $v, X \in \mathfrak{g}$.*

Proof. The statement that $\frac{d}{dt}p(v+t\mathrm{ad}_X(v))|_{t=0}=0$ is just an infinitesimal form of the G-invariance of p. Alternatively, one can write the above expression as $\frac{1}{(d-1)!}$ times

$$(\partial^d p)(\mathrm{ad}_X(v)\otimes v^{\otimes d-1}) = \frac{1}{d}\frac{d}{dt}\left(\partial^d p\left(\left(Ad(e^{tX})v\right)^{\otimes d}\right)\right)\Big|_{t=0} = 0.$$

□

Lemma 3.3. *Let $V=\bigoplus_{i=0}^{k-1}V_i$, $F\in T^d(V^\vee)^\Sigma$, and $L\in\prod_i GL(V_i)$. The decomposition of V induces a decomposition of $S^d(V^\vee)$, indexed by ordered partitions of d of length k. The projection of $F\circ(L\otimes 1^{\otimes d-1})$ onto the subspace corresponding to $(d-k+1,1,\dots,1)$ maps $v\otimes X_1\otimes\dots\otimes X_{k-1}$ to*

$$\frac{d!}{(d-k)!}F(L(v)\otimes X_1\otimes\dots\otimes X_{k-1}\otimes v^{\otimes d-k})+$$

$$\sum_{\sigma\in S(1,k-2)}\frac{d!}{(d-k+1)!}F(L(X_{\sigma(1)})\otimes X_{\sigma(2)}\otimes\dots\otimes X_{\sigma(k)}\otimes v^{\otimes d-k}).$$

Proof. The proof amounts to expanding $F(L(v+\sum_i X_i), v+\sum_i X_i,\dots,v+\sum_i X_i)$ in powers of X_i, and counting the number of terms, containing exactly one of each X_i. □

Corollary 3.4. *Let $p\in\mathbb{C}[\mathfrak{g}]^G$ be a homogeneous G-invariant polynomial of degree d. Let $2\le k\le d$, and let $v,Y,X_1,\dots,X_{k-1}\in\mathfrak{g}$. Then*

$$\begin{aligned}&(\partial^k p)([Y,v]\otimes X_1\dots X_{k-1}\otimes v^{d-k})+\\ &\qquad\sum_{\sigma\in S(1,k-2)}(\partial^{k-1}p)([Y,X_{\sigma_1}]\otimes\dots\otimes X_{\sigma_{k-1}}\otimes v^{d-k+1})=0.\end{aligned}$$

Proof. We apply Lemma 3.3 to $F=(\partial^d p)$ and $L=\mathrm{ad}Y$ and use Lemma 3.2 to argue that $F\circ(L\otimes 1)$ is zero. □

In the next section we will apply the various operators $\mathscr{P}_d^{k,d-k}$ to sections of $\mathcal{A}^{0,\bullet}(\mathrm{ad}\mathbf{P}\otimes S^k\Omega^1_X)$ without changing the notation. For example, given sections s_i expressed locally as $s_i=\alpha_i\otimes X_i$ and $v\in H^0(X,\mathrm{ad}P\otimes\Omega^1_X)$, we write

$$(\partial^k p)(s_1\otimes\dots\otimes s_k\otimes v^{d-k})=\alpha_1\wedge\dots\wedge\alpha_k(\partial^k p)(X_1\otimes\dots\otimes X_k\otimes v^{d-k}).$$

3.3. The L-infinity Morphism

The main result of this section is the following

Proposition 3.5. *Let* $\mathbf{p}_0 : \overline{S^\bullet(C^\bullet[1])} \to \overline{S^\bullet(C^1)}$ *denote the homomorphism induced by the projection* $pr_2 : C^\bullet[1] = \mathfrak{g}[1] \oplus \mathfrak{g} \to C^1 = \mathfrak{g}$. *The collection of maps*

$$\bigoplus_i h_k^{d_i} = (\partial^k p_i)(\_\otimes v^{d_i-k}) \circ p_0 : \qquad S^k(C^\bullet[1]) \longrightarrow \mathbb{C}^N$$

$$(a_1,b_1)\cdot \ldots \cdot (a_k,b_k) \longmapsto \bigoplus_i (\partial^k p_i)(b_1 \otimes \ldots \otimes b_k \otimes v^{d_i-k})$$

induces an L_∞*-morphism* $h_\infty : C^\bullet \to B^\bullet = \mathbb{C}^N[-1]$. *Under the identifications* $MC_{B^\bullet} \simeq Def_{\chi(v),\mathbb{C}^n}$ *and* $MC_{C^\bullet} \simeq Def_{v,\mathfrak{g}}$, $MC(h_\infty) : MC_{C^\bullet} \to MC_{B^\bullet}$ *coincides with* $\chi : \mathfrak{g} \to \mathbb{C}^N$.

Proof. To show that this collection of maps determines an L_∞-morphism, it suffices to verify that for each fixed d_i, the maps $\{h_k^{d_i}\}$ determine an L_∞-morphism $C^\bullet \to \mathbb{C}[-1]$. We prove this in Lemma 3.6. Assuming that, let $s = (0,b) \in \mathrm{MC}_{C^\bullet}(A)$, $b \in \mathfrak{g} \otimes \mathfrak{m}_A$ for $A \in \mathrm{Art}_\mathbb{C}$. Then, by (5), $\mathrm{MC}(h_\infty)(s) = \sum_{d=1}^\infty \frac{1}{d!} h_\infty(s^d)$, which equals $\oplus_i (p_i(v+b) - p_i(v)) = \chi(v+b) - \chi(v)$ by (8). The specified identifications amount to affine transformations translating the origin, which carry $\mathrm{MC}(h_\infty)$ to the map $v + b \mapsto \chi(v+b)$, hence the last statement. □

Lemma 3.6. *Let* $p \in \mathbb{C}[\mathfrak{g}]^G$ *be a homogeneous polynomial of degree* d. *The collection of maps*

$$h_k^d = (\partial^k p)(\_\otimes v^{d-k}) \circ p_0 : \qquad S^k(C^\bullet[1]) \longrightarrow \mathbb{C}$$

$$(a_1,b_1)\cdot \ldots \cdot (a_k,b_k) \longmapsto (\partial^k p)(b_1 \otimes \ldots \otimes b_k \otimes v^{d-k})$$

induces an L_∞*-morphism*

$$h_\infty^d : C^\bullet \longrightarrow \mathbb{C}[-1].$$

Proof. We start with condition (2). The differentials of the two dgla's are, respectively, $\mathrm{ad}v$ and 0, so we have to show that, for any $s = (a,b) \in \mathfrak{g}^{\oplus 2}$, $h_1^d([v,s]) = 0$. But this means $(\partial p)([v,b] \otimes v^{d-1}) = 0$, which is the conclusion of Lemma 3.2. We turn to (3), whose right hand side is identically zero (since $B^\bullet$ is formal). The left side is zero on $S^k(C^1)$, since $[C^1, C^1] = 0$. It is also zero on $S^r(C^0)\cdot S^{k-r}(C^1)$ for $r \geq 2$, since h_k^d factors through p_0. So we only have to verify (3) on $C^0 \cdot S^{k-1}(C^1)$, in which case Q_k^{k-1} contributes via the bracket and Q_k^k via $\mathrm{ad}v$. Take homogeneous elements $s_j = (0, b_j)$, $j \geq 2$ and $s_1 = (a, 0)$. In the first summand of (4), unshuffles with $\sigma(1) \neq 1$ give zero, while $\sigma(1) = 1$

means $\sigma = id$, so we have $h_k^d\left([v, s_1] \cdot s_2 \cdot \ldots \cdot s_k\right) = (-1)(\partial^k p_i)([a,v] \otimes b_2 \otimes \ldots \otimes b_k \otimes v^{d-k})$. The second summand of (4) is $h_{k-1}^d \circ Q_k^{k-1}(s_1 \cdot \ldots \cdot s_k)$ and the non-vanishing terms correspond to $(2, k-2)$ unshuffles σ, for which $\sigma(1) = 1$. Hence the summation is in fact over $(1, k-2)$ unshuffles and we have

$$h_{k-1}^d \left(\sum_{\sigma \in S(1,k-2)} (-1)\epsilon(\sigma)[s_1, s_{\sigma(1)}] \cdot \ldots \cdot s_{\sigma(k-1)} \right) =$$

$$(-1) \sum_{\sigma \in S(1,k-2)} (\partial^{k-1} p_i) \left([a, b_{\sigma(1)}] \otimes \ldots \otimes b_{\sigma(k-1)} \otimes v^{d-k+1}\right).$$

Note that $C^1 = C^\bullet[1]^0$, so $\epsilon(\sigma) = 1$. The two summands add up to zero by Corollary 3.4. ☐

4. The Hitchin Map

We prove now the two main results of this note by suitably adapting the calculation of the previous section, thus extending the results of [19] to arbitrary reductive structure groups.

4.1. Proof of Theorem 1.2

Proof. To prove that the collection$\{h_k\}$ determines an L_∞-morphism, it suffices to prove that for each fixed homogeneous polynomial p_i of degree d_i, the given collection of maps induces an L_∞-morphism $h_\infty^{d_i} : \mathscr{C}^\bullet \to A^{0,\bullet}(S^{d_i}\Omega_X^1)[-1]$. This is shown in Lemma 4.1 below. Assuming that, suppose $s = (s', s'') \in \mathrm{MC}_{\mathscr{C}^\bullet}(A)$, $A \in \mathrm{Art}_{\mathbb{C}}$. By (5) $\mathrm{Def}(h_\infty)(s) = \sum_{d=1}^\infty \frac{1}{d!} h_\infty(s^d)$, which by formula (8) equals $\oplus_i p_i(\theta + s') - p_i(\theta) = H(P_A, \theta_A) - H(P, \theta)$. This is exactly what we want to prove, in view of the identification (7), which amounts to "shifting the origin". ☐

Lemma 4.1. *Let $p \in \mathbb{C}[\mathfrak{g}]^G$ be a homogeneous polynomial of degree d. Let $\mathbf{p}_0 : \overline{S^\bullet(\mathscr{C}^\bullet)} \to \overline{S^\bullet\left(A^{0,\bullet}(ad\mathbf{P} \otimes \Omega_X^1)\right)}$ denote the homomorphism induced by $\bigoplus_{p+q=\bullet} s^{p,q} \mapsto s^{1,q}$, where $s^{p,q} \in A^{0,q}(ad\mathbf{P} \otimes \Omega_X^p)$. Then the collection of maps*

$$h_k^d = (\partial^k p)(\_ \otimes \theta^{d-k}) \circ p_0 : \qquad S^k(\mathscr{C}^\bullet[1]) \longrightarrow A^{0,\bullet}(S^d\Omega_X^1)$$

$$\bigoplus_{p_1,q_1} s_1^{p_1,q_1} \cdot \bigoplus_{p_2,q_2} s_2^{p_2,q_2} \cdot \ldots \cdot \bigoplus_{p_k,q_k} s_k^{p_k,q_k} \longmapsto \sum_{q_1,\ldots,q_k} (\partial^k p)(s_1^{1,q_1} \otimes \ldots \otimes s_k^{1,q_k} \otimes \theta^{d-k})$$

induces an L_∞-morphism

$$h_\infty^d : \mathscr{C}^\bullet = \bigoplus_{r+s=\bullet} A^{0,r}(ad\mathbf{P} \otimes \Omega_X^s) \to A^{0,\bullet}(S^d\Omega_X^1)[-1].$$

Proof. We check the conditions (2),(3). The differentials are $\overline{\partial}_{\mathbf{P}} + \mathrm{ad}\theta$ and $\overline{\partial}_{\mathbf{P}}$, so (2) is equivalent to $(\partial p)([\theta, s] \otimes \theta^{d-1}) = 0$, which holds by Lemma 3.2. Next assume $k \geq 2$. Since by definition h^d_k factors through $\mathbf{p}_0$, both sides of (3) are identically zero, except possibly for *two* cases. Case 1: when evaluated on $S^k\left(A^{0,\bullet}(\mathrm{ad}P \otimes \Omega^1)\right)$ and Case 2: when evaluated on $A^{0,\bullet}(\mathrm{ad}\mathbf{P}) \cdot S^{k-1}\left(A^{0,\bullet}(\mathrm{ad}\mathbf{P} \otimes \Omega^1)\right)$. Notice that in $\mathscr{C}^\bullet[1]$, the degree of a homogeneous element in $A^{0,n}(\mathrm{ad}P \otimes \Omega^1_X)$ is n. We start with Case 1, evaluating on decomposable homogeneous elements $s_i = \alpha_i \otimes X_i$, $i = 1 \dots k$. Since $[s_{\sigma 1}, s_{\sigma 2}]$ and $\mathrm{ad}\theta(s_{\sigma 1})$ belong to $A^{0,\bullet}(\mathrm{ad}\mathbf{P} \otimes \Omega^2_X)$, they do not contribute to the left side of (4). And since $\sum_{\sigma \in S(1,k-1)} \epsilon(\sigma)\overline{\partial}(\alpha_{\sigma(1)}) \wedge \dots \wedge \alpha_{\sigma(k)} = \overline{\partial}(\alpha_1 \wedge \dots \wedge \alpha_k)$, the left side of (4) gives

$$-\overline{\partial}(\alpha_1 \wedge \dots \wedge \alpha_k) \otimes (\partial^k p)(X_1 \otimes \dots \otimes X_k \otimes \theta^{d-k}) = \hat{q}_1 \circ h^d_k(s_1 \cdot \dots \cdot s_k),$$

which we wanted to show. Next we proceed to Case 2, and take decomposable homogeneous elements $s_i = \alpha_i \otimes X_i$, $s_1 \in A^{0,\bullet}(\mathrm{ad}\mathbf{P})$, $s_2, \dots, s_k \in A^{0,\bullet}(\mathrm{ad}\mathbf{P} \otimes \Omega^1_X)$. The right hand side of (4) is zero on their product, so we just compute the left side. The terms with $\sigma(1) \neq 1$ are identically zero, and $\sigma(1) = 1$ implies $\sigma = id$, so we obtain

$$\begin{aligned} &h^d_k\left([\theta, s_1] \cdot s_2 \cdot \dots \cdot s_k\right) = \\ &\qquad = (-1)^{\deg s_1} \alpha_1 \wedge \dots \wedge \alpha_k (\partial^k p)([X_1, \theta] \otimes X_2 \otimes \dots \otimes X_k \otimes \theta^{d-k}). \end{aligned}$$

The non-vanishing contributions from $h^d_{k-1} \circ Q^{k-1}_k$ in (4) correspond to $(2, k-2)$ unshuffles for which $\sigma_1 = 1$, so the summation is in fact over $(1, k-2)$ unshuffles and we have

$$h^d_{k-1}\left(\sum_{\sigma \in S(1,k-2)} (-1)^{\deg s_1} \epsilon(\sigma)[s_1, s_{\sigma(1)}] \cdot \dots \cdot s_{\sigma(k-1)}\right).$$

By the shift, the Koszul sign is traded for reordering the forms and we get

$$(-1)^{\deg s_1} \alpha_1 \wedge \dots \wedge \alpha_k \sum_{\sigma \in S(1,k-1)} (\partial^{k-1} p)([X_1, X_{\sigma(1)}] \otimes \dots \otimes X_{\sigma(k-1)} \otimes \theta^{d-k+1}).$$

Then the sum of the two terms is zero by Corollary 3.4. □

4.2. Obstructions to smoothness

While Higgs bundles on *curves* have been extensively studied, fairly little is known about their moduli if $\dim X > 1$, apart from the general results of [22], partially due to scarcity of examples. By formality ([21, Lemma 2.2]), Simpson's moduli spaces have at most quadratic singularities ([22, Theorem 10.4]).

It is known that whenever $H^2(\mathscr{C}^\bullet) = 0$, the functor $\mathrm{Def}_{\mathscr{C}^\bullet}$ is smooth (the representing complete local algebra is regular), see [2, Theorem 3.1], [3, Proposition 3.7], [1, Remark 2.8]. We recall now the description of the obstruction space $O_{\mathrm{Def}_{\mathscr{C}^\bullet}} \subset H^2(\mathscr{C}^\bullet)$.

Recall ([5], [18, §4]) that an *obstruction theory* for a deformation functor $F : \mathrm{Art}_{\mathbb{C}} \to \mathrm{Sets}$ is a pair (V, v). Here V is a vector space (*obstruction space*), and v assigns to any small extension $e : 0 \longrightarrow M \longrightarrow B \longrightarrow A \longrightarrow 0$, an *obstruction map* $v_e : F(A) \to V \otimes M$, respecting base change, with $\mathrm{Im}\,(F(B) \to F(A)) \subset \ker v_e$. The obstruction theory is *complete*, if this containment is an equality. A *universal* obstruction theory is an obstruction theory (O_F, o), admitting a unique morphism to any other obstruction theory (V, v). The vector space O_F is called *the obstruction space* of F.

Proof of Theorem 1.1. The proof is essentially a standard argument in deformation theory, and can be considered as a form of the so-called "Kodaira principle". By [18, Theorem 4.6 and Corollary 4.8] (see also [5]), any deformation functor F admits a universal obstruction theory, and, if (V, v) is any complete obstruction theory, then O_F is isomorphic to the space, generated by $v_e(F(A))$, where e ranges over all principal (i.e., with $M = \mathbb{C}$) small extensions. By [18, Example 4.4], for any dgla L, the functor $F = \mathrm{MC}_L$ admits a complete obstruction theory $(H^2(L), v)$. Here the map $v_e : \mathrm{MC}_L(A) \to H^2(L) \otimes M$ is defined by $v_e(x) = [h]$, where $h = d\widetilde{x} + \frac{1}{2}[\widetilde{x}, \widetilde{x}]$, and $\widetilde{x} \in \mathrm{MC}_L(B)$ is a lift of $x \in \mathrm{MC}_L(A)$. Also, by [18, Corollary 4.13], the functors MC_L and Def_L have isomorphic obstruction theories. In particular, $O_{\mathrm{Def}_L} \subset H^2(L)$ and if L *abelian*, then $O_{\mathrm{Def}_L} = (0)$. Now consider the abelian dgla $\mathscr{B}^\bullet$ and $h_\infty : \mathscr{C}^\bullet \to \mathscr{B}^\bullet$. By equation (2), h_1 is a morphism of complexes, and one can show ([16]) that $H^2(h_1)$ is a morphism of obstruction spaces, hence the result.

There is a more direct argument if X is Kähler and G is semi-simple (so that $d_i = 0$ is not an exponent), or if $H^1(X, \mathcal{O}_X) = 0$. Indeed, in that case $\mathscr{B}^\bullet \simeq_{qis} \mathbb{C}^N[-1]$ (see Section 2.3), so $H^2(\mathscr{B}^\bullet) = (0) = H^2(h_1)(O_{\mathscr{C}^\bullet})$. $\square$

Acknowledgements

I am grateful to Anya Kazanova for encouragement, and to E. Cattani, C. Daenzer, S. Guerra, E. Markman, T. Pantev and J. Stasheff for support and helpful comments. I would like to thank ICTP and the organisers of the *Conference on Hodge Theory and Related Topics* (2010), during which an early version of this material was completed. This work was partially supported by a CERES-CEI Research Fellowship (FP-7), Grant agreement 229660.

References

[1] I. Biswas, *A remark on a deformation theory of Green and Lazarsfeld*, J. Reine Angew. Math **449** (1994), 103–124.

[2] I. Biswas and S. Ramanan, *An infinitesimal study of the moduli of Hitchin pairs*, J. London Math. Soc. (2) **49** (1994), no. 2, 219–231.

[3] I. Biswas and G. Schumacher, *Geometry of moduli spaces of Higgs bundles*, Comm. Anal. Geom. **14** (2006), no. 4, 765–793.

[4] R. Donagi and D. Gaitsgory, *The gerbe of Higgs bundles*, Transform. Groups **7** (2002), no. 2, 109–153.

[5] B. Fantechi and M. Manetti, *Obstruction calculus for functors of Artin rings. I*, J. Algebra **202** (1998), no. 2, 541–576.

[6] D. Fiorenza and M. Manetti, *L1-algebras, cartan homotopies and period maps*, (2006), ArXiv AG/0605297.

[7] A. Fujiki, *Hyper-Kähler structure on the moduli space of flat bundles*, Prospects in complex geometry (Katata and Kyoto, 1989), Lecture Notes in Math., vol. 1468, Springer, Berlin, 1991, pp. 1–83.

[8] W. Goldman and J. Millson, *The deformation theory of representations of fundamental groups of compact Kähler manifolds*, Inst. Hautes Études Sci. Publ. Math. (1988), no. 67, 43–96.

[9] W. M. Goldman and J. J. Millson, *The homotopy invariance of the Kuranishi space*, Illinois J. Math. **34** (1990), no. 2, 337–367.

[10] Ph. Griffiths and J. Harris, *Principles of algebraic geometry*, Wiley-Interscience [John Wiley & Sons], New York, 1978, Pure and Applied Mathematics.

[11] N. Hitchin, *The self-duality equations on a Riemann surface*, Proc. London Math. Soc. **3** (1987), no. 55, 59–126.

[12] N. Hitchin, *Stable bundles and integrable systems*, Duke Math. J. **54** (1987), no. 1, 91–114.

[13] M. Kontsevich, *Topics in algebra: deformation theory*, Notes by A.Weinstein, 1994.

[14] M. Kontsevich, *Deformation quantization of Poisson manifolds*, Lett. Math. Phys. **66** (2003), no. 3, 157–216.

[15] T. Lada and J. Stasheff, *Introduction to SH Lie algebras for physicists*, Internat. J. Theoret. Phys. **32** (1993), no. 7, 1087–1103.

[16] M. Manetti, *Deformation theory via differential graded Lie algebras*, Algebraic Geometry Seminars, 1998–1999 (Italian) (Pisa), Scuola Norm. Sup., Pisa, 1999, pp. 21–48.

[17] M. Manetti, *Lectures on deformations of complex manifolds*, Rend. Mat. Appl. **7** (2004), no. 24, 1–183.

[18] M. Manetti, *Differential graded Lie algebras and formal deformation theory*, Algebraic geometry—Seattle 2005. Part 2, Proc. Sympos. Pure Math., vol. 80, Amer. Math. Soc., Providence, RI, 2009, pp. 785–810.

[19] E. Martinengo, *Infinitesimal deformations of Hitchin pairs and Hitchin map*, Internat. J. Math. **23** (2012), no. 7, 1250053, 30.

[20] M. Schlessinger and J. Stasheff, *Deformation theory and rational homotopy*

type, (2012), ArXiv QA/1211.1647.

[21] C. Simpson, *Higgs bundles and local systems*, Inst. Hautes Études Sci. Publ. Math. (1992), no. 75, 5–95.

[22] C. Simpson, *Moduli of representations of the fundamental group of a smooth projective variety, II*, Inst. Hautes Études Sci. Publ. Math. (1994), no. 80, 5–79.

[23] C. Simpson, *The Hodge filtration on nonabelian cohomology*, Algebraic geometry—Santa Cruz 1995, Proc. Sympos. Pure Math., vol. 62, Amer. Math. Soc., Providence, RI, 1997, pp. 217–281.

Author's address:

Peter Dalakov
Institute of Mathematics & Informatics, Bulg. Acad. Sci.
Acad. G. Bontchev str., bl.8
1113 Sofia, Bulgaria
E-mail: `dalakov@math.bas.bg`

Received October 24, 2013

Rend. Istit. Mat. Univ. Trieste
Volume 46 (2014), 19–41

Decompositions of PDE over Cayley-Dickson algebras

Sergey V. Ludkowski

Abstract. *The article is devoted to decompositions of partial differential operators (PDO) into products of lower order PDO over octonions and Cayley-Dickson algebras. Partial differential equations (PDE) with generalized and discontinuous coefficients are considered as well.*

Keywords: hypercomplex numbers, Cayley-Dickson algebra, operator, partial differential equation, generalized function
MS Classification 2010: 30G35; 17A05; 35D05; 46F10

1. Introduction

In the preceding articles of the author [21, 23] line integration methods of partial differential equations (PDE) over octonions and Cayley-Dickson algebras were described [1, 2, 10]. Such technique is based on decompositions of partial differential operators (PDO) into products of PDO of lower order. The present paper is devoted to investigations of such decompositions. Besides PDO with differentiable coefficients, PDO with generalized and discontinuous coefficients are studied as well. This permits to integrate not only elliptic, but also hyperbolic and parabolic PDE of the second and higher orders developing further Dirac's approach. It is important for many-sided applications of PDE [4, 9, 11, 25, 28, 29], where differential equations over the real and complex fields were considered. But recently substantial interest was evoked by PDE over Clifford algebras [5, 6, 7, 8, 13].

In previous articles of the author (super)-differentiable functions of Cayley-Dickson variables and their non-commutative line integrals were investigated [14, 15, 19, 24]. Furthermore, in the works [18, 20, 22] differential equations and their systems over octonions and quaternions were studied.

Main results of this paper are obtained for the first time.

2. Decompositions of PDE

Henceforward, notations and definitions of the article [21] are used.

2.1. Transformations of the first order PDO over the Cayley-Dickson algebras

We consider the first order Dirac's type operator in the form:

$$\Upsilon f = \sum_{j=0}^{2^v-1} (\partial f / \partial z_j)\, \eta_j(z)\,, \tag{1}$$

with either $\eta_j(z) = i_j^* \psi_j(z)$ or $\eta_j(z) = \phi_j^*(z) \in \mathcal{A}_v$ for each j (see Theorems 2.4.1 and 2.5.2 in [21]). To simplify the operator Υ one can use a change of variables. For it we seek the change of variables $x = x(z)$ satisfying the conditions:

$$\sum_{j=0}^{2^v-1} (\partial x_l / \partial z_j)\, \omega_j(z) = t_l\,, \tag{2}$$

where $t_l \in \mathcal{A}_v$ is a constant for each l, while each yet unknown function ω_j is supposed to be z-differentiable subjected to the condition that the resulting matrix Ω is not degenerate, i.e. its rows are real-independent as vectors (see below), when η_j is not identically zero. Certainly, $(\partial x_l / \partial z_j) \in \mathbf{R}$ are real partial derivatives, since x_l and z_j are real coordinates, where $z = z_0 i_0 + \ldots + z_{2^v-1} i_{2^v-1}$, while $x_j, z_j \in \mathbf{R}$ for each j, whilst $i_0,\ldots,i_{2^v-1}$ are the standard basis generators of the Cayley-Dickson algebra $\mathcal{A}_v$ over the real field $\mathbf{R}$ so that $i_0 = 1$, $i_j^2 = -1$ and $i_j^* = -i_j$ and $i_j i_k = -i_k i_j$ for each $j \ne k \ge 1$. We suppose that the functions $\eta_j(z)$ are linearly independent over the real field for each Cayley-Dickson number z in the domain U in $\mathcal{A}_v$. Using the standard basis of generators $\{i_j :\ j = 0, \ldots, 2^v - 1\}$ of the Cayley-Dickson algebra $\mathcal{A}_v$ and the decompositions

$$\omega_j = \sum_k \omega_{j,k}\, i_k \text{ and } t_j = \sum_k t_{j,k}\, i_k$$

with real elements $\omega_{j,k}$ and $t_{j,k}$ for all j and k we rewrite system (2) in the matrix form:

$$(\partial x_l / \partial z_j)_{l,j=0,\ldots,2^v-1} \Omega = T, \tag{3}$$

where

$$\Omega = (\omega_{j,k})_{j,k=0,\ldots,2^v-1}, \quad T = (t_{j,k})_{j,k=0,\ldots,2^v-1}.$$

It is supposed that the functions $\omega_j(z)$ are arranged into the family $\{\omega_j : j = 0, \ldots, 2^v - 1\}$ as above and are such that the matrix $\Omega = \Omega(z)$ is non-degenerate for all Cayley-Dickson numbers z in the domain U. For example, this is always the case, when $|\omega_j(z)| > 0$ and $Re[\omega_j(z) \omega_k(z)^*] = 0$ for each $j \ne k$ and $z \in U$. There, particularly $\omega_j(z) = \eta_j(z)$ can also be taken for all $j = 0, \ldots, 2^v - 1$ and $z \in U$. Therefore, equality (3) becomes equivalent to

$$(\partial x_l / \partial z_j)_{l,j=0,\ldots,2^v-1} = T \Omega^{-1}\,. \tag{4}$$

We take the real matrix $T = T(z)$ of the same rank as the real matrix $\Omega = (\omega_{j,k})_{j,k=0,\dots,2^v-1}$. Thus (4) is the linear system of PDE of the first order over the real field $\mathbf{R}$. It can be solved by the standard methods (see, for example, [25]).

We remind how each linear partial differential equation (3) or (4) can be resolved. One writes it in the form:

$$X_1(x_1,\dots,x_n,u)\partial u/\partial x_1 + \dots + X_n(x_1,\dots,x_n,u)\partial u/\partial x_n \\ = R(x_1,\dots,x_n,u)\,, \tag{5}$$

with u and $x_1,\dots,x_n$ here instead of x_l and $z_0,\dots,z_{2^v-1}$ in (3) seeking simultaneously a suitable function R corresponding to $t_{l,k}$. A function $u = u(x_1,\dots,x_n)$ continuous with its partial derivatives $\partial u/\partial x_1$, …,$\partial u/\partial x_n$ and defined in some domain V of variables $x_1,\dots,x_n$ in $\mathbf{R}^n$ making (5) the identity is called a solution of this linear equation. If the right side $R = 0$ identically, then the equation is called homogeneous. A solution $u = const$ of the homogeneous equation

$$X_1(x_1,\dots,x_n,u)\partial u/\partial x_1 + \dots + X_n(x_1,\dots,x_n,u)\partial u/\partial x_n = 0 \tag{6}$$

is called trivial. Then one composes the equations:

$$dx_1/X_1(x) = dx_2/X_2(x) = \dots = dx_n/X_n(x)\,, \tag{7}$$

where $x = (x_1,\dots,x_n)$. This system is called the system of ordinary differential equations in the symmetric form corresponding to the homogeneous linear equation in partial derivatives. It is supposed that the coefficients $X_1,\dots,X_n$ are defined and continuous together with their first order partial derivatives by $x_1,\dots,x_n$ and that $X_1,\dots,X_n$ are not simultaneously zero in a neighborhood of some point x^0. Such point x^0 is called non singular. For example when the function X_n is non-zero, then system (7) can be written as:

$$dx_1/dx_n = X_1/X_n, \dots, dx_{n-1}/dx_n = X_{n-1}/X_n\,. \tag{8}$$

A system of n differential equations

$$dy_k/dx = f_k(x,y_1,\dots,y_n)\,, \quad k = 1,\dots,n\,, \tag{9}$$

is called normal of the n-th order. It is called linear if all functions f_k depend linearly on $y_1,\dots,y_n$. Any family of functions $y_1,\dots,y_n$ satisfying the system of n differential equations (9) in some interval (a,b) is called its solution.

A function $g(x,y_1,\dots,y_n)$ different from a constant identically and differentiable in a domain D and such that its partial derivatives $\partial g/\partial y_1$,…,$\partial/\partial y_n$ are not simultaneously zero in D is called an integral of system (9) in D if the total differential $dg = (\partial g/\partial x)dx + (\partial g/\partial y_1)dy_1 + \dots + (\partial g/\partial y_n)dy_n$ becomes identically zero, when the differentials dy_k are substituted on their values from (9), that is $(\partial g(x,y)/\partial x) + (\partial g/\partial y_1) f_1(x,y) + \dots + (\partial g(x,y)/\partial y_n) f_n(x,y) = 0$ for each $(x,y) \in D$, where $y = (y_1, \dots, y_n)$. The equality $g(x,y) = const$ is called the first integral of system (9). Thus system (8) satisfies conditions of the theorem about an existence of integrals of the normal system.

It is supposed that each function $f_k(x,y)$ is continuous on D and satisfies the Lipschitz conditions by variables $y_1, \dots, y_n$:

$$(L) \qquad |f_k(x,y) - f_k(x,z)| \le C_k |y - z|$$

for all (x,y) and $(x,z) \in D$, where C_k are positive constants. Then system (9) has exactly n independent integrals in some neighborhood D^0 of a marked point (x^0, y^0) in D, when the Jacobian $\partial(g_1, \dots, g_n)/\partial(y_1, \dots, y_n)$ is not zero on D^0 (see Section 5.3.3 [25]).

In accordance with Theorem 12.1,2 [25] each integral of system (7) is a non-trivial solution of equation (6) and vice versa each non-trivial solution of equation (6) is an integral of (7). If $g_1(x_1, \dots, x_n), \dots, g_{n-1}(x_1, \dots, x_n)$ are independent integrals of (7), then the function

$$u = \Phi(g_1, \dots, g_{n-1}) \,, \tag{10}$$

where Φ is an arbitrary function continuously differentiable by $g_1, \dots, g_{n-1}$, is the solution of (6). A solution provided by formula (10) is called a general solution of equation (6).

To the non-homogeneous equation (5) the system

$$dx_1/X_1 = \dots = dx_n/X_n = du/R \,, \tag{11}$$

is posed. System (11) gives n independent integrals $g_1, \dots, g_n$ and the general solution

$$\Phi(g_1(x_1, \dots, x_n, u), \dots, g_n(x_1, \dots, x_n, u)) = 0 \tag{12}$$

of (5), where Φ is any continuously differentiable function by $g_1, \dots, g_n$. If the latter equation is possible to resolve relative to u this gives the solution of (5) in the explicit form $u = u(x_1, \dots, x_n)$ which generally depends on Φ and $g_1, \dots, g_n$. Therefore, formula (12) for different R and u and X_j corresponding to $t_{l,k}$ and x_l and $\omega_{j,k}$ respectively can be satisfied, so that to solve equation (3) or (4), where the variables x_j are used in (12) instead of z_j in (3) and (4), $k = 0, \dots, 2^v - 1$.

Thus after the change of the variables the operator Υ takes the form:

$$\Upsilon f = \sum_{j=0}^{2^v - 1} (\partial f/\partial x_j)\, t_j \tag{13}$$

with constants $t_j \in \mathcal{A}_v$. Undoubtedly, the operator Υ with $j = 0, \dots, n$, $2^{v-1} \le n \le 2^v - 1$ instead of $2^v - 1$ can also be reduced to the form $\Upsilon f =$

$\sum_{j=0}^{n}(\partial f/\partial x_j)t_j$, when the rank is $rank(\omega_{j,k}) = n+1$ in a basis of generators $N_0,\dots,N_n$, where $N_0,\dots,N_{2^v-1}$ is a generator basis of the Cayley-Dickson algebra $\mathcal{A}_v$ over the real field $\mathbf{R}$.

In particular, if the rank is $rank(\omega_{j,k}) = m \le 2^v$ and a matrix T contains the unit upper left $m \times m$ block and zeros outside it, then $t_j = N_j$ for each $j = 0,\dots,m-1$ can be chosen.

One can mention that direct algorithms of Theorems 2.4.1 and 2.5.2 [21] may be simpler for finding the anti-derivative operator $\mathcal{I}_\Upsilon$, than this preliminary transformation of the partial differential operator Υ to the standard form (13).

2.2. Some notations

Let X and Y be two $\mathbf{R}$ linear normed spaces which also are left and right $\mathcal{A}_r$ modules, where $1 \le r$. Let Y be complete relative to its norm. We put $X^{\otimes k} := X \otimes_{\mathbf{R}} \dots \otimes_{\mathbf{R}} X$ to be the k times ordered tensor product over $\mathbf{R}$ of X. By $L_{q,k}(X^{\otimes k}, Y)$ we denote a family of all continuous k times $\mathbf{R}$ poly-linear and $\mathcal{A}_r$ additive operators from $X^{\otimes k}$ into Y. Then $L_{q,k}(X^{\otimes k}, Y)$ is also a normed $\mathbf{R}$ linear and left and right $\mathcal{A}_r$ module complete relative to its norm. In particular, $L_{q,1}(X,Y)$ is denoted also by $L_q(X,Y)$.

We present a normed space X as the direct sum $X = X_0 i_0 \oplus \dots \oplus X_{2^r-1} i_{2^r-1}$, where $X_0,\dots,X_{2^r-1}$ are pairwise isomorphic real normed spaces. Moreover, if $A \in L_q(X,Y)$ and $A(xb) = (Ax)b$ or $A(bx) = b(Ax)$ for each $x \in X_0$ and $b \in \mathcal{A}_r$, then an operator A we call right or left $\mathcal{A}_r$-linear respectively.

An $\mathbf{R}$ linear space of left (or right) k times $\mathcal{A}_r$ poly-linear operators is denoted by $L_{l,k}(X^{\otimes k}, Y)$ (or $L_{r,k}(X^{\otimes k}, Y)$ respectively).

As usually a support of a function $g : S \to \mathcal{A}_r$ on a topological space S is by the definition $supp(g) = cl\{t \in S : \ g(t) \ne 0\}$, where the closure ($cl$) is taken in S.

We consider a space of test function $\mathcal{D} := \mathcal{D}(\mathbf{R}^n, Y)$ consisting of all infinite differentiable functions $f : \mathbf{R}^n \to Y$ on $\mathbf{R}^n$ with compact supports.

The following convergence is considered. A sequence of functions $f_n \in \mathcal{D}$ tends to zero, if all f_n are zero outside some compact subset K in the Euclidean space $\mathbf{R}^n$, while on it for each $k = 0,1,2,\dots$ the sequence $\{f_n^{(k)} : \ n \in \mathbf{N}\}$ converges to zero uniformly. Here as usually $f^{(k)}(t)$ denotes the k-th derivative of f, which is a k times $\mathbf{R}$ poly-linear symmetric operator from $(\mathbf{R}^n)^{\otimes k}$ to Y, that is $f^{(k)}(t).(h_1,\dots,h_k) = f^{(k)}(t).(h_{\sigma(1)},\dots,h_{\sigma(k)}) \in Y$ for each $h_1,\dots,h_k \in \mathbf{R}^n$ and every transposition $\sigma : \{1,\dots,k\} \to \{1,\dots,k\}$, σ is an element of the symmetric group S_k, $t \in \mathbf{R}^n$. For convenience one puts $f^{(0)} = f$. In particular, $f^{(k)}(t).(e_{j_1},\dots,e_{j_k}) = \partial^k f(t)/\partial t_{j_1} \dots \partial t_{j_k}$ for all $1 \le j_1,\dots,j_k \le n$, where $e_j = (0,\dots,0,1,0,\dots,0) \in \mathbf{R}^n$ with 1 on the j-th place.

Such convergence in $\mathcal{D}$ defines closed subsets in this space $\mathcal{D}$, their complements by the definition are open, that gives the topology on $\mathcal{D}$. The space $\mathcal{D}$

is $\mathbf{R}$ linear and right and left $\mathcal{A}_r$ module.

By a generalized function of class $\mathcal{D}' := [\mathcal{D}(\mathbf{R}^n, Y)]'$ we call a continuous $\mathbf{R}$-linear $\mathcal{A}_r$-additive function $g : \mathcal{D} \to \mathcal{A}_r$. The set of all such functionals is denoted by $\mathcal{D}'$. That is, g is continuous, if for each sequence $f_n \in \mathcal{D}$, converging to zero, a sequence of numbers $g(f_n) =: [g, f_n) \in \mathcal{A}_r$ converges to zero while n tends to the infinity.

A generalized function g is zero on an open subset V in $\mathbf{R}^n$, if $[g, f) = 0$ for each $f \in \mathcal{D}$ equal to zero outside V. By a support of a generalized function g is called the family, denoted by $supp(g)$, of all points $t \in \mathbf{R}^n$ such that in each neighborhood of each point $t \in supp(g)$ the functional g is different from zero. The addition of generalized functions g, h is given by the formula:

$$[g + h, f) := [g, f) + [h, f)\,. \tag{14}$$

The multiplication $g \in \mathcal{D}'$ on an infinite differentiable function w is given by the equality:

$$[gw, f) = [g, wf) \tag{15}$$

either for $w : \mathbf{R}^n \to \mathcal{A}_r$ and each test function $f \in \mathcal{D}$ with a real image $f(\mathbf{R}^n) \subset \mathbf{R}$, where $\mathbf{R}$ is embedded into Y; or $w : \mathbf{R}^n \to \mathbf{R}$ and $f : \mathbf{R}^n \to Y$. A generalized function g' prescribed by the equation:

$$[g', f) := -[g, f') \tag{16}$$

is called a derivative g' of a generalized function g, where $f' \in \mathcal{D}(\mathbf{R}^n, L_q(\mathbf{R}^n, Y))$, $g' \in [\mathcal{D}(\mathbf{R}^n, L_q(\mathbf{R}^n, Y))]'$.

Another space $\mathcal{B} := \mathcal{B}(\mathbf{R}^n, Y)$ of test functions consists of all infinite differentiable functions $f : \mathbf{R}^n \to Y$ such that the limit $\lim_{|t| \to +\infty} |t|^m f^{(j)}(t) = 0$ exists for each $m = 0, 1, 2, \dots$, $j = 0, 1, 2, \dots$. Then analogously, a sequence $f_n \in \mathcal{B}$ is called converging to zero, if the sequence $|t|^m f_n^{(j)}(t)$ converges to zero uniformly on $\mathbf{R}^n \setminus B(\mathbf{R}^n, 0, R)$ for each $m, j = 0, 1, 2, \dots$ and each $0 < R < +\infty$, where $B(Z, z, R) := \{y \in Z : \ \rho(y, z) \le R\}$ denotes a ball with center at z of radius $R > 0$ in a metric space Z with a metric ρ, whilst the Euclidean space $\mathbf{R}^n$ is supplied with the standard norm. The family of all $\mathbf{R}$-linear and $\mathcal{A}_r$-additive functionals on $\mathcal{B}$ is denoted by $\mathcal{B}'$.

In particular we can take $X = \mathcal{A}_r^\alpha$, $Y = \mathcal{A}_r^\beta$ with $1 \le \alpha, \beta \in \mathbf{Z}$. Furthermore, analogous spaces $\mathcal{D}(U, Y)$, $[\mathcal{D}(U, Y)]'$, $\mathcal{B}(U, Y)$ and $[\mathcal{B}(U, Y)]'$ are defined for domains U in $\mathbf{R}^n$. For definiteness we write $\mathcal{B}(U, Y) = \{f|_U : \ f \in \mathcal{B}(\mathbf{R}^n, Y)\}$ and $\mathcal{D}(U, Y) = \{f|_U : \ f \in \mathcal{D}(\mathbf{R}^n, Y)\}$.

It is said, that a function $g : U \to \mathcal{A}_v$ is locally integrable, if it is absolutely integrable on each bounded λ measurable sub-domain V in U, i.e.

$\int_V |g(z)| \lambda(dz) < \infty$, where λ denotes the Lebesgue measure on U induced by that of on its real shadow.

A generalized function f is called regular if locally integrable functions ${}_{j,k}f^1, {}_lf^2 : U \to \mathcal{A}_v$ exist such that

$$[f,\omega) = \int_U \sum_{j,k,l} \left\{ {}_{j,k}f^1(z)_k\omega(z)_lf^2(z) \right\}_{q(3)} \lambda(dz)$$

for each test function either $\omega \in \mathcal{B}(U,Y)$ or $\omega \in \mathcal{D}(U,Y)$ correspondingly, where $\omega = ({}_1\omega, \dots, {}_\beta\omega)$, ${}_k\omega(z) \in \mathcal{A}_v$ for each $z \in U$ and all k, $q(3)$ is a vector indicating on an order of the multiplication in the curled brackets and it may depend on the indices $j,l = 1,\dots,\alpha$, $k = 1,\dots,\beta$.

We supply the space $\mathcal{B}(\mathbf{R}^n, Y)$ with the countable family of semi-norms

$$p_{\alpha,k}(f) := \sup_{x\in\mathbf{R}^n} \left|(1+|x|)^k \partial^\alpha f(x)\right| \tag{17}$$

inducing its topology, where $k = 0,1,2,\dots$; $\alpha = (\alpha_1,\dots,\alpha_n)$, $0 \le \alpha_j \in \mathbf{Z}$. On this space we take the space $\mathcal{B}'(\mathbf{R}^n, Y)_l$ of all Y valued continuous generalized functions (functionals) of the form

$$f = f_0 i_0 + \ldots + f_{2^v-1} i_{2^v-1} \quad \text{and} \quad g = g_0 i_0 + \ldots + g_{2^v-1} i_{2^v-1}\,, \tag{18}$$

where f_j and $g_j \in \mathcal{B}'(\mathbf{R}^n, Y)$, with restrictions on $\mathcal{B}(\mathbf{R}^n, \mathbf{R})$ being real- or $\mathbf{C_i} = \mathbf{R} \oplus \mathbf{iR}$- valued generalized functions $f_0,\dots,f_{2^v-1}$, $g_0,\dots,g_{2^v-1}$ respectively. Let $\phi = \phi_0 i_0 + \ldots + \phi_{2^v-1} i_{2^v-1}$ with $\phi_0,\dots,\phi_{2^v-1} \in \mathcal{B}(\mathbf{R}^n, \mathbf{R})$, then

$$[f,\phi) = \sum_{k,j=0}^{2^v-1} [f_j,\phi_k) i_k i_j\,. \tag{19}$$

Let their convolution be defined in accordance with the formula:

$$[f*g,\phi) = \sum_{j,k=0}^{2^v-1} \big([f_j * g_k, \phi) i_j\big) i_k \tag{20}$$

for each $\phi \in \mathcal{B}(\mathbf{R}^n, Y)$. Particularly,

$$(f*g)(x) = f(x-y)*g(y) = f(y)*g(x-y) \tag{21}$$

for all $x, y \in \mathbf{R}^n$ due to formula (20), since the latter equality is satisfied for each pair f_j and g_k (see also [3]).

2.3. The decomposition theorem of PDO over the Cayley-Dickson algebras

We consider a partial differential operator of order u:

$$Af(x) = \sum_{|\alpha|\le u} \mathbf{a}_\alpha(x) \partial^\alpha f(x)\,, \tag{22}$$

where $\partial^\alpha f = \partial^{|\alpha|} f(x)/\partial x_0^{\alpha_0} \dots \partial x_n^{\alpha_n}$, $x = x_0 i_0 + \dots x_n i_n$, $x_j \in \mathbf{R}$ for each j, $1 \le n = 2^r - 1$, $\alpha = (\alpha_0, \dots, \alpha_n)$, $|\alpha| = \alpha_0 + \dots + \alpha_n$, $0 \le \alpha_j \in \mathbf{Z}$. By the definition this means that the principal symbol

$$A_0 := \sum_{|\alpha|=u} \mathbf{a}_\alpha(x) \partial^\alpha \tag{23}$$

has α so that $|\alpha| = u$ and $\mathbf{a}_\alpha(x) \in \mathcal{A}_r$ is not identically zero on a domain U in $\mathcal{A}_r$. As usually $C^k(U, \mathcal{A}_r)$ denotes the space of k times continuously differentiable functions by all real variables $x_0, \dots, x_n$ on U with values in $\mathcal{A}_r$, while the x-differentiability corresponds to the super-differentiability by the Cayley-Dickson variable x.

Speaking about locally constant or locally differentiable coefficients we shall undermine that a domain U is the union of sub-domains U^j satisfying conditions 2.2$(D1, i{-}vii)$ [23] (or see Section 2.4 below) and $U^j \cap U^k = \partial U^j \cap \partial U^k$ for each $j \ne k$. All coefficients $\mathbf{a}_\alpha$ are either constant or differentiable of the same class on each $Int(U^j)$ with the continuous extensions on U^j. More generally it is up to a C^u or x-differentiable diffeomorphism of U respectively.

If an operator A is of the odd order $u = 2s - 1$, then an operator E of the even order $u + 1 = 2s$ by variables (t, x) exists so that

$$Eg(t,x)|_{t=0} = Ag(0,x) \tag{24}$$

for any $g \in C^{u+1}([c,d] \times U, \mathcal{A}_r)$, where $t \in [c,d] \subset \mathbf{R}$, $c \le 0 < d$, for example, $Eg(t,x) = \partial(tAg(t,x))/\partial t$.

Therefore, it remains the case of the operator A of the even order $u = 2s$. Take $z = z_0 i_0 + \dots + z_{2^v-1} i_{2^v-1} \in \mathcal{A}_v$, $z_j \in \mathbf{R}$. Moreover, operators depending on a less set $z_{l_1}, \dots, z_{l_n}$ of variables can be considered as restrictions of operators by all variables on spaces of functions constant by variables z_s with $s \notin \{l_1, \dots, l_n\}$.

Theorem 2.1. *Let $A = A_u$ be a partial differential operator of an even order $u = 2s$ with (locally) either constant or variable $C^{s'}$ or x-differentiable on U coefficients $\mathbf{a}_\alpha(x) \in \mathcal{A}_r$ such that it has the form*

$$Af = c_{u,1}(B_{u,1}f) + \dots + c_{u,k}(B_{u,k}f)\,, \tag{25}$$

where each

$$B_{u,p} = B_{u,p,0} + Q_{u-1,p} \tag{26}$$

is a partial differential operator of the order u by variables $x_{m_{u,1}+\dots+m_{u,p-1}+1}$, $\dots$, $x_{m_{u,1}+\dots+m_{u,p}}$, $m_{u,0} = 0$, $c_{u,k}(x) \in \mathcal{A}_r$ for each k, its principal part

$$B_{u,p,0} = \sum_{|\alpha|=s} \mathbf{a}_{p,2\alpha}(x) \partial^{2\alpha} \tag{27}$$

is elliptic with real coefficients $\mathbf{a}_{p,2\alpha}(x) \geq 0$*, either* $0 \leq r \leq 3$ *and* $f \in C^u(U, \mathcal{A}_r)$*, or* $r \geq 4$ *and* $f \in C^u(U, \mathbf{R})$*. Then three partial differential operators* Υ^s *and* Υ^s_1 *and* Q *of orders* s *and* p *with* $p \leq u-1$ *with (locally) either constant or variable of the class* $C^{s'}$ *or* x*-differentiable correspondingly on* U *coefficients with values in* $\mathcal{A}_v$ *exist and coefficients of the third operator* Q *may be generalized functions, when coefficients of* A *are locally either constant or of the class* $C^{s'}$ *or* x*-differentiable and discontinuous on the entire domain* U *or when* $s' < s$*,* $r \leq v$*, such that*

$$Af = \Upsilon^s(\Upsilon^s_1 f) + Qf\,. \tag{28}$$

Proof. Certainly, we have $ordQ_{u-1,p} \leq u-1$, $ord(A - A_0) \leq u-1$. We choose the following operators:

$$\Upsilon^s f(x) = \sum_{p=1}^{k} \sum_{\substack{|\alpha| \leq s;\ \alpha_q = 0 \forall q < (m_{u,1}+\ldots+m_{u,p-1}+1) \\ \text{and } \alpha_q = 0 \forall q > (m_{u,1}+\ldots+m_{u,p})}} (\partial^\alpha f(x))[w_p^* \psi_{p,\alpha}] \tag{29}$$

and

$$\Upsilon^s_1 f(x) = \sum_{p=1}^{k} \sum_{\substack{|\alpha| \leq s;\ \alpha_q = 0 \forall q < (m_{u,1}+\ldots+m_{u,p-1}+1) \\ \text{and } \alpha_q = 0 \forall q > (m_{u,1}+\ldots+m_{u,p})}} (\partial^\alpha f(x))[w_p \psi^*_{p,\alpha}]\,, \tag{30}$$

where $w_p^2 = c_{u,p}$ for all p and $\psi^2_{p,\alpha}(x) = -\mathbf{a}_{p,2\alpha}(x)$ for each p and x, $w_p \in \mathcal{A}_r$, $\psi_{p,\alpha}(x) \in \mathcal{A}_{r,v}$ and $\psi_{p,\alpha}(x)$ is purely imaginary for $\mathbf{a}_{p,2\alpha}(x) > 0$ for all α and x, $Re(w_p Im(\psi_{p,\alpha})) = 0$ for all p and α, $Im(x) = (x - x^*)/2$, $v > r$. There $\mathcal{A}_{r,v} = \mathcal{A}_v/\mathcal{A}_r$ is the real quotient algebra. The algebra $\mathcal{A}_{r,v}$ is considered with the generators i_{j2^r}, $j = 0, \ldots, 2^{v-r}-1$. Then a natural number v satisfying the condition:

$$2^{v-r} - 1 \geq \sum_{p=1}^{k} \sum_{q=0}^{u} \binom{m_p + q - 1}{q}$$

is sufficient, since as it is known the number of different solutions of the equation $\alpha_1 + \ldots + \alpha_m = q$ in non-negative integers α_j is $\binom{m+q-1}{q}$, where

$$\binom{m}{q} = \frac{m!}{q!(m-q)!} \text{ denotes the binomial coefficient.}$$

We have either $\partial^{\alpha+\beta} f \in \mathcal{A}_r$ for $0 \leq r \leq 3$ or $\partial^{\alpha+\beta} f \in \mathbf{R}$ for $r \geq 4$. Therefore, we can take $\psi_{p,\alpha}(x) \in i_{2^r q}\mathbf{R}$, where $q = q(p,\alpha) \geq 1$, $\quad q(p^1, \alpha^1) \neq q(p,\alpha)$ when $(p,\alpha) \neq (p^1, \alpha^1)$.

Thus decomposition (28) is valid due to the following. For $b = \partial^{\alpha+\beta} f(z)$ and $\mathbf{l} = i_{2^r p}$ and $w \in \mathcal{A}_r$ one has the identities:

$$(b(w\mathbf{l}))(w^*\mathbf{l}) = ((wb)\mathbf{l})(w^*\mathbf{l}) = -w(wb) = -w^2 b \tag{31}$$

and

$$(((b\mathbf{l})w^*)\mathbf{l})w = (((bw)\mathbf{l})\mathbf{l})w = -(bw)w = -bw^2 \tag{32}$$

in the considered in this section cases, since $\mathcal{A}_r$ is alternative for the parameter $r \leq 3$, while $\mathbf{R}$ is the center of the Cayley-Dickson algebra (see formulas 2.2(13, 14) [21]).

This decomposition of the operator A_{2s} is generally up to a partial differential operator of order not greater, than $(2s-1)$:

$$Qf(x) = \sum_{p=1}^{k} c_{u,p} Q_{u-1,p}$$
$$+ \sum_{\substack{|\alpha|\leq s, |\beta|\leq s; \gamma\leq\alpha, \\ \epsilon\leq\beta, |\gamma+\epsilon|>0}} \left[\prod_{j=0}^{2^v-1} \binom{\alpha_j}{\gamma_j}\binom{\beta_j}{\epsilon_j}\right] (\partial^{\alpha+\beta-\gamma-\epsilon} f(x))[(\partial^\gamma \eta_\alpha(x))((\partial^\epsilon \eta^1_\beta(x)], \tag{33}$$

where operators Υ^s and Υ^s_1 are already written in accordance with the general form

$$\Upsilon^s f(x) = \sum_{|\alpha|\leq s} (\partial^\alpha f(x))\eta_\alpha(x)\,; \tag{34}$$

$$\Upsilon^s_1 f(x) = \sum_{|\beta|\leq s} (\partial^\beta f(x))\eta^1_\beta(x)\,. \tag{35}$$

The coefficients of the operator Q may be generalized functions, since they are calculated with the participation of partial derivatives of the coefficients of the operator Υ^s_1, but the coefficients of the operators Υ^s and Υ^s_1 may be locally either constant or of class $C^{s'}$ or x-differentiable and discontinuous on the entire U or $s' < s$ when for the initial operator A they are such.

When the operator A in formula (24) is with constant coefficients, then the coefficients w_p and $\psi_{p,\alpha}$ for Υ^m and Υ^m_1 can also be chosen constant and hence $Q - \sum_{p=1}^k c_{u,p} Q_{u-1,p} = 0$. □

Corollary 2.2. *Let suppositions of Theorem 2.1 be satisfied. Then a change of variables (locally) either affine or variable C^1 or x-differentiable on U correspondingly on U exists so that the principal part $A_{2,0}$ of A_2 becomes with constant coefficients, when $\mathbf{a}_{p,2\alpha} > 0$ for each p, α and x.*

Corollary 2.3. *If two operators $E = A_{2s}$ and $A = A_{2s-1}$ are related by equation* (24)*, and A_{2s} is presented in accordance with formulas* (25) *and* (26)*, then three operators Υ^s, Υ^{s-1} and Q of orders s, $s-1$ and $p \leq 2s-2$ exist so that*

$$A_{2s-1} = \Upsilon^s \Upsilon^{s-1} + Q\,. \tag{36}$$

Proof. It remains to verify the inequality $ord(Q) \leq 2s-2$ in the case of A_{2s-1}, where $Q = \{\partial(tA_{2s-1})/\partial t - \Upsilon^s \Upsilon_1^s\}|_{t=0}$. Indeed, the form $\lambda(E)$ corresponding to E is of degree $2s-1$ by x and each addendum of degree $2s$ in it is of degree not less than 1 by t, consequently, the product of forms $\lambda(\Upsilon_s)\lambda(\Upsilon_1^s)$ corresponding to Υ^s and Υ_1^s is also of degree $2s-1$ by x and each addendum of degree $2s$ in it is of degree not less than 1 by t. But the principal parts of $\lambda(E)$ and $\lambda(\Upsilon_s)\lambda(\Upsilon_1^s)$ coincide identically by variables (t, x), hence the order satisfies the inequality $ord(\{E - \Upsilon^s \Upsilon_1^s\}|_{t=0}) \leq 2s-2$. Let $a(t, x)$ and $h(t, x)$ be coefficients from Υ_1^s and Υ^s. Using the identities

$$a(t,x)\partial_t \partial^\gamma t g(x) = a(t,x)\partial^\gamma g(x)$$

and

$$h(t,x)\partial^\beta \partial_t [a(t,x)\partial^\gamma g(x)] = h(t,x)\partial^\beta [(\partial_t a(t,x))\partial^\gamma g(x)]$$

for any functions $g(x) \in C^{2s-1}$ and $a(t,x) \in C^s$,

$$ord[(h(t,x)\partial^\beta), (a(t,x)\partial^\gamma)]|_{t=0} \leq 2s-2$$

where $\partial_t = \partial/\partial t$, $|\beta| \leq s-1$, $|\gamma| \leq s$, $[A, B] := AB - BA$ denotes the commutator of two operators, we reduce the term $(\Upsilon^s \Upsilon_1^s + Q_1)|_{t=0}$ from formula (28)to the form prescribes by equation (36). □

Remark 2.4. *We consider operators of the form:*

$$(\Upsilon^k + \beta I_r) f(z) := \left\{ \sum_{0<|\alpha|\leq k} (\partial^\alpha f(z))\eta_\alpha(z) \right\} + f(z)\beta(z)\,,$$

with $\eta_\alpha(z) \in \mathcal{A}_v$, $\alpha = (\alpha_0, \dots, \alpha_{2^r-1})$, $0 \leq \alpha_j \in \mathbf{Z}$ *for each* j, $|\alpha| = \alpha_0 + \dots + \alpha_{2^r-1}$, $\beta I_r f(z) := f(z)\beta$, $\partial^\alpha f(z) := \partial^{|\alpha|} f(z)/\partial z_0^{\alpha_0} \dots \partial z_{2^r-1}^{\alpha_{2^r-1}}$, $2 \leq r \leq v < \infty$, $\beta(z) \in \mathcal{A}_v$, $z_0, \dots, z_{2^r-1} \in \mathbf{R}$, $z = z_0 i_0 + \dots + z_{2^r-1} i_{2^r-1}$.

Proposition 2.5. *The operator* $(\Upsilon^k + \beta)^*(\Upsilon^k + \beta)$ *is elliptic on the space* $C^{2k}(\mathbf{R}^{2^r}, \mathcal{A}_v)$, *where* $(\Upsilon^k + \beta)^*$ *denotes the adjoint operator (i.e. with adjoint coefficients).*

Proof. In view of formulas (1) and (29) the form corresponding to the principal symbol of the operator $(\Upsilon^k + \beta)^*(\Upsilon^k + \beta)$ is with real coefficients, of degree $2k$ and non-negative definite, consequently, the operator $(\Upsilon^k + \beta)^*(\Upsilon^k + \beta)$ is elliptic. □

Example 2.6. *Let* Υ^* *be the adjoint operator defined on differentiable* $\mathcal{A}_v$ *valued functions* f *given by the formula:*

$$(\Upsilon + \beta)^* f = \left[\sum_{j=0}^{n} (\partial f(z)/\partial z_j)\phi_j(z) \right] + f(z)\beta(z)^*\,. \tag{37}$$

Thus we can consider the operator

$$\Xi_\beta := (\Upsilon + \beta)(\Upsilon + \beta)^* \,. \tag{38}$$

From Proposition 2.5 we have that the operator Ξ_β is elliptic as classified by its principal symbol with real coefficients. Put $\Xi = \Xi_0$. In the x coordinates from Section 2.1 it has the simpler form:

$$\begin{aligned}(\Upsilon + \beta)(\Upsilon + \beta)^* f = \ & \sum_{j=0}^{n} (\partial^2 f/\partial x_j^2)|t_j|^2 \\ & +2 \sum_{0 \le j < k \le n} (\partial^2 f/\partial x_j \partial x_k) Re(t_j t_k^*) \\ & +2 \sum_{j=0}^{n} (\partial f/\partial x_j) Re(t_j^* \beta) + \{f|\beta|^2 \\ & + \sum_{j=0}^{n} [f(\partial \beta^*/\partial x_j)] t_j\} \,,\end{aligned} \tag{39}$$

because the coefficients t_j are already constant. After a change of variables reducing the corresponding quadratic form to the sum of squares $\sum_j \epsilon_j s_j^2$ we get the formula:

$$\Upsilon \Upsilon^* f = \sum_{j=1}^{m} (\partial^2 f/\partial s_j^2) \epsilon_j \,, \tag{40}$$

where $s_j \in \mathbf{R}$, $\epsilon_j = 1$ for $1 \le j \le p$ and $\epsilon_j = -1$ for each $p < j \le m$, $m \le 2^v$, $1 \le p \le m$ depending on the signature $(p, m-p)$.

Generally (see Formula (36)*) we have*

$$A = (\Upsilon + \beta)(\Upsilon_1 + \beta^1) f(z) = B_0 f(z) + Q f(z) \,, \tag{41}$$

where the decomposition PDOs are given by the formulas:

$$B_0 f(z) = \sum_{j,k} \left[(\partial^2 f(z)/\partial z_j \partial z_k) \phi_j^1(z)^*\right] \phi_k^*(z) + \left[f(z) \beta^1(z)\right] \beta(z) \,, \tag{42}$$

$$\begin{aligned}Q f(z) \ & = \sum_{j,k} \left[(\partial f(z)/\partial z_j)(\partial \phi_j^1(z)^*/\partial z_k)\right] \phi_k^*(z) \\ & + \sum_j \left[(\partial f(z)/\partial z_j) \phi_j^1(z)^*\right] \beta(z) \\ & + \sum_k \left[f(z)(\partial \beta^1(z)/\partial z_k)\right] \phi_k^*(z)\end{aligned} \tag{43}$$

and

$$(\Upsilon_1 + \beta^1) f(z) = \left[\sum_j (\partial f(z)/\partial z_j) \phi_j^1(z)^*\right] + f(z) \beta^1(z) \,. \tag{44}$$

The latter equations show that coefficients of the operator Q may be generalized functions, when $\phi^1_j(z)$ for some j or $\beta^1(z)$ are locally C^0 or locally C^1 functions, while $\phi_k(z)$ for each k and $\beta(z)$ are locally C^0 functions on U. We consider this in more details in the next section.

2.4. Partial differential operators with generalized coefficients

Let an operator Q be given by the formula:

$$[Af, \omega^{\otimes(u+1)}) = [\Upsilon^s(\Upsilon^s_1 f) + Qf, \omega^{\otimes(u+1)}) \tag{45}$$

for each real-valued test function ω on a domain U. Initially it is considered on a domain in the Cayley-Dickson algebra $\mathcal{A}_v$. But in the case when Q and f depend on smaller number of real coordinates $z_0, \dots, z_{n-1}$ we can take the real shadow of U and its sub-domain V of variables $(z_0, \dots, z_{n-1})$, where z_k are marked for example being zero for all $n \le k \le 2^v - 1$. Thus we take a domain V which is a canonical closed subset in the Euclidean space $\mathbf{R}^n$, where $2^{v-1} \le n \le 2^v - 1$, $v \ge 2$.

A canonical closed subset P of the Euclidean space $X = \mathbf{R}^n$ is called a quadrant if it can be given by the condition $P := \{x \in X : q_j(x) \ge 0\}$, where $(q_j : j \in \Lambda_P)$ are linearly independent elements of the topologically adjoint space X^*. Here $\Lambda_P \subset \mathbf{N}$ (with $card(\Lambda_P) = k \le n$) and k is called the index of P. If $x \in P$ and exactly j of functionals q_i's satisfy the equality $q_i(x) = 0$ then x is called a corner of index j. That is a quadrant P is affine diffeomorphic with the domain $P^n = \prod_{j=1}^n [\mathsf{a}_j, b_j]$, where $-\infty \le \mathsf{a}_j < b_j \le \infty$, $[\mathsf{a}_j, b_j] := \{x \in \mathbf{R} : \ \mathsf{a}_j \le x \le b_j\}$ denotes the segment in $\mathbf{R}$. This means that there exists a vector $p \in \mathbf{R}^n$ and a linear invertible mapping C on $\mathbf{R}^n$ so that $C(P) - p = P^n$. We put $t^{j,1} := (t_1, \dots, t_j, \dots, t_n : \ t_j = \mathsf{a}_j)$, $t^{j,2} := (t_1, \dots, t_j, \dots, t_n : \ t_j = b_j)$. Consider $t = (t_1, \dots, t_n) \in P^n$.

Then a manifold M with corners is defined as follows. It is a metric separable space modelled on the Euclidean space $X = \mathbf{R}^n$ and it is supposed to be of class C^s, where $1 \le s$. Charts on the manifold M are denoted by (U_l, u_l, P_l), that is, $u_l : U_l \to u_l(U_l) \subset P_l$ is a C^s-diffeomorphism for each l, where a subset U_l is open in M, the composition $u_l \circ {u_j}^{-1}$ is of C^s class of smoothness from the domain $u_j(U_l \cap U_j) \ne \emptyset$ onto $u_l(U_l \cap U_j)$, that is, $u_j \circ u_l^{-1}$ and $u_l \circ u_j^{-1}$ are bijective, $\bigcup_j U_j = M$.

A point $x \in M$ is called a corner of index j if there exists a chart (U, u, P) of M with $x \in U$ and $u(x)$ is of index $ind_M(x) = j$ in $u(U) \subset P$. A set of all corners of index $j \ge 1$ is called a border ∂M of M, x is called an inner point of M if $ind_M(x) = 0$, so $\partial M = \bigcup_{j \ge 1} \partial^j M$, where $\partial^j M := \{x \in M : ind_M(x) = j\}$ (see also [26]). We consider that

(D1) V is a canonical closed subset in the Euclidean space $\mathbf{R}^n$, that is $V = cl(Int(V))$, where $Int(V)$ denotes the interior of V and $cl(V)$ denotes the closure of V.

Particularly, the entire Euclidean space $\mathbf{R}^n$ may also be taken. Let a manifold W be satisfying the following conditions $(i - v)$.

(i) The manifold W is continuous and piecewise C^α, where C^l denotes the family of l times continuously differentiable functions. This means by the definition that W as the manifold is of class $C^0 \cap C^\alpha_{loc}$. That is W is of class C^α on open subsets $W_{0,j}$ in W and $W \setminus (\bigcup_j W_{0,j})$ has a codimension not less than one in W.

(ii) $W = \bigcup_{j=0}^m W_j$, where $W_0 = \bigcup_k W_{0,k}$, $W_j \cap W_k = \emptyset$ for each $k \neq j$, $m = dim_{\mathbf{R}} W$, $dim_{\mathbf{R}} W_j = m - j$, $W_{j+1} \subset \partial W_j$.

(iii) Each W_j with $j = 0, \dots, m-1$ is an oriented C^α-manifold, W_j is open in $\bigcup_{k=j}^m W_k$. An orientation of W_{j+1} is consistent with that of ∂W_j for each $j = 0, 1, \dots, m-2$. For $j > 0$ the set W_j is allowed to be void or non-void.

(iv) A sequence W^k of C^α orientable manifolds embedded into the Euclidean space $\mathbf{R}^n$, with $\alpha \geq 1$, exists such that W^k uniformly converges to W on each compact subset in $\mathbf{R}^n$ relative to the metric $dist$. For two subsets B and E in a metric space X with a metric ρ we put

$$dist(B, E) := \max \left\{ \sup_{b \in B} dist(\{b\}, E), \sup_{e \in E} dist(B, \{e\}) \right\} \tag{46}$$

where $dist(\{b\}, E) := \inf_{e \in E} \rho(b, e)$, $dist(B, \{e\}) := \inf_{b \in B} \rho(b, e)$, $b \in B$, $e \in E$. Generally, $dim_{\mathbf{R}} W = m \leq n$. Let $(e_1^k(x), \dots, e_m^k(x))$ be a basis in the tangent space $T_x W^k$ at $x \in W^k$ consistent with the orientation of W^k, $k \in \mathbf{N}$. We suppose that the sequence of orientation frames $(e_1^k(x_k), \dots, e_m^k(x_k))$ of W^k at x_k converges to $(e_1(x), \dots, e_m(x))$ for each $x \in W_0$, where $\lim_k x_k = x \in W_0$, while $e_1(x),\dots,e_m(x)$ are linearly independent vectors in $\mathbf{R}^n$.

(v) Let a sequence of Riemann volume elements λ_k on W^k (see §XIII.2 [30]) induce a limit volume element λ on W, that is, $\lambda(B \cap W) = \lim_{k\to\infty}(B \cap W^k)$ for each compact canonical closed subset B in $\mathbf{R}^n$, consequently, $\lambda(W \setminus W_0) = 0$.

(vi) We consider surface integrals of the second kind, i.e. by the oriented surface W (see (iv)), where W_j is oriented for each $j = 0, \dots, m-1$ (see also §XIII.2.5 [30]).

Suppose that a boundary ∂U of U satisfies Conditions $(i-v)$ and

(vii) let the orientations of ∂U^k and U^k be consistent for each $k \in \mathbf{N}$ (see Proposition 2 and Definition 3 §XIII.2.5 [30]).

Particularly, the Riemann volume element λ_k on ∂U^k is consistent with the Lebesgue measure on U^k induced from $\mathbf{R}^n$ for each k. These conditions provide the measure λ on ∂U as in (v).

The consideration of this section certainly encompasses the case of a domain U with a C^α boundary ∂U as well.

Suppose that $U_1,\dots,U_l$ are domains in the Euclidean space $\mathbf{R}^n$ satisfying conditions $(D1, i-vii)$ and such that $U_j \cap U_k = \partial U_j \cap \partial U_k$ for each $j \ne k$, $U = \bigcup_{j=1}^l U_j$. Consider a function $g : U \to \mathcal{A}_v$ such that each its restriction $g|_{U_j}$ is of class C^s, but g on the entire domain U may be discontinuous or not C^k, where $0 \le k \le s$, $1 \le s$. If $x \in \partial U_j \cap \partial U_k$ for some $j \ne k$, $x \in Int(U)$, such that x is of index $m \ge 1$ in U_j (and in U_k also). Then there exists canonical C^α local coordinates $(y_1,\dots,y_n)$ in a neighborhood V_x of x in U such that $S := V_x \cap \partial^m U_j = \{y: \ y \in V_x;\ y_1 = 0,\dots,y_m = 0\}$. Using locally finite coverings of the locally compact topological space $\partial U_j \cap \partial U_k$ without loss of generality we suppose that C^α functions $P_1(z),\dots,P_m(z)$ on $\mathbf{R}^n$ exist with $S = \{z: \ z \in \mathbf{R}^n;\ P_1(z) = 0,\dots,P_m(z) = 0\}$. Therefore, on the surface S the delta-function $\delta(P_1,\dots,P_m)$ exists, for $m = 1$ denoting them $P = P_1$ and $\delta(P)$ respectively (see §III.1 [3]).

It is possible to choose $y_j = P_j$ for $j = 1,\dots,m$. Using generalized functions with definite supports, for example compact supports, there is possible without loss of generality to consider that $y_1,\dots,y_n \in \mathbf{R}$ are real variables. Let $\theta(P_j)$ be the characteristic function of the domain $\mathcal{P}_j := \{z : P_j(z) \ge 0\}$, $\theta(P_j) := 1$ for $P_j \ge 0$ and $\theta(P_j) = 0$ for $P_j < 0$. Then the generalized function $\theta(P_1,\dots,P_m) := \theta(P_1)\dots\theta(P_m)$ can be considered as the direct product of generalized functions $\theta(y_1),\dots,\theta(y_m)$, $1(y_{m+1},\dots,y_n) \equiv 1$, since variables $y_1,\dots,y_n$ are independent. Then in the class of generalized functions the following formulas are valid:

$$\partial\theta(P_j)/\partial z_k = \delta(P_j)(\partial P_j/\partial z_k) \tag{47}$$

for each $k = 1,\dots,n$, consequently,

$$grad[\theta(P_1,\dots,P_m)] = \\ = \sum_{j=1}^m \left[\theta(P_1)\dots\theta(P_{j-1})\delta(P_j)(grad(P_j))\theta(P_{j+1})\dots\theta(P_m)\right], \tag{48}$$

where $grad\ g(z) := (\partial g(z)/\partial z_1,\dots,\partial g(z)/\partial z_n)$ (see Formulas III.1.3$(1,7,7',9)$ and III.1.9(6) [3]).

Let for the domain U in the Euclidean space $\mathbf{R^n}$ the set of internal surfaces $cl_U[Int_{\mathbf{R}^n}(U)\cap\bigcup_{j\ne k}(\partial U_j\cap\partial U_k)]$ in U on which a function $g:U\to\mathcal{A}_v$ or its derivatives may be discontinuous is presented as the disjoint union of surfaces Γ_j, where each surface Γ^j is the boundary of the sub-domain

$$\mathcal{P}^j := \{P_{j,1}(z)\ge 0,\dots,P_{j,m_j}(z)\ge 0\}\,,\quad \Gamma^j=\partial\mathcal{P}^j=\bigcup_{k=1}^{m_j}\partial^k\mathcal{P}^j\,, \tag{49}$$

$m_j\in\mathbf{N}$ for each j, $cl_X(V)$ denotes the closure of a subset V in a topological space X, $Int_X(V)$ denotes the interior of V in X. By its construction $\{\mathcal{P}^j:\ j\}$ is the covering of U which is the refinement of the covering $\{U_k:\ k\}$, i.e. for each $\mathcal{P}^j$ a number k exists so that $\mathcal{P}^j\subset U_k$ and $\partial\mathcal{P}^j\subset\partial U_k$ and $\bigcup_j\mathcal{P}^j=\bigcup_k U_k=U$. We put

$$\begin{aligned} h_j(z(x)) = &\quad h(x)|_{x\in\Gamma^j}\\ := &\lim_{y_{j,1}\downarrow 0,\dots,y_{j,n}\downarrow 0} g(z(x+y)) - \lim_{y_{j,1}\downarrow 0,\dots,y_{j,n}\downarrow 0} g(z(x-y)) \end{aligned} \tag{50}$$

in accordance with the supposition made above that g can have only discontinuous of the first kind, i.e. the latter two limits exist on each Γ^j, where x and y are written in coordinates in $\mathcal{P}^j$, $z=z(x)$ denotes the same point in the global coordinates z of the Euclidean space $\mathbf{R}^n$. Then we take a new continuous function

$$g^1(z)=g(z)-\sum_j h_j(z)\theta(P_{j,1}(z),\dots,P_{j,m_j}(z)). \tag{51}$$

Let the partial derivatives and the gradient of the function g^1 be calculated piecewise one each U_k. Since g^1 is the continuous function, it is the regular generalized function by the definition, consequently, its partial derivatives exist as the generalized functions. If $g^1|_{U_j}\in C^1(U_j,\mathcal{A}_v)$, then $\partial g^1(z)/\partial z_k$ is the continuous function on U_j. The latter means that in such case $\partial g^1(z)\chi_{U_j}(z)/\partial z_k$ is the regular generalized function on U_j for each k, where $\chi_G(z)$ denotes the characteristic function of a subset G in $\mathcal{A}_v$, $\chi_G(z)=1$ for each $z\in G$, while $\chi(z)=0$ for $z\in\mathcal{A}_v\setminus G$. Therefore, one gets:

$$g^1(z)=\sum_j g^1(z)\chi_{U_j\setminus\bigcup_{k<j}U_k}(z),$$

where $U_0=\emptyset$, $j,k\in\mathbf{N}$.

On the other hand, the function $g(z)$ is locally continuous on U, consequently, it defines the regular generalized function on the space $\mathcal{D}(U,\mathcal{A}_v)$ of test functions by the formula:

$$[g,\omega) := \int_U g(z)\omega(z)\lambda(dz)\,,$$

where λ is the Lebesgue measure on U induced by the Lebesgue measure on the real shadow $\mathbf{R}^{2^v}$ of the Cayley-Dickson algebra $\mathcal{A}_v$, $\omega \in \mathcal{D}(U, \mathcal{A}_v)$. Thus partial derivatives of g exist as generalized functions.

In accordance with formulas (47), (48) and (51) we infer that the gradient of the function $g(z)$ on the domain U is the following:

$$grad\ g(z) = grad\ g^1(z) + \sum_j h_j(z) grad\ \theta(P_{j,1}, \dots, P_{j,m_j})\,. \tag{52}$$

Thus formulas (48) and (52) permit calculations of coefficients of the partial differential operator Q given by formula (43).

2.5. Line generalized functions

Let U be a domain satisfying conditions $2.1(D1, D2)$ [21] and $(D1, i-vii)$. We embed the Euclidean space $\mathbf{R}^n$ into the Cayley-Dickson algebra $\mathcal{A}_v$, $2^{v-1} \le n \le 2^v - 1$, as the $\mathbf{R}$ affine sub-space putting $\mathbf{R}^n \ni x = (x_1, \dots, x_n) \mapsto x_1 i_{j_1} + \dots + x_n i_{j_n} + x^0 \in \mathcal{A}_v$, where $j_k \ne j_l$ for each $k \ne l$, x^0 is a marked Cayley-Dickson number, for example, $j_k = k$ for each k, $x^0 = 0$. Moreover, each z_j can be written in the z-representation in accordance with formulas $2.1(1-3)$ [21].

We denote by $\mathbf{P} = \mathbf{P}(U)$ the family of all rectifiable paths $\gamma : [a_\gamma, b_\gamma] \to U$ supplied with the metric

$$\rho(\gamma, \omega) := |\gamma(a) - \omega(a_\omega)| + \inf_\phi V_a^b(\gamma(t)) - \omega(\phi(t)) \tag{53}$$

where the infimum is taken by all diffeomorphisms $\phi : [a_\gamma, b_\gamma] \to [a_\omega, b_\omega]$ so that $\phi(a_\gamma) = a_\omega$, $a = a_\gamma < b_\gamma = b$.

Let us introduce a continuous mapping $g : \mathcal{B}(U, \mathcal{A}_v) \times \mathbf{P}(U) \times \mathcal{V}(U, \mathcal{A}_v) \to Y$ such that its values are denoted by $[g; \omega, \gamma; \nu]$, where Y is a module over the Cayley-Dickson algebra $\mathcal{A}_v$, $\omega \in \mathcal{B}(U, \mathcal{A}_v)$, $\gamma \in \mathbf{P}(U)$, while $\mathcal{V}(U, \mathcal{A}_v)$ denotes the family of all functions on U with values in the Cayley-Dickson algebra of bounded variation (see §2.3 [21]), $\nu \in \mathcal{V}(U, \mathcal{A}_v)$. For the identity mapping $\nu(z) = id(z) = z$ values of this functional will be denoted shortly by $[g; \omega, \gamma]$. Suppose that this mapping g satisfies the following properties $(G1 - G5)$:

(G1) $[g; \omega, \gamma; \nu]$ is bi- $\mathbf{R}$ homogeneous and $\mathcal{A}_v$ additive by a test function ω and by a function of bounded variation ν;

(G2) $[g; \omega, \gamma; \nu] = [g; \omega, \gamma^1; \nu] + [g; \omega, \gamma^2; \nu]$ for each γ, γ^1 and $\gamma^2 \in \mathbf{P}(U)$ such that $\gamma(t) = \gamma^1(t)$ for all $t \in [a_{\gamma^1}, b_{\gamma^1}]$ and $\gamma(t) = \gamma^2(t)$ for any $t \in [a_{\gamma^2}, b_{\gamma^2}]$ and $a_{\gamma^1} = a_\gamma$ and $a_{\gamma^2} = b_{\gamma^1}$ and $b_\gamma = b_{\gamma^2}$.

(G3) If a rectifiable curve γ does not intersect a support of a test function ω or a function of bounded variation ν, $\gamma([a, b]) \cap (supp(\omega) \cap supp(\nu)) = \emptyset$, then $[g; \omega, \gamma; \nu] = 0$, where $supp(\omega) := cl\{z \in U : \ \omega(z) \ne 0\}$.

Further we put

(G4) $[\partial^{|m|}g(z)/\partial z_0^{m_0}\dots\partial z_{2^v-1}^{m_{2^v-1}};\omega,\gamma]:=(-1)^{|m|}[g;\partial^{|m|}\omega(z)/\partial z_0^{m_0}\dots\partial z_{2^v-1}^{m_{2^v-1}},\gamma]$ for each $m=(m_0,\dots,m_{2^v-1})$, m_j is a non-negative integer $0\le m_j\in\mathbf{Z}$ for each j, $|m|:=m_0+\dots+m_{2^v-1}$. In the case of a super-differentiable function ω and a generalized function g we also put

(G5) $[(d^kg(z)/dz^k).(h_1,\dots,h_k);\omega,\gamma]:=(-1)^k[g;(d^k\omega(z)/dz^k).(h_1,\dots,h_k),\gamma]$ for any natural number $k\in\mathbf{N}$ and Cayley-Dickson numbers $h_1,..,h_k\in\mathcal{A}_v$.

Then g is called the Y valued line generalized function on $\mathcal{B}(U,\mathcal{A}_v)\times\mathbf{P}(U)\times\mathcal{V}(U,\mathcal{A}_v)$. Analogously it can be defined on $\mathcal{D}(U,\mathcal{A}_v)\times\mathbf{P}(U)\times\mathcal{V}(U,\mathcal{A}_v)$. In the case $Y=\mathcal{A}_v$ we call it simply the line generalized function, while for $Y=L_q(\mathcal{A}_v^k,\mathcal{A}_v^l)$ we call it the line generalized operator valued function, $k,l\ge 1$, omitting "on $\mathcal{B}(U,\mathcal{A}_v)\times\mathbf{P}(U)\times\mathcal{V}(U,\mathcal{A}_v)$" or "line" for short, when it is specified. Their spaces we denote by $L_q(\mathcal{B}(U,\mathcal{A}_v)\times\mathbf{P}(U)\times\mathcal{V}(U,\mathcal{A}_v);Y)$. Thus if g is a generalized function, then formula $(G5)$ defines the operator valued generalized function $d^kg(z)/dz^k$ with $k\in\mathbf{N}$ and $l=1$.

If g is a continuous function on U, then the formula

$$[g;\omega,\gamma;\nu]=\int_\gamma g(y)\omega(y)d\nu(y) \tag{54}$$

defines the generalized function. If $\hat{f}(z)$ is a continuous $L_q(\mathcal{A}_v,\mathcal{A}_v)$ valued function on U, then it defines the generalized operator valued function with $Y=L_q(\mathcal{A}_v,\mathcal{A}_v)$ such that

$$[\hat{f};\omega,\gamma;\nu]=\int_\gamma\{\hat{f}(z).\omega(z)\}d\nu(z)\,. \tag{55}$$

Particularly, for $\nu=id$ the equality $d\nu(z)=dz$ is satisfied.

We consider on $L_q(\mathcal{B}(U,\mathcal{A}_v)\times\mathbf{P}(U)\times\mathcal{V}(U,\mathcal{A}_v);Y)$ the strong topology:

(G6) $\lim_l f^l=f$ means by the definition that for each marked test function $\omega\in\mathcal{B}(U,\mathcal{A}_v)$ and rectifiable path $\gamma\in\mathbf{P}(U)$ and function of bounded variation $\nu\in\mathcal{V}(U,\mathcal{A}_v)$ the limit relative to the norm in Y exists $\lim_l[f^l;\omega,\gamma;\nu]=[f;\omega,\gamma;\nu]$.

2.6. Poly-functionals

Let $\mathbf{a}_k:\mathcal{B}(U,\mathcal{A}_r)^k\to\mathcal{A}_r$ or $\mathbf{a}_k:\mathcal{D}(U,\mathcal{A}_r)^k\to\mathcal{A}_r$ be a continuous mapping satisfying the following three conditions:

(P1) $[\mathbf{a}_k, \omega^1 \otimes \ldots \otimes \omega^k)$ is $\mathbf{R}$-homogeneous $[\mathbf{a}_k, \omega^1 \otimes \ldots \otimes (\omega^l t) \otimes \ldots \otimes \omega^k) = [\mathbf{a}_k, \omega^1 \otimes \ldots \otimes \omega^l \otimes \ldots \otimes \omega^k)t = [\mathbf{a}_k t, \omega^1 \otimes \ldots \otimes \omega^k)$ for each $t \in \mathbf{R}$ and $\mathcal{A}_r$-additive $[\mathbf{a}_k, \omega^1 \otimes \ldots \otimes (\omega^l + \kappa^l) \otimes \ldots \otimes \omega^k) = [\mathbf{a}_k, \omega^1 \otimes \ldots \otimes \omega^l \otimes \ldots \otimes \omega^k) + [\mathbf{a}_k, \omega^1 \otimes \ldots \otimes \kappa^l \otimes \ldots \otimes \omega^k)$ by any $\mathcal{A}_r$ valued test functions ω^l and κ^l, when other functions are marked, $l = 1, \ldots, k$, i.e. it is k $\mathbf{R}$-linear and k $\mathcal{A}_r$-additive, where $[\mathbf{a}_k, \omega^1 \otimes \ldots \otimes \omega^k)$ denotes a value of $\mathbf{a}_k$ on given test $\mathcal{A}_r$ valued functions $\omega^1, \ldots, \omega^k$;

(P2) $[\mathbf{a}_k \alpha, \omega^1 \otimes \ldots \otimes (\omega^l \beta) \otimes \ldots \otimes \omega^k) = ([\mathbf{a}_k, \omega^1 \otimes \ldots \otimes \omega^l \otimes \ldots \otimes \omega^k)\alpha)\beta = [(\mathbf{a}_k \alpha)\beta, \omega^1 \otimes \ldots \otimes \omega^l \otimes \ldots \otimes \omega^k)$ for all real-valued test functions and $\alpha, \beta \in \mathcal{A}_r$;

(P3) $[\mathbf{a}_k, \omega^{\sigma(1)} \otimes \ldots \otimes \omega^{\sigma(k)}) = [\mathbf{a}_k, \omega^1 \otimes \ldots \otimes \omega^k)$ for all real-valued test functions and each transposition σ, i.e. bijective surjective mapping $\sigma : \{1, \ldots, k\} \to \{1, \ldots, k\}$.

Then $\mathbf{a}_k$ will be called the symmetric k $\mathbf{R}$-linear k $\mathcal{A}_r$-additive continuous functional, $1 \le k \in \mathbf{Z}$. The family of all such symmetric functionals is denoted by $\mathcal{B}'_{k,s}(U, \mathcal{A}_v)$ or $\mathcal{D}'_{k,s}(U, \mathcal{A}_r)$ correspondingly. A functional satisfying conditions $(P1, P2)$ is called a continuous k-functional over $\mathcal{A}_r$ and their family is denoted by $\mathcal{B}'_k(U, \mathcal{A}_r)$ or $\mathcal{D}'_k(U, \mathcal{A}_r)$ respectively. When a situation is outlined we may omit for short "continuous" or "k $\mathbf{R}$-linear k $\mathcal{A}_v$-additive".

The sum of two k-functionals over the Cayley-Dickson algebra $\mathcal{A}_r$ is prescribed by the equality:

$$[\mathbf{a}_k + \mathbf{b}_k, \omega^1 \otimes \ldots \otimes \omega^k) = [\mathbf{a}_k, \omega^1 \otimes \ldots \otimes \omega^k) + [\mathbf{b}_k, \omega^1 \otimes \ldots \otimes \omega^k) \tag{56}$$

for each test functions. Using formula (56) each k-functional can be written as

$$\begin{aligned}&[\mathbf{a}_k, \omega^1 \otimes \ldots \otimes \omega^k) = \\ &\quad = [\mathbf{a}_{k,0} i_0, \omega^1 \otimes \ldots \otimes \omega^k) + \ldots + [\mathbf{a}_{k,2^r-1} i_{2^r-1}, \omega^1 \otimes \ldots \otimes \omega^k)\,,\end{aligned} \tag{57}$$

where $[\mathbf{a}_{k,j}, \omega^1 \otimes \ldots \otimes \omega^k) \in \mathbf{R}$ is real for all real-valued test functions $\omega^1, \ldots, \omega^k$ and each j; $i_0, \ldots, i_{2^r-1}$ denote the standard generators of the Cayley-Dickson algebra $\mathcal{A}_r$.

The direct product $\mathbf{a}_k \otimes \mathbf{b}_p$ of two functionals $\mathbf{a}_k$ and $\mathbf{b}_p$ for the same space of test functions is a $k + p$-functional over $\mathcal{A}_r$ given by the following three conditions:

(P4) $[\mathbf{a}_k \otimes \mathbf{b}_p, \omega^1 \otimes \ldots \otimes \omega^{k+p}) = [\mathbf{a}_k, \omega^1 \otimes \ldots \otimes \omega^k)[\mathbf{b}_p, \omega^{k+1} \otimes \ldots \otimes \omega^{k+p})$ for any real-valued test functions $\omega^1, \ldots, \omega^{k+p}$;

(P5) if $[\mathbf{b}_p, \omega^{k+1} \otimes \ldots \otimes \omega^{k+p}) \in \mathbf{R}$ is real for any real-valued test functions, then $[(\mathbf{a}_k N_1) \otimes (\mathbf{b}_p N_2), \omega^1 \otimes \ldots \otimes \omega^{k+p}) = ([\mathbf{a}_k \otimes \mathbf{b}_p, \omega^1 \otimes \ldots \otimes \omega^{k+p})N_1)N_2$ for any real-valued test functions $\omega^1, \ldots, \omega^{k+p}$ and Cayley-Dickson numbers $N_1, N_2 \in \mathcal{A}_r$;

(P6) if $[\mathbf{a}_k, \omega^1 \otimes \ldots \otimes \omega^k) \in \mathbf{R}$ and $[\mathbf{b}_p, \omega^{k+1} \otimes \ldots \otimes \omega^{k+p}) \in \mathbf{R}$ are real for any real-valued test functions, then $[\mathbf{a}_k \otimes \mathbf{b}_p, \omega^1 \otimes \ldots \otimes (\omega^l N_1) \otimes \ldots \otimes \omega^{k+p}) = [\mathbf{a}_k \otimes \mathbf{b}_p, \omega^1 \otimes \ldots \otimes \omega^{k+p}) N_1$ for any real-valued test functions $\omega^1, \ldots, \omega^{k+p}$ and each Cayley-Dickson number $N_1 \in \mathcal{A}_r$ for each $l = 1, \ldots, k+p$.

Therefore, we can now consider a partial differential operator of order u acting on a generalized function $f \in \mathcal{B}'(U, \mathcal{A}_r)$ or $f \in \mathcal{D}'(U, \mathcal{A}_r)$ and with generalized coefficients either $\mathbf{a}_\alpha \in \mathcal{B}'_{|\alpha|}(U, \mathcal{A}_r)$ or all $\mathbf{a}_\alpha \in \mathcal{D}'_{|\alpha|}(U, \mathcal{A}_r)$ correspondingly:

$$Af(x) = \sum_{|\alpha| \le u} (\partial^\alpha f(x)) \otimes [(\mathbf{a}_\alpha(x)) \otimes 1^{\otimes(u-|\alpha|)}], \tag{58}$$

where $\partial^\alpha f = \partial^{|\alpha|} f(x) / \partial x_0^{\alpha_0} \ldots \partial x_n^{\alpha_n}$, $x = x_0 i_0 + \ldots x_n i_n$, $x_j \in \mathbf{R}$ for each j, $1 \le n = 2^r - 1$, $\alpha = (\alpha_0, \ldots, \alpha_n)$, $|\alpha| = \alpha_0 + \ldots + \alpha_n$, $0 \le \alpha_j \in \mathbf{Z}$, $[1, \omega) := \int_U \omega(y) \lambda(dy)$, λ denotes the Lebesgue measure on U, for convenience $1^{\otimes 0}$ means the multiplication on the unit $1 \in \mathbf{R}$. The partial differential equation $Af = g$ in terms of generalized functions has a solution f means by the definition that

$$[Af, \omega^{\otimes(u+1)}) = [g, \omega^{\otimes(u+1)}) \tag{59}$$

for each real-valued test function ω on U, where $\omega^{\otimes k} = \omega \otimes \ldots \otimes \omega$ denotes the k times direct product of a test functions ω.

Theorem 2.7. *Let $A = A_u$ be a partial differential operator with generalized over the Cayley-Dickson algebra $\mathcal{A}_r$ coefficients of an even order $u = 2s$ on U such that each $\mathbf{a}_\alpha$ is symmetric for $|\alpha| = u$ and A has the form*

$$Af = (B_{u,1} f) c_{u,1} + \ldots + (B_{u,k} f) c_{u,k}, \tag{60}$$

where each

$$B_{u,p} = B_{u,p,0} + Q_{u-1,p} \tag{61}$$

is a partial differential operator of the order u by variables $x_{m_{u,1}+\ldots+m_{u,p-1}+1}$, $\ldots$, $x_{m_{u,1}+\ldots+m_{u,p}}$, $m_{u,0} = 0$, $c_{u,k}(x) \in \mathcal{A}_r$ for each k, its principal part

$$B_{u,p,0} f = \sum_{|\alpha|=s} (\partial^{2\alpha} f) \otimes \mathbf{a}_{p,2\alpha}(x) \tag{62}$$

is elliptic, that is

$$\sum_{|\alpha|=s} y^{2\alpha} [\mathbf{a}_{p,2\alpha}, \omega^{\otimes 2s}) \ge 0$$

for all $y_{k(1)}, \ldots, y_{k(m_{u,p})}$ in $\mathbf{R}$ with $k(1) = m_{u,1} + \ldots + m_{u,p-1} + 1, \ldots, k(m_{u,p}) = m_{u,1} + \ldots + m_{u,p}$, $y^\beta := y_{k(1)}^{\beta_{k_1}} \ldots y_{k(m_{u,p})}^{\beta_{k(m_{u,p})}}$ and $[\mathbf{a}_{p,2\alpha}, \omega^{\otimes 2s}) \ge 0$ for each real *test function ω, either $0 \le r \le 3$ and f is with values in $\mathcal{A}_r$, or $r \ge 4$ and f is real-valued on real-valued test functions. Then three partial differential operators Υ^s and Υ^s_1 and Q of orders s and p with $p \le u-1$ with generalized on U coefficients with values in $\mathcal{A}_v$ exist such that*

$$[Af, \omega^{\otimes(u+1)}) = [\Upsilon^s(\Upsilon^s_1 f) + Qf, \omega^{\otimes(u+1)}) \tag{63}$$

for each real-valued test function ω on U.

Proof. If a_{2s} is a symmetric functional and $[\mathbf{c}_s, \omega^{\otimes s}) = [\mathbf{a}_{2s}, \omega^{\otimes 2s})^{1/2}$ for each real-valued test function ω, then by formulas $(P1, P2)$ this functional $\mathbf{c}_s$ has an extension up to a continuous s-functional over the Cayley-Dickson algebra $\mathcal{A}_r$. This is sufficient for Formula (63), where only real-valued test functions ω are taken.

Consider a continuous p-functional $\mathbf{c}_p$ over $\mathcal{A}_v$, $p \in \mathbf{N}$. Supply the domain U with the metric induced from either the corresponding Euclidean space or the Cayley-Dickson algebra in which U is embedded depending on the considered case. It is possible to take a sequence of non-negative test functions ${}_l\omega$ on U with a support $supp({}_l\omega)$ contained in the ball $B(U, z, 1/l)$ with center z and radius $1/l$ and ${}_l\omega$ positive on some open neighborhood of a marked point z in U so that

$$\int_U {}_l\omega(z)\lambda(dz) = 1$$

for each $l \in \mathbf{N}$. If the p-functional $\mathbf{c}_p$ is regular and realized by a continuous $\mathcal{A}_v$ valued function g on U^p, then the limit exists:

$$\lim_l [\mathbf{c}_p, {}_l\omega^{\otimes p}) = g(z, \dots, z).$$

Thus the partial differential equation (47) for regular functionals and their derivatives implies the classical partial differential equation (22).

The considered above spaces of real-valued test functions are dense in the corresponding spaces of real-valued generalized functions (see [3]). Moreover, there is the decomposition of each generalized function g into the sum of the form $g = \sum_j g_j i_j$ with real-valued generalized functions g_j, where $i_0, \dots, i_{2^v}$ are the standard basis generators of the Cayley-Dickson algebra. In this section and Section 10 generalized functions are considered on real valued test functions ω. Therefore, the statement of this theorem follows from Theorem 2.1, Example 2.6, Sections 2.4 and 2.6. $\square$

Corollary 2.8. *If*

$$\begin{aligned} Af = & \sum_{j,k} (\partial^2 f(z)/\partial z_k \partial z_j) \otimes a_{j,k}(z) \\ & + \sum_j (\partial f(z)/\partial z_j) \otimes b_j(z) \otimes 1 + f(z) \otimes \eta(z) \otimes 1 \end{aligned}$$

is a second order partial differential operator with generalized coefficients in either $\mathcal{B}'(U,\mathcal{A}_r)$ or $\mathcal{D}'(U,\mathcal{A}_r)$, where each $a_{j,k}$ is symmetric, f and $\mathcal{A}_r$ are as in Section 2.6, then three partial differential operators $\Upsilon+\beta$, $\Upsilon_1+\beta_1$ and Q of the first order with generalized coefficients with values in $\mathcal{A}_v$ for suitable $v\ge r$ of the same class exist such that

$$[Af,\omega^{\otimes 3})=[(\Upsilon+\beta)(\Upsilon_1+\beta_1)f+Qf,\omega^{\otimes 3}) \tag{64}$$

for each real-valued test function ω on U.

Proof. This follows from §2.2 [21], Theorem 2.7, Corollary 2.3, Example 2.6 and Section 2.1 above. □

Remark 2.9. *An integration technique and examples of PDE with generalized and discontinuous coefficients are planned to be presented in the next paper using results of this article.*

References

[1] J.C. Baez, *The octonions*, Bull. Amer. Math. Soc. **39** (2002), 145–205.
[2] L.E. Dickson, *The collected mathematical papers*, Chelsea Publishing Co., New York, 1975.
[3] I.M. Gelfand and G.E. Shilov, *Generalized functions and operations with them*, Fiz.-Mat. Lit., Moscow, 1958.
[4] G Grubb, *Functional calculus of pseudodifferential boundary problems*, Birkhäuser, Boston, 1996.
[5] K. Gürlebeck and W. Sprössig, *Quaternionic analysis and elliptic boundary value problem*, Birkhäuser, Basel, 1990.
[6] K. Gürlebeck and W. Sprössig, *Quaternionic and clifford calculus for physicists and engineers*, John Wiley and Sons, Inc., Chichester, 1997.
[7] K. Gürlebeck, K. Habetha and W. Sprössig, *Holomorphic functions in the plane and the n-dimensional space*, Birkhäuser, Basel, 2008.
[8] F. Gürsey and C.-H. Tze, *On the role of division, jordan and related algebras in particle physics*, World Scientific Publ. Co., Singapore, 1996.
[9] L. Hörmander, *The analysis of linear partial differential operators*, Springer, Berlin, 1985.
[10] I.L. Kantor and A.S. Solodovnikov, *Hypercomplex numbers*, Springer, Berlin, 1989.
[11] A. Kneschke, *Differentialgleichungen und randwertprobleme*, B.G. Teubner Verlaggesellschaft, Leipzig, 1965.
[12] A.G. Kurosh, *Lectures on general algebra*, Nauka, Moscow, 1973.
[13] H.B. Lawson and M.-L. Michelson., *Spin geometry*, Princ. Univ. Press, Princeton, 1989.
[14] S.V. Ludkovsky, *Differentiable functions of cayley-dickson numbers and line integration*, J. Math. Sci., N.Y. **141** (2007), 1231–1298.

[15] S.V. Ludkovsky, *Functions of several cayley-dickson variables and manifolds over them*, J. Math. Sci., N.Y. **141** (2007), 1299–1330.
[16] S.V. Ludkovsky, *Normal families of functions and groups of pseudoconformal diffeomorphisms of quaternion and octonion variables*, J. Math. Sci., N.Y. **150** (2008), 2224–2287.
[17] S.V. Ludkovsky, *Quasi-conformal functions of quaternion and octonion variables, their integral transformations*, Far East J. Math. Sci. (FJMS) **28** (2008), 37–88.
[18] S.V. Ludkovsky, *The two-sided laplace transformation over the cayley-dickson algebras and its applications*, J. Math. Sci., N.Y. **151** (2008), 3372–3430.
[19] S.V. Ludkovsky, *Residues of functions of octonion variables*, Far East J. Math. Sci. (FJMS) **39** (2010), 65–104.
[20] S.V. Ludkovsky, *Differential equations over octonions*, Adv. Appl. Clifford Algebr. **21** (2011), 773–797.
[21] S.V. Ludkovsky, *Line integration of dirac operators over octonions and cayley-dickson algebras*, Comput. Methods Funct. Theory **12** (2012), 279–306.
[22] S.V. Ludkovsky, *Multidimensional laplace transforms over quaternions, octonions and cayley-dickson algebras, their applications to pde*, Adv. Pure Math. **2** (2012), 63–103.
[23] S.V. Ludkovsky, *Dirac operators with generalized coefficients and their line integration*, Commun. Math. Appl. **4** (2013), 149–168.
[24] S.V. Ludkovsky and F. van Oystaeyen, *Differentiable functions of quaternion variables*, Bull. Sci. Math. (Paris) **127** (2003), 755–796.
[25] N.M. Matveev, *Methods of integration of ordinary differential equations*, Lan, St. Petersbourg, 2003.
[26] P.W. Michor, *Manifolds of differentiable mappings*, Shiva, Boston, 1980.
[27] I.G. Petrovsky, *Lectures on partial differential equations*, Gos. Izd. Techn. and Theoret. Liter., Moscow, 1953.
[28] I. Rubinstein and L. Rubinstein, *Partial differential equations in classical mathematical physics*, Cambridge Univ. Press, Cambridge, 1998.
[29] V.S. Vladimirov, *Equations of mathematical physics*, Nauka, Moscow, 1971.
[30] V.A. Zorich, *Mathematical analysis*, Nauka, Moscow, 1984.

Author's address:

Sergey V. Ludkowski
Dep. Appl. Mathematics
Moscow State Technical University MIREA
av. Vernadsky 78, Moscow, 119454, Russia
E-mail: `Ludkowski@mirea.ru`

Received August 25, 2013
Revised December 15, 2013

Rend. Istit. Mat. Univ. Trieste
Volume 46 (2014), 43–70

Existence and uniqueness of the gradient flow of the Entropy in the space of probability measures

Stefano Bianchini and Alexander Dabrowski

Abstract. *After a brief introduction on gradient flows in metric spaces and on geodesically convex functionals, we give an account of the proof (following the outline of [3, 7]) of the existence and uniqueness of the gradient flow of the Entropy in the space of Borel probability measures over a compact geodesic metric space with Ricci curvature bounded from below.*

Keywords: Entropy, gradient flows
MS Classification 2010: 35K90, 60B05, 28A33

1. Introduction

This paper aims to give an account of some of the main ideas from recent developments on gradient flows in metric measure spaces, examining the special case of the gradient flow of the Entropy functional in the space of probability measures. The results presented in this work are published in several texts, mainly [2, 3, 4, 7]; our aim is to give to the interested reader a single self-contained paper with both the proofs of existence and uniqueness of the gradient flow of the Entropy. We prove technical results when needed, however, to avoid excessive difficulties, in Section 4 we restrict our analysis to the case of a compact metric space.

We assume the reader has some familiarity with standard tools in measure theory; we recall the fundamental ones in Section 2.

In Section 3 we introduce the main concepts of the theory of gradient flows in a purely metric setting. We begin with a generalization to metric spaces of a property of gradient flows in the smooth setting, namely the Energy Dissipation Equality, which relies only on the norm of the differential of the functional and the norm of the derivative of the curve which solves the gradient flow. To make

This paper resumes the main part of the Bachelor thesis of the second author, discussed in 2013 at the University of Trieste.

sense of these two concepts in a metric space we introduce the *metric speed* of a curve and the *descending slope* of a functional. We proceed defining K-convexity in geodesic metric spaces and state a useful formula for computing the descending slope and an important weak form of the chain rule for K-convex functionals.

Section 4 is dedicated to the proof of existence and uniqueness of the gradient flow of the Entropy in the space of probability measures over a compact metric space. After defining the fundamental *Wasserstein distance* between probability measures, we introduce the Entropy functional and outline two interesting cases where a solution to a PDE is obtained as a solution of a gradient flow of a functional: the Dirichlet Energy in $L^2(\mathbb{R}^n)$ and the Entropy in the space of probability measures on the torus $\mathbb{T}^n$. We then give the definition of *geodesic* metric space with *Ricci curvature bounded from below*, a concept which will allow us to apply the theory developed for K-convex functionals to the Entropy. We proceed by proving the existence of a curve solving the gradient flow using a discrete approximation scheme and showing its convergence to a curve which satisfies the Energy Dissipation Equality.
The rest of the section deals with the uniqueness of the gradient flow of the Entropy. A deeper understanding of the nature of the curve solving the gradient flow is achieved introducing the concept of *push-forward via a plan* and restricting our analysis to plans with *bounded deformation*. After proving some preliminary properties concerning the approximation of the Entropy and the convexity of the squared descending slope of the Entropy, we conclude showing the uniqueness of the gradient flow.

2. Measure theoretic preliminaries

From now on, if not otherwise stated, $(X, \mathfrak{d})$ will be a complete and separable metric space with distance $\mathfrak{d}$. We will indicate as $\mathcal{P}(X)$ the set of Borel probability measures on X.

We recall two concepts we will often use: the push-forward of a measure through a map and narrow convergence of measures.

Definition 2.1. *Let X, Y be metric spaces, $\mu \in \mathcal{P}(X)$, $T : X \to Y$ a Borel map. We define the* push-forward *of μ through T as*

$$T_*\mu(E) := \mu(T^{-1}(E))$$

for every Borel subset $E \subseteq Y$.

The push-forward of a measure satisfies the following property: for every Borel function $f : Y \to \mathbb{R} \cup \{\infty\}$ it holds

$$\int_Y f \, T_*\mu = \int_X f \circ T \, \mu,$$

where the equality means that if one of the integrals exists so does the other one and their value is the same.

We give a useful weak notion of convergence in $\mathcal{P}(X)$.

DEFINITION 2.2. *Given $(\mu_n)_n$ a sequence of measures in $\mathcal{P}(X)$, μ_n* narrowly *converges to $\mu \in \mathcal{P}(X)$, and we write $\mu_n \rightharpoonup \mu$, if for every $\varphi \in C_b(X, \mathbb{R})$ it holds*

$$\int_X \varphi\, \mu_n \to \int_X \varphi\, \mu.$$

Notice that if X is compact, weak* convergence and narrow convergence on $\mathcal{P}(X)$ are the same, thanks to the Riesz-Markov-Kakutani Representation Theorem.

We pass to examine absolutely continuous measures.

DEFINITION 2.3. *Let λ, μ be measures on a σ-algebra $\mathcal{A}$. λ is* absolutely continuous *with respect to μ, and we write $\lambda \ll \mu$, if for every $E \in \mathcal{A}$ such that $\mu(E) = 0$, it also holds $\lambda(E) = 0$.*

The following two classic theorems will be widely used in the last section of this paper.

THEOREM 2.4 (Radon-Nikodym). *Let λ, μ be finite measures on X measurable space, $\lambda \ll \mu$. Then there exists a unique $h \in L^1(\mu)$ such that*

$$\lambda = h\mu. \tag{1}$$

To express (1) we also write synthetically

$$\frac{d\lambda}{d\mu} = h.$$

The analogue of Radon-Nikodym Theorem in the space of probability measures is the Disintegration Theorem.

THEOREM 2.5 (Disintegration Theorem). *Let $(X, \mathfrak{d}_X), (Y, \mathfrak{d}_Y)$ be complete and separable metric spaces, let $\gamma \in \mathcal{P}(X \times Y)$ and $\pi^1 : X \times Y \to X$ the projection on the first coordinate. Then there exists a $\pi^1_*\gamma$-almost everywhere uniquely determined family of probability measures $\{\gamma_x\}_{x\in X} \subseteq \mathcal{P}(Y)$ such that*

1. *the function $x \mapsto \gamma_x(B)$ is a Borel map for every Borel $B \subseteq Y$,*
2. *for every Borel function $f : X \times Y \to [0, \infty]$ it holds*

$$\int_{X\times Y} f(x, y)\, \gamma(dx, dy) = \int_X \left(\int_Y f(x, y)\, \gamma_x(dy) \right) \pi^1_*\gamma(dx). \tag{2}$$

We also express property (2) by writing

$$\gamma = \int_X \gamma_x \ \pi^1_* \gamma(dx).$$

The theorem obviously holds *mutatis mutandis* for the projection π^2 on Y.

For a proof of Theorem 2.5 and a broader view on the topic see [6, Chapter 45]. A simpler proof for vector valued measures can be found in [1, Theorem 2.28].

3. Gradient flows in metric spaces

The following observation is the starting point to extend the notion of gradient flows to metric spaces.

From now on we adopt the subscript notation for curves (i.e. $u_t = u(t)$).

REMARK 3.1. *Let H be a Hilbert space, $E : H \to \mathbb{R} \cup \{\infty\}$ a Fréchet differentiable functional. If $u : [0,\infty) \to \mathbb{R}$ is a gradient flow of E, i.e. $\dot{u}_t = -\nabla E(u_t)$, then*

$$\frac{d}{dt} E(u_t) = \langle \nabla E(u_t), \dot{u}_t \rangle = -\frac{1}{2}\|\dot{u}_t\|^2 - \frac{1}{2}\|\nabla E(u_t)\|^2.$$

Integrating with respect to t we obtain

$$E(u_s) - E(u_0) = -\frac{1}{2}\int_0^s \|\dot{u}_t\|^2 \ dt - \frac{1}{2}\int_0^s \|\nabla E(u_t)\|^2 \ dt \quad \forall s > 0.$$

This last equality is called Energy Dissipation Equality*.*

By extending appropriately the concepts of the norm of the derivative of a curve and the norm of the gradient of a functional we can make sense of this last equality even in metric spaces.

We restrict our analysis to a special class of curves.

DEFINITION 3.2. *A curve $u : [0,1] \to X$ is* absolutely continuous *if there exists $g \in L^1(I)$ such that for every $t < s$ it holds*

$$\mathfrak{d}(u_t, u_s) \leq \int_t^s g(r) \ dr. \tag{3}$$

For absolutely continuous curves we are able to define a corresponding concept of speed of a curve.

PROPOSITION 3.3. *If u is an absolutely continuous curve, there exists a minimal (in the L^1-sense) g which satisfies (3); this function is given for almost every t by*

$$|\dot{u}_t| := \lim_{s\to t} \frac{\mathfrak{d}(u_s, u_t)}{|s-t|}.$$

The function $|\dot{u}_t|$ is called *metric derivative* or *metric speed* of u.

Proof. Let $(y_n)_n$ be dense in $u(I)$ and define

$$h_n(t) := \mathfrak{d}(y_n, u_t) \quad \forall n \in \mathbb{N}.$$

Let $g \in L^1$ be such that

$$|h_n(t) - h_n(s)| \leq \mathfrak{d}(u_t, u_s) \leq \int_t^s g(r)\, dr \quad \forall n \in \mathbb{N}.$$

Therefore $h_n(t)$ are absolutely continuous for every n, so by the Lebesgue Fundamental Theorem of Calculus there exists $h_n' \in L^1$ such that

$$h_n(t) - h_n(s) = \int_t^s h_n'(r)\, dr.$$

We have that $|h_n'(t)| \leq g(t)$ a.e. and a fairly easy calculation shows that

$$\limsup_{s \to t} \frac{\mathfrak{d}(u_s, u_t)}{|s-t|} \leq \sup_n |h_n'(t)| \leq \liminf_{s \to t} \frac{\mathfrak{d}(u_s, u_t)}{|s-t|},$$

therefore we can take $\sup_n |h_n'(t)|$ as the metric derivative. □

We pass to define the concept which will substitute the norm of the differential of a function in the metric case.

We indicate by $(\,\cdot\,)^+$, $(\,\cdot\,)^-$ the standard positive and negative parts, i.e. $x^+ = \max\{x, 0\}$, $x^- = \max\{-x, 0\}$.

Definition 3.4. *Let $E : X \to \mathbb{R} \cup \{\infty\}$. The* descending slope *of E at x is*

$$|D^- E|(x) := \limsup_{y \to x} \frac{\big(E(x) - E(y)\big)^+}{\mathfrak{d}(x,y)}.$$

We are now ready to define gradient flows using the Energy Dissipation Equality (EDE).

For a functional $E : X \to \mathbb{R} \cup \{\infty\}$, we write $D(E) := \{x : E(x) < \infty\}$.

Definition 3.5 (Gradient flow - EDE). *Let E be a functional from X to $\mathbb{R} \cup \{\infty\}$ and let $x_0 \in D(E)$. A locally absolutely continuous curve $x : [0, \infty) \to X$ is an* EDE-gradient flow, *or simply a* gradient flow, *of E starting from x_0 if x takes values in $D(E)$ and it holds*

$$E(x_s) = E(x_t) - \frac{1}{2}\int_t^s |\dot{x}_r|^2\, dr - \frac{1}{2}\int_t^s |D^- E|^2(x_r)\, dr, \quad \forall s > t,$$

or equivalently

$$E(x_s) = E(x_0) - \frac{1}{2}\int_0^s |\dot{x}_r|^2\, dr - \frac{1}{2}\int_0^s |D^- E|^2(x_r)\, dr, \quad \forall s > 0.$$

3.1. K-convexity

A class of functionals with useful properties is that of K-convex functionals.

In $\mathbb{R}^n$ the standard definition is that the distributional derivative of a function $E : \mathbb{R}^n \mapsto \mathbb{R}$ satisfies

$$D^2E - KI_n \geq 0,$$

where I_n is the $n \times n$-identity matrix. To extend the definition from the smooth setting to metric spaces we will use geodesics.

Definition 3.6. *A metric space X is* geodesic *if $\forall x_0, x_1 \in X, \exists g : [0,1] \to X$ such that $g_0 = x_0, g_1 = x_1$ and*

$$\mathfrak{d}(g_t, g_s) = |t-s|\mathfrak{d}(x_0, x_1), \quad \forall s,t \in [0,1].$$

Such a g is called constant speed geodesic *between x_0 and x_1.*

It is natural to extend the definition of K-convexity by requiring the K-convexity of the functional along geodesics.

Definition 3.7. *Let $(X, \mathfrak{d})$ be a geodesic space, $E : X \to \mathbb{R} \cup \{\infty\}$. E is* K-geodesically convex*, or simply* K-convex*, if $\forall x_0, x_1 \in Y, \exists g : [0,1] \to X$ constant speed geodesic between x_0 and x_1 and for every $t \in [0,1]$ it holds*

$$E(g_t) \leq (1-t)E(x_0) + tE(x_1) - \frac{K}{2}t(1-t)\mathfrak{d}^2(x_0, x_1).$$

Remark 3.8. *Notice that if E is K-convex then for every $K' \leq K, E$ is K'-convex.*

We prove a useful formula for computing the descending slope of K-convex functionals.

Lemma 3.9. *If E is K-convex then*

$$|D^-E|(x) = \sup_{y \neq x} \left(\frac{E(x) - E(y)}{\mathfrak{d}(x,y)} + \frac{K}{2}\mathfrak{d}(x,y) \right)^+. \tag{4}$$

Proof. " $\leq$ ". This inequality holds trivially.
" $\geq$ ". Fix $y \neq x$. Let g be a constant speed geodesic from x to y such that

$$\begin{aligned} \frac{E(x) - E(g_t)}{\mathfrak{d}(x, g_t)} &\geq \frac{t}{\mathfrak{d}(x, g_t)} \left(E(x) - E(y) + \frac{K}{2}(1-t)\mathfrak{d}^2(x,y) \right) \\ &= \frac{E(x) - E(y)}{\mathfrak{d}(x,y)} + \frac{K}{2}(1-t)\mathfrak{d}(x,y). \end{aligned}$$

Therefore as $t \to 0$,

$$
\begin{aligned}
|D^- E|(x) &\geq \limsup_{t\to 0^+} \left(\frac{E(x) - E(g_t)}{\mathfrak{d}(x, g_t)} \right)^+ \\
&\geq \left(\limsup_{t\to 0^+} \left(\frac{E(x) - E(y)}{\mathfrak{d}(x, y)} + \frac{K}{2}(1-t)\mathfrak{d}(x,y) \right) \right)^+ \\
&= \left(\frac{E(x) - E(y)}{\mathfrak{d}(x, y)} + \frac{K}{2}\mathfrak{d}(x,y) \right)^+ .
\end{aligned}
$$

We conclude by taking the supremum w.r.t. y. □

For K-convex functionals we have a useful weak form of chain rule.

THEOREM 3.10. *Let $E : X \to \mathbb{R}\cup\{\infty\}$ be a K-convex and lower semicontinuous functional. Then for every absolutely continuous curve $x : [0,1] \to X$ such that $E(x_t) < \infty$ for every $t \in [0,1]$, it holds*

$$
|E(x_s) - E(x_t)| \leq \int_t^s |\dot{x}_r||D^- E|(x_r)\, dr, \tag{5}
$$

with $t < s$.

Proof. We follow the reasoning of [2, Proposition 3.19].

Step 0. By linear scaling we may reduce to the case $t = 0$ and $s = 1$. We may also assume that

$$
\int_0^1 |\dot{x}_r||D^- E|(x_r)\, dr < +\infty,
$$

otherwise the inequality holds trivially. By the standard arc-length reparametrization we may furthermore assume $|\dot{x}_t| = 1$ for almost every t, so x_t is 1-Lipschitz and the function $t \mapsto |D^- E|(x_t)$ is in $L^1([0,1])$.

Step 1. Notice that it is sufficient to prove absolute continuity of the function $t \mapsto E(x_t)$, then the thesis follows from the inequality

$$
\begin{aligned}
&\limsup_{h\to 0} \frac{E(x_{t+h}) - E(x_t)}{h} \leq \limsup_{h\to 0} \frac{\big(E(x_{t+h}) - E(x_t)\big)^+}{|h|} \\
&\leq \limsup_{h\to 0} \frac{\big(E(x_{t+h}) - E(x_t)\big)^+}{\mathfrak{d}(x_{t+h}, x_t)} \limsup_{h\to 0} \frac{\mathfrak{d}(x_{t+h}, x_t)}{|h|} \leq |D^- E|(x_t)|\dot{x}_t|
\end{aligned}
$$

and the fact that for a.c. f it holds

$$
f(s) - f(t) = \int_t^s \frac{df}{d\tau} d\tau.
$$

Step 2. We define $f, g : [0,1] \to \mathbb{R}$ as

$$
\begin{aligned}
f(t) &:= E(x_t),\\
g(t) &:= \sup_{s \neq t} \frac{\big(f(t) - f(s)\big)^+}{|t-s|}.
\end{aligned}
$$

From the fact that $|\dot{x}_t| = 1$ and the trivial inequality $a^+ \leq (a+b)^+ + b^-$ valid for any $a, b \in \mathbb{R}$ we obtain

$$
\begin{aligned}
g(t) &\leq \sup_{s \neq t} \frac{\big(f(t) - f(s)\big)^+}{\mathfrak{d}(x_t, x_s)}\\
&\leq \left(\sup_{s \neq t} \frac{f(t) - f(s)}{\mathfrak{d}(x_t, x_s)} + \frac{K}{2}\mathfrak{d}(x_t, x_s)\right)^+ + \left(\frac{K}{2}\mathfrak{d}(x_t, x_s)\right)^- .
\end{aligned}
$$

Since $\{x_t\}_{t \in [0,1]}$ is compact, there exists a $D \in \mathbb{R}^+$ such that $\mathfrak{d}(x_t, x_s) \leq D$; applying then (4) we obtain

$$
g(t) \leq |D^- E|(x_t) + \frac{K^-}{2} D.
$$

Therefore the thesis is proven if we show that

$$
|f(s) - f(t)| \leq \int_t^s g(r)\, dr.
$$

Step 3. Fix $M, \varepsilon > 0$ and define $f^M := \min\{f, M\}$, $\rho_\varepsilon : \mathbb{R} \to \mathbb{R}$ a smooth mollifier with support in $[-\varepsilon, \varepsilon]$ and $f^M_\varepsilon, g^M_\varepsilon : [\varepsilon, 1-\varepsilon] \to \mathbb{R}$ such that

$$
\begin{aligned}
f^M_\varepsilon(t) &:= (f^M * \rho_\varepsilon)(t),\\
g^M_\varepsilon(t) &:= \sup_{s \neq t} \frac{\big(f^M_\varepsilon(t) - f^M_\varepsilon(s)\big)^+}{|s-t|}.
\end{aligned}
$$

Since f^M_ε is smooth and $g^M_\varepsilon \geq \big|(f^M_\varepsilon)'\big|$,

$$
\big|f^M_\varepsilon(s) - f^M_\varepsilon(t)\big| \leq \int_t^s g^M_\varepsilon(r)\, dr.
$$

Therefore we have

$$\begin{aligned}
g_\varepsilon^M(t) &= \sup_{s\neq t}\frac{1}{|s-t|}\left(\int_0^1 \big(f^M(t-r)-f^M(s-r)\big)\rho_\varepsilon(r)\,dr\right)^+\\
&\le \sup_{s\neq t}\frac{1}{|s-t|}\int_0^1 \big(f(t-r)-f(s-r)\big)^+\rho_\varepsilon(r)\,dr\\
&\le \sup_{s\neq t}\int_0^1 \frac{\big(f(t-r)-f(s-r)\big)^+}{|(s-r)-(t-r)|}\rho_\varepsilon(r)\,dr\\
&\le \int_0^1 g(t-r)\rho_\varepsilon(r)\,dr = (g*\rho_\varepsilon)(t),
\end{aligned}$$

thus the family $\left\{g_\varepsilon^M\right\}_\varepsilon$ is uniformly integrable in $L^1((0,1))$. In fact, for $A\subset(\varepsilon,1-\varepsilon)$,

$$\begin{aligned}
\int_A g_\varepsilon^M(t)\,dt \le \int_A (g*\rho_\varepsilon)(t)\,dt = \int_A\int_0^1 g(t-y)\rho_\varepsilon(y)\,dy\,dt\\
= \int_0^1 \frac{1}{\varepsilon}\rho\Big(\frac{y}{\varepsilon}\Big)\int_{A-y} g(t)\,dt\,dy = \int_0^1 \rho(z)\left[\int_{A-\varepsilon z} g(t)\,dt\right]dz \le \omega(|A|),
\end{aligned}$$

where $\omega(|A|)=\sup_{\mathcal{L}(B)=\mathcal{L}(A)}\{\int_B g\}$.
Since

$$|f_\varepsilon^M(y)-f_\varepsilon^M(x)|\le\omega(|y-x|)\quad\text{and}\quad \lim_{z\to 0}\omega(z)=0,$$

the family $\left\{f_\varepsilon^M\right\}_\varepsilon$ is equicontinuous in $C([\delta,1-\delta])$ for every δ fixed. Hence by Arzelà-Ascoli Theorem, up to subsequences, the family $\left\{f_\varepsilon^M\right\}_\varepsilon$ uniformly converges to a function $\tilde f^M$ on $(0,1)$ as $\varepsilon\to 0$ for which it holds

$$|\tilde f^M(s)-\tilde f^M(t)|\le\int_t^s g(r)\,dr.$$

By the fact that $f_\varepsilon^M\to f^M$ in L^1, $f^M=\tilde f^M$ on a $A\subseteq[0,1]$ such that $[0,1]\setminus A$ has negligible Lebesgue measure.

Step 4. Now we prove that $f^M=\tilde f^M$ everywhere. f^M is lower semicontinuous and $\tilde f^M$ is continuous, hence $f^M\le\tilde f^M$ in $[0,1]$. Suppose by contradiction that there are $t_0\in(0,1)$, $c,C\in\mathbb{R}$ such that $f^M(t_0)<c<C<\tilde f^M(t_0)$, so there exists $\delta>0$ such that $\tilde f^M(t)>C$ for $t\in[t_0-\delta,t_0+\delta]$. Thus $f^M(t)>C$ for $t\in[t_0-\delta,t_0+\delta]\cap A$, so

$$\int_0^1 g(t)\,dt\ge\int_{[t_0-\delta,t_0+\delta]\cap A} g(t)\,dt\ge\int_{[t_0-\delta,t_0+\delta]\cap A}\frac{C-c}{|t-t_0|}\,dt=+\infty,$$

which is absurd since $g\in L^1(\mathbb{R})$.

Conclusion. Thus we proved that if $g \in L^1((0,1))$,

$$|f^M(s) - f^M(t)| \leq \int_t^s g(r)\, dr, \quad \forall t < s, \forall M > 0.$$

Letting $M \to \infty$ the thesis is proven. □

Notice that an application of Young's inequality on (5) gives

$$E(x_s) - E(x_t) \geq -\frac{1}{2}\int_t^s |\dot{x}_r|^2\, dr - \frac{1}{2}\int_t^s |D^- E|^2(x_r)\, dr, \quad \forall t < s, \tag{6}$$

Therefore, a K-convex functional satisfies the Energy Dissipation Equality if we require only a minimum dissipation of E along the curve, in particular

$$E(x_t) \geq E(x_s) + \frac{1}{2}\int_t^s |\dot{x}_r|^2\, dr + \frac{1}{2}\int_t^s |D^- E|^2(x_r)\, dr, \quad \forall t < s. \tag{7}$$

4. The Entropy functional in the space of probability measures and its gradient flow

For simplicity, in all this section we will restric our analysis to a compact metric space $(X, \mathfrak{d})$.

4.1. The Wasserstein distance

We can equip the space of probability measures with a natural distance obtained by the minimization problem of Optimal Transport theory.

Definition 4.1. *Given $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$, we define the set of* admissible plans *from μ to ν as*

$$\mathrm{Adm}(\mu, \nu) := \Big\{\gamma \in \mathcal{P}(X \times Y) : \pi^X_* \gamma = \mu, \pi^Y_* \gamma = \nu\Big\},$$

where π^X, π^Y is the projection on X, Y.

Definition 4.2. *Given $\mu, \nu \in \mathcal{P}(X)$, the* Wasserstein distance *between μ and ν is*

$$W_2(\mu, \nu) := \sqrt{\inf_{\gamma \in \mathrm{Adm}(\mu,\nu)} \int \mathfrak{d}^2(x, y)\, \gamma(dx, dy)}.$$

The space of probability measures $\mathcal{P}(X)$ endowed with the Wasserstein distance inherits many of the properties of the underlying space X. We point out just two of them:

- Given a sequence $(\mu_n)_n$ in $\mathcal{P}(X)$, it holds

$$W_2(\mu_n, \mu) \to 0 \quad \Leftrightarrow \quad \mu_n \rightharpoonup \mu. \tag{8}$$

  Notice that as a consequence sequential narrow compactness and narrow compactness coincide. (In the case of non compact metric spaces, one needs an additional condition on the right hand side of the equivalence (8).)

- If X is geodesic then $\mathcal{P}(X)$ is also geodesic.

For the proofs and a generalization to non-compact spaces see [2, Theorems 2.7 and 2.10].

Before passing to the definition of Entropy, we prove a property of the metric derivative which we will use later on.

Remark 4.3. *If μ_t, ν_t are absolutely continuous curves in $(\mathcal{P}(X), W_2)$, then $\forall k \in [0,1], \eta_t := (1-k)\mu_t + k\nu_t$ is absolutely continuous and it holds*

$$|\dot{\eta}_t|^2 \leq (1-k)|\dot{\mu}_t|^2 + k|\dot{\nu}_t|^2.$$

In fact, it is easy to prove the convexity of the squared Wasserstein distance w.r.t. linear interpolation of measures, i.e.

$$W_2^2\big((1-t)\mu_0 + t\mu_1, (1-t)\nu_0 + t\nu_1\big) \leq (1-t)W_2^2(\mu_0, \nu_0) + tW_2^2(\mu_1, \nu_1)$$

for arbitrary $\mu_0, \mu_1, \nu_0, \nu_1 \in \mathcal{P}(X)$. Applying the definition of metric speed, the estimate above follows immediately.

4.2. Entropy: definition and properties

Definition 4.4. *The* Entropy functional $\mathrm{Ent}_m : \mathcal{P}(X) \to [0, \infty]$ *relative to $m \in \mathcal{P}(X)$ is defined as*

$$\mathrm{Ent}_m(\mu) := \begin{cases} \displaystyle\int_X f \log f \; m & \textit{if } \exists f \in L^1(m) : \mu = fm, \\ \infty & \textit{otherwise.} \end{cases}$$

Example 4.6 below suggest that the gradient flow of the Entropy functional in measure metric spaces is the natural extension of the heat flow from the smooth setting. As a consequence, it is possible to construct the analogue of the Laplace operator, which is the starting point to construct several tools used in Analysis on measure metric spaces.

A very interesting fact is that gradient flows of certain functionals in carefully selected spaces generate solutions to well known PDEs (see the seminal papers [8], [10]). In this context we show two interesting examples of solutions to a PDE generated by a gradient flow.

EXAMPLE 4.5. *The gradient flow of the Dirichlet Energy functional* $D : L^2(\mathbb{R}^n) \to \mathbb{R}$ *defined as*

$$D(f) := \frac{1}{2}\|\nabla f\|^2_{L^2}$$

produces a solution of the heat equation in $\mathbb{R}^n$.

We sketch the proof assuming that the functions are smooth. Differentiating $D(f)$ *along* v *we obtain*

$$\lim_{t\to 0}\frac{1}{2}\frac{\|\nabla(f+tv)\|^2_{L^2} - \|\nabla f\|^2_{L^2}}{t} = \int \langle \nabla f, \nabla v\rangle,$$

which can be rewritten, using Green's first identity, as $-\int v\Delta f$. *Therefore we conclude that*

$$-\nabla D(f) = \Delta f.$$

EXAMPLE 4.6. *The heat equation is also obtained as a solution of the Entropy gradient flow in* $(\mathcal{P}_2(X), W_2)$. *We give an informal proof of this fact in the case* $X = \mathbb{T}^n$ *the* n*-torus. We refer to [3, Chapters 8-10] and to [10] for a more detailed approach.*

Let $f : \mathbb{T}^n \to \mathbb{R}$ *be integrable and s.t.* $\int_{\mathbb{T}^n} f\, \mathcal{L} = 1$ *where* $\mathcal{L}$ *is the Lebesgue measure on the torus* $\mathbb{T}^n$. *Define* $\mu := f\mathcal{L}$. *The natural space of perturbations (tangent vector fields) in the metric space* $(\mathcal{P}(\mathbb{T}^n), W_2)$ *in the point* $\mu = f\mathcal{L}$ *are vector fields* $v : \mathbb{T}^n \to \mathbb{T}^n$ *square integrable w.r.t.* μ, *corresponding to the perturbations (see Definition 2.31 of [2])*

$$\partial_t f = -\mathrm{div}_x(vf).$$

Inserting these perturbations in $\mathrm{Ent}_{\mathcal{L}}(f)$ *we obtain*

$$\frac{d}{dt}\mathrm{Ent}_{\mathcal{L}}\big((I+tv)_*(f\mathcal{L})\big) = -\int \mathrm{div}_x(vf)\big(\log f + 1\big)\,dx$$
$$= -\int \mathrm{div}_x(vf)\log f\,dx = \int (vf)\cdot\nabla_x \log f\,dx = \int v\cdot\nabla_x f\,dx.$$

To find the norm of the gradient we have to maximize it with respect to v *with the restriction* $\int |v|^2 f\,dx \le 1$. *With a fairly easy calculation we obtain*

$$\bar{v} = -\frac{1}{\alpha}\nabla_x \log f \quad \textit{where} \quad \alpha = \sqrt{\int f|\nabla_x \log f|^2\,dx}.$$

From

$$\frac{d}{dt}\mathrm{Ent}_{\mathcal{L}}\big((I+t\bar{v})_*(f\mathcal{L})\big) = -\alpha,$$

we conclude that the gradient flow is $\nabla\,\mathrm{Ent}_{\mathcal{L}}(f) = -\nabla_x \log f$, *and finally that*

$$\partial_t f = -\mathrm{div}_x((\nabla_x \log f)f) = \mathrm{div}_x(\nabla_x f) = \Delta f.$$

In the following two propositions we prove strict convexity w.r.t. linear interpolation and lower semicontinuity of the Entropy.

PROPOSITION 4.7. *The Entropy functional is strictly convex with respect to linear interpolation of measures, i.e. given* $\mu_0, \mu_1 \in D(\mathrm{Ent}_m)$*,*

$$\mathrm{Ent}_m\big((1-t)\mu_0 + t\mu_1\big) \le (1-t)\,\mathrm{Ent}_m(\mu_0) + t\,\mathrm{Ent}_m(\mu_1), \quad \forall t \in [0,1],$$

and equality holds if and only if $\mu_0 = \mu_1$*.*

Proof. Let $\mu_0 = f_0 m$, $\mu_1 = f_1 m$, $u(z) := z\log z$. Since u is strictly convex, it holds

$$u((1-t)f_0 + tf_1) \le (1-t)u(f_0) + tu(f_1), \quad \forall t \in [0,1],$$

and equality holds if and only if $f = g$. Integrating we obtain the thesis. □

PROPOSITION 4.8. *The Entropy functional is lower semicontinuous with respect to narrow convergence of measures.*

Proof. Given $\varphi \in C(X)$ let $G_\varphi : \mathcal{P}(X) \to \mathbb{R}$ be such that

$$G_\varphi(\mu) := \int_X \varphi\,\mu - \int_X e^{\varphi-1}\,m.$$

Notice that G_φ is continuous with respect to narrow convergence by definition. Define now

$$F(\mu) := \sup_{\phi\in C(X)} G_\phi(\mu).$$

If $\mu \perp m$, by varying ϕ we can obtain arbitrary large values for $\int \phi\ \mu$ without increasing $\int e^{\phi-1}\ m$ more than 1; therefore $F(\mu) = \infty$.

If $\mu \ll m$, there exists a non-negative $f \in L^1(m)$ s.t. $\mu = fm$. It is easily verified that $1 + \log f$ maximizes F, therefore by a standard approximation technique we have

$$F(\mu) = \int (1+\log f)f - e^{\log f}\ m = \int f\log f\ m,$$

and

$$F(\mu) = \mathrm{Ent}_m(\mu).$$

But F is the supremum of continuous functions, therefore it is l.s.c.. □

We now define boundedness from below of the Ricci curvature, relying on the definition by Sturm and Lott-Villani (see [9] and [11]).

Definition 4.9. *Let $(X, \mathfrak{d}, m)$ be a compact geodesic metric space with $m \in \mathcal{P}(X)$. X has* Ricci curvature *bounded from below by $K \in \mathbb{R}$, and we write $(X, \mathfrak{d}, m)$ is a $CD(K, \infty)$* space, *if* Ent_m *is K-convex in $(\mathcal{P}(X), W_2)$, i.e. for every pair of points $\mu_0, \mu_1 \in D(\mathrm{Ent}_m)$, there exists $\mu_t : [0,1] \to \mathcal{P}(X)$ constant speed geodesic between μ_0 and μ_1, such that*

$$\mathrm{Ent}_m(\mu_t) \leq (1-t)\,\mathrm{Ent}_m(\mu_0) + t\,\mathrm{Ent}_m(\mu_1) - \frac{K}{2} t(1-t) W_2^2(\mu_0, \mu_1), \quad \forall t \in [0,1].$$

The previous definition will allow us to apply the properties of K-convex functionals to the Entropy.

4.3. Existence of the gradient flow of the Entropy

The proof of the existence of a gradient flow of the Entropy functional relies on a variational approach which dates back to De Giorgi (see [5]). We will define a recursive scheme and prove that it converges to a solution of the gradient flow of the Entropy. The recursive scheme is obtained with the following functional.

Definition 4.10. *For $\tau > 0$, and $\mu \in \mathcal{P}(X)$ define $J_\tau \mu$ as the probability $\sigma \in \mathcal{P}(X)$ which minimizes*

$$\sigma \mapsto \mathrm{Ent}_m(\sigma) + \frac{W_2^2(\sigma, \mu)}{2\tau}.$$

Since $P(X)$ is compact w.r.t. narrow convergence, the existence and uniqueness of the minimizer is obtained through lower semicontinuity and strict convexity of the Entropy.

We prove an important estimate on the curve $t \mapsto J_t \mu$, which is almost the EDE we are looking for.

Lemma 4.11. *Let $\mu \in \mathcal{P}(X)$. Then it holds*

$$\mathrm{Ent}_m(\mu) = \mathrm{Ent}_m(J_t \mu) + \frac{W_2^2(J_t \mu, \mu)}{2t} + \int_0^t \frac{W_2^2(J_r \mu, \mu)}{2r^2}\, dr. \tag{9}$$

Proof. By the definition of $J_t \mu$ we have

$$\begin{aligned}
&\frac{W_2^2(J_t\mu, \mu)}{2t} - \frac{W_2^2(J_t\mu, \mu)}{2s} \\
\leq\ & \mathrm{Ent}_m(J_t\mu) + \frac{W_2^2(J_t\mu, \mu)}{2t} - \left(\mathrm{Ent}_m(J_s\mu) + \frac{W_2^2(J_s\mu, \mu)}{2s} \right) \\
\leq\ & \frac{W_2^2(J_s\mu, \mu)}{2t} - \frac{W_2^2(J_s\mu, \mu)}{2s},
\end{aligned}$$

and passing to the limit we obtain

$$\lim_{s\to t}\frac{1}{t-s}\left(\mathrm{Ent}_m(J_t\mu)+\frac{W_2^2(J_t\mu,\mu)}{2t}-\left(\mathrm{Ent}_m(J_s\mu)+\frac{W_2^2(J_s\mu,\mu)}{2s}\right)\right)$$
$$= -\frac{W_2^2(J_t\mu,\mu)}{2t^2}.$$

Since the left hand side is the derivative of

$$r\mapsto \mathrm{Ent}_m(J_r\mu)+\frac{W_2^2(J_r\mu,\mu)}{2r},$$

by integration we obtain

$$\int_0^t \frac{d}{dr}\left(\mathrm{Ent}_m(J_r\mu)+\frac{W_2^2(J_r\mu,\mu)}{2r}\right)dr = -\int_0^t \frac{W_2^2(J_r\mu,\mu)}{2r^2}\,dr.$$

If we show that

$$\lim_{x\to 0^+}\frac{W_2^2(J_x\mu,\mu)}{2x}=0,$$

we will thus have the thesis. In fact, since $\mathrm{Ent}_m(\mu)<\infty$, then by definition of $J_r\mu$ we have

$$0\le \mathrm{Ent}_m(J_r\mu)+\frac{W_2^2(J_r\mu,\mu)}{2r}\le \mathrm{Ent}_m(\mu),$$

and the lower semicontinuity of the Entropy yields $\mathrm{Ent}_m(J_r\mu)\ge \mathrm{Ent}_m(\mu)$ as $r\to 0$. □

We are ready to prove the first main result of this chapter, i.e. the existence of an EDE-gradient flow of the Entropy for metric measure spaces with Ricci curvature bounded from below.

THEOREM 4.12 (Existence). *If $(X,\mathfrak{d},m)$ has Ricci curvature bounded from below by K then for every $\tilde\mu\in D(\mathrm{Ent}_m)$ there exists a gradient flow of* Ent_m *starting from $\tilde\mu$.*

Proof. The proof will be given in several steps. Notice that in order to prove the EDE, by (7) it is enough to show there exists an absolutely continuous curve $t\mapsto\mu_t$ such that $\mu_0=\tilde\mu$ and

$$\mathrm{Ent}_m(\tilde\mu)\ge \mathrm{Ent}_m(\mu_t)+\frac12\int_0^t|\dot\mu|^2(r)\,dr+\frac12\int_0^t|D^-\mathrm{Ent}_m|^2(\mu_r)\,dr,\quad \forall t\ge 0.$$

Step 1. The approximate solution is constructed by defining recursively

$$\mu_0^\tau := \tilde\mu,$$
$$\mu_{n+1}^\tau := J_\tau(\mu_n^\tau).$$

Then define the curve $t \mapsto \mu^\tau(t)$ as

$$\mu^\tau(n\tau) := \mu_n^\tau,$$
$$\mu^\tau(t) := J_{t-n\tau}(\mu_n^\tau), \quad \forall t \in (n\tau, (n+1)\tau),$$

and let

$$|\dot\mu^\tau|(t) := \frac{W_2(\mu_n^\tau, \mu_{n+1}^\tau)}{2\tau}, \quad \forall t \in [n\tau, (n+1)\tau).$$

Step 2. We give an estimate on the descending slope of the Entropy. Given $\nu, \sigma \in \mathcal{P}(X)$, since $J_t(\sigma)$ is the minimizer, we have

$$\mathrm{Ent}_m(J_t\sigma) + \frac{W_2^2(\sigma, J_t\sigma)}{2t} \le \mathrm{Ent}_m(\nu) + \frac{W_2^2(\sigma, \nu)}{2t}.$$

Hence by triangle inequality

$$\begin{aligned}
\mathrm{Ent}_m(J_t\sigma) - \mathrm{Ent}_m(\nu) &\le \frac{1}{2t}\Big(W_2^2(\sigma,\nu) - W_2^2(\sigma, J_t\sigma)\Big) \\
&= \frac{1}{2t}\Big(W_2(\sigma,\nu) - W_2(\sigma, J_t\sigma)\Big)\Big(W_2(\nu,\sigma) + W_2(\sigma, J_t\sigma)\Big) \\
&\le \frac{W_2(J_t\sigma, \nu)}{2t}\Big(W_2(\nu,\sigma) + W_2(\sigma, J_t\sigma)\Big).
\end{aligned}$$

If $J_t\sigma = \nu$ the inequality holds trivially. Otherwise, dividing by $W_2(J_t\sigma, \nu)$ both sides and passing to the limit

$$\begin{aligned}
|D^- \mathrm{Ent}_m|(J_t\sigma) &= \limsup_{\nu \to J_t\sigma} \frac{\big(\mathrm{Ent}_m(J_t\sigma) - \mathrm{Ent}_m(\nu)\big)^+}{W_2(J_t\sigma, \nu)} \\
&\le \limsup_{\nu \to J_t\sigma} \frac{W_2(\nu,\sigma) + W_2(\sigma, J_t\sigma)}{2t} = \frac{W_2(\sigma, J_t\sigma)}{t}.
\end{aligned}$$

Step 3. Using now the curve $\mu^\tau(t)$, the definition of its time derivative $|\dot\mu^\tau|$ and the previous inequality, we can thus rewrite (9) as

$$\mathrm{Ent}_m(\mu_n^\tau) \ge \mathrm{Ent}_m(\mu_{n+1}^\tau) + \frac{1}{2}(2\tau)|\dot\mu^\tau|^2(t) + \frac{1}{2}\int_{n\tau}^{(n+1)\tau} |D^- \mathrm{Ent}_m|^2\big(\mu^\tau(s)\big)\, ds.$$

where $t \in [n, n+1)\tau$. Adding these inequalities from 0 to $T = N\tau$ we obtain

$$\mathrm{Ent}_m(\tilde\mu) \ge \mathrm{Ent}_m(\mu_T^\tau) + \frac{1}{2}\int_0^T |\dot\mu^\tau|^2(t)\, dt + \frac{1}{2}\int_0^T |D^- \mathrm{Ent}_m|^2\big(\mu^\tau(t)\big)\, dt. \tag{10}$$

Notice that the definition of $|\dot\mu^\tau|$ implies that we rescale time as $t \mapsto t/2$, i.e. we take 2τ to pass from μ_n^τ to μ_{n+1}^τ.

The last part of the proof concerns the compactness of the family of curves $\{t \mapsto \mu_t^\tau\}_\tau$ in the space $C(X, \mathcal{P}(X))$ and the lower semicontinuity (with respect to τ in 0) of the right hand side of (10). These results clearly will conclude the proof.

Step 4. We address the convergence of the curve $t \mapsto \mu^\tau(t)$ as $\tau \to 0$. First, since the starting point is in the domain of the Entropy we have by (10) that

$$\int_0^T |\dot{\mu}^\tau|^2(t)\, dt < 2\,\mathrm{Ent}_m(\tilde{\mu}).$$

By Hölder inequality, for all Borel $A \subset [0, T]$

$$\begin{aligned}\int_A |\dot{\mu}^\tau|(t)\, dt = \int_0^T \chi_A |\dot{\mu}^\tau|(t)\, dt &\leq \sqrt{\int_0^T \chi_A^2\, dt}\sqrt{\int_0^T |\dot{\mu}^\tau|^2(t)\, dt} \\ &\leq \sqrt{\mathcal{L}(A)}\sqrt{2\,\mathrm{Ent}_m(\tilde{\mu})},\end{aligned}$$

which gives the uniform integrability of $|\dot{\mu}^\tau(t)|$. Since

$$W_2(\mu^\tau(s), \mu^\tau(t)) \leq \int_t^s |\dot{\mu}^\tau|(r)\, dr,$$

the family of curves $(\mu^\tau)_\tau$ is uniformly continuous.
Up to subsequences we can pass to the limit as $\tau_n \searrow 0$, obtaining by Arzelà-Ascoli Theorem that μ^{τ_n} converges uniformly to a curve $t \mapsto \mu(t)$ such that $\mu(0) = \tilde{\mu}$. Since $|\dot{\mu}^\tau(t)|$ is uniformly integrabile, up to subsequences, also $|\dot{\mu}^{\tau_n}(t)|$ L^1-weakly converges to a function g. It follows easily from the definition of metric derivative that $|\dot{\mu}(t)| \leq g(t)$, therefore $\mu(t)$ is locally absolutely continuous.

Step 5. We prove the lower semicontinuity of the right hand side of (10). By Hölder's inequality we have

$$\int_0^T |\dot{\mu}|^2(t)\, dt \leq \int_0^T g(t)^2\, dt \leq \liminf_n \int_0^T |\dot{\mu}^{\tau_n}|^2(t)\, dt,$$

thus the l.s.c. of the first integral.

Notice that being the supremum of lower semicontinuous functions by formula (4), $|D^- \mathrm{Ent}_m|$ is lower semicontinuous too. Then define for every $k \in \mathbb{N}, \nu \in \mathcal{P}(X)$

$$e_k(\nu) := \inf_{\sigma \in \mathcal{P}(X)} \Big\{ |D^- \mathrm{Ent}_m|^2(\sigma) + kW_2(\nu, \sigma) \Big\}.$$

Notice that $\sup_k e_k(\nu) = |D^- \mathrm{Ent}_m|^2(\nu)$ for all ν.
The infimum of Lipschitz functions bounded from below is a Lipschitz function;

therefore e_k is Lipschitz. By uniform convergence

$$\lim_n \int_0^T e_k(\mu^{\tau_n}(t))\, dt = \int_0^T e_k(\mu(t))\, dt.$$

By Fatou Lemma

$$\int e_k(\mu(t))\, dt \le \liminf_n \int |D^- \operatorname{Ent}_m|^2(\mu^{\tau_n}(t))\, dt,$$

for every $k \in \mathbb{N}$. By Monotone Convergence Theorem we finally have

$$\int |D^- \operatorname{Ent}_m|^2(\mu(t))\, dt = \sup_k \int e_k(\mu(t))\, dt \le \liminf_n \int |D^- \operatorname{Ent}_m|^2(\mu^{\tau_n}(t))\, dt.$$

□

4.4. Entropy: uniqueness of its gradient flow

For the proof of the uniqueness we will follow the argumentation of [4, Section 5] and [7, Section 3].

We start by introducing two new concepts, the push-forward via a plan, and plans with bounded deformation. We prove different interesting auxiliary properties which correlate these two new concepts with the Entropy functional and its descending slope, and the key result in Proposition 4.22. We conclude the section proving the uniqueness of the gradient flow of the Entropy.

4.4.1. Plans with bounded deformation and push-forward via a plan

We extend the notion of push-forward via a map as follows.

Definition 4.13 (Push-forward via a plan). *Let $\mu \in \mathcal{P}(X), \gamma \in \mathcal{P}(X^2)$ be such that $\mu \ll \pi^1_*\gamma$. The measures $\gamma_\mu \in \mathcal{P}(X^2)$ and $\gamma_*\mu \in \mathcal{P}(X)$ are defined as*

$$\begin{aligned}\gamma_\mu(dx,dy) &:= \frac{d\mu}{d\pi^1_*\gamma}(x)\gamma(dx,dy),\\ \gamma_*\mu(dy) &:= \pi^2_*\gamma_\mu(dy).\end{aligned}$$

Remark 4.14. *Since $\mu \ll \pi^1_*\gamma$, there exists $f \in L^1(\pi^1_*\gamma)$ such that $\mu = f\pi^1_*\gamma$. By Disintegration Theorem, considering $\{\gamma_y\}_{y\in X}$ the disintegration of γ with respect to its second marginal we obtain*

$$\gamma_*\mu(dy) = \left(\int_X f(x)\, \gamma_y(dx)\right) \pi^2_*\gamma(dy). \tag{11}$$

The plans which are particularly useful in our analysis belong to the following category.

Definition 4.15. *The plan* $\gamma \in \mathcal{P}(X^2)$ *has* bounded deformation *if* $\exists c \in \mathbb{R}^+$ *such that* $\frac{1}{c}m \leq \pi^1_*\gamma, \pi^2_*\gamma \leq cm$.

We show now some preliminary properties. The following is a useful estimate.

Proposition 4.16. $\forall \mu, \nu \in \mathcal{P}(X), \forall \gamma \in \mathcal{P}(X^2)$ *such that* $\mu, \nu \ll \pi^1_*\gamma$,

$$\mathrm{Ent}_{\gamma_*\nu}(\gamma_*\mu) \leq \mathrm{Ent}_\nu(\mu).$$

Proof. We assume $\mu \ll \nu$, otherwise $\mathrm{Ent}_\nu(\mu) = \infty$ and there is nothing to prove. Then there exists $f \in L^1(\nu)$ such that $\mu = f\nu$ and since $\nu \ll \pi^1_*\gamma$ by hypothesis, $\exists \theta \in L^1(\pi^1_*\theta)$ such that $\nu = \theta\pi^1_*\gamma$. Disintegrating $\gamma_*\nu, \gamma_*\mu$ as in (11) we obtain

$$\gamma_*\mu = \left(\int_X f(x)\theta(x)\ \gamma_y(dx)\right)\pi^2_*\gamma, \tag{12}$$

$$\gamma_*\nu = \left(\int_X \theta(x)\ \gamma_y(dx)\right)\pi^2_*\gamma. \tag{13}$$

It is easily verified that $\gamma_*\mu \ll \gamma_*\nu$. Therefore by Radon-Nikodym Theorem there exists $\eta \in L^1(\gamma_*\mu)$ such that $\gamma_*\mu = \eta\gamma_*\nu$, and considering (12), (13), we have

$$\eta(y) = \frac{\int f\theta\ \gamma_y(dx)}{\int \theta\ \gamma_y(dx)} = \int f \frac{\theta}{\int \theta\ \gamma_y(dx)}\ \gamma_y(dx). \tag{14}$$

Defining

$$\tilde{\gamma} := (\theta \circ \pi^1)\gamma,$$

its disintegration with respect to its second marginal $\gamma_*\nu$ is

$$\tilde{\gamma}_y = \frac{\theta}{\int \theta\ \gamma_y(dx)}\gamma_y,$$

so we can rewrite (14) as

$$\eta(y) = \int f\ \tilde{\gamma}_y(dx).$$

Now let $u(z) := z\log z$. From the convexity of $u(z)$ and Jensen's inequality,

$$u(\eta(y)) \leq \int u(f(x))\ \tilde{\gamma}_y(dx).$$

Integrating both sides with respect to $\gamma_*\nu$ we get

$$\mathrm{Ent}_{\gamma_*\nu}(\gamma_*\mu) = \int u(\eta(y))\ \gamma_*\nu(dy) \leq \int \left(\int u(f(x))\ \tilde{\gamma}_y(dx)\right)\ \gamma_*\nu(dy),$$

and from Disintegration Theorem,

$$\int \left(\int u(f(x))\, \tilde{\gamma}_y(dx) \right) \gamma_* \nu(dy) = \int u(f(x))\, \nu(dx) = \mathrm{Ent}_\nu(\mu).$$

$\square$

The following formula will be useful in the proof of the next proposition.

LEMMA 4.17. *If $\mu, \nu, \sigma \in \mathcal{P}(X)$ and σ is such that there exists $c > 0 : \frac{1}{c}\nu \leq \sigma \leq c\nu$, then it holds*

$$\mathrm{Ent}_\nu(\mu) = \mathrm{Ent}_\sigma(\mu) + \int_X \log\left(\frac{d\sigma}{d\nu}\right) \mu(dx). \tag{15}$$

Proof. From the hypothesis on σ we can deduce there exists $\frac{1}{c} \leq g \leq c$ such that $\sigma = g\nu$. If μ is not absolutely continuous with respect to ν we obtain $\infty = \infty + C$ and (15) holds.

Otherwise if $\mu \ll \nu$ take $\mu = f\nu$; therefore

$$\mu = \frac{f}{g}\sigma$$

and

$$\mathrm{Ent}_\sigma(\mu) + \int_X \log\left(\frac{d\sigma}{d\nu}\right) \mu = \int_X f \log f\ \nu = \mathrm{Ent}_\nu(\mu).$$

$\square$

The push-forward of a measure via a plan with bounded deformation allows us to remain in the domain of the Entropy, as it is proved in the next results.

PROPOSITION 4.18. *If $\mu \in D(\mathrm{Ent}_m)$ and $\gamma \in \mathcal{P}(X^2)$ has bounded deformation, then $\gamma_*\mu \in D(\mathrm{Ent}_m)$.*

Proof. Since γ has bounded deformation there exist $c, C > 0$ such that $cm \leq \pi^1_*\gamma, \pi^2_*\gamma \leq Cm$. Using identity (15) we obtain

$$\begin{aligned} \mathrm{Ent}_m(\gamma_*\mu) = &\ \mathrm{Ent}_{\pi^2_*\gamma}(\gamma_*\mu) + \int_X \log\left(\frac{d\pi^2_*\gamma}{dm}\right) \gamma_*\mu \\ \leq &\ \mathrm{Ent}_{\pi^2_*\gamma}(\gamma_*\mu) + \log(C). \end{aligned}$$

From the fact that $\gamma_*(\pi^1_*\gamma) = \pi^2_*(\gamma_{\pi^1_*\gamma}) = \pi^2_*(\gamma)$ and Proposition 4.16,

$$\mathrm{Ent}_{\pi^2_*\gamma}(\gamma_*\mu) = \mathrm{Ent}_{\gamma_*(\pi^1_*\gamma)}(\gamma_*\mu) \leq \mathrm{Ent}_{\pi^1_*\gamma}(\mu).$$

Then using again identity (15)

$$\begin{aligned}\operatorname{Ent}_{\pi^1_*\gamma}(\mu) &= \operatorname{Ent}_m(\mu) + \int_X \log\left(\frac{dm}{d\pi^1_*\gamma}\right)\,\mu \\ &\leq \operatorname{Ent}_m(\mu) + \mu(X)\log\frac{1}{c} = \operatorname{Ent}_m(\mu) - \log c.\end{aligned}$$

In conclusion

$$\operatorname{Ent}_m(\gamma_*\mu) \leq \operatorname{Ent}_m(\mu) - \log c + \log C < \infty.$$

□

The following proposition gives a quite unexpected property of convexity of the Entropy.

Proposition 4.19. *If $\gamma \in \mathcal{P}(X^2)$ has bounded deformation then the map*

$$D(\operatorname{Ent}_m) \ni \mu \mapsto \operatorname{Ent}_m(\mu) - \operatorname{Ent}_m(\gamma_*\mu)$$

is convex with respect to linear interpolation of measures

Proof. Let $\mu_0 = f_0 m, \mu_1 = f_1 m$. Define for every $t \in (0,1)$

$$\begin{aligned}\mu_t &:= (1-t)\mu_0 + t\mu_1,\\ f_t &:= (1-t)f_0 + tf_1.\end{aligned}$$

We compute

$$\begin{aligned}&(1-t)\operatorname{Ent}_{\mu_t}(\mu_0) + t\operatorname{Ent}_{\mu_t}(\mu_1)\\ &= (1-t)\int_X \frac{f_0}{f_t}\log\left(\frac{f_0}{f_t}\right)\,\mu_t + t\int_X \frac{f_1}{f_t}\log\left(\frac{f_1}{f_t}\right)\,\mu_t\\ &= (1-t)\int_X f_0\log f_0\, m + t\int_X f_1 \log f_1\, m - \int_X f_t\log f_t\, m\\ &= (1-t)\operatorname{Ent}_m(\mu_0) + t\operatorname{Ent}_m(\mu_1) - \operatorname{Ent}_m(\mu_t).\end{aligned}$$

Since $\mu_i \in D(\operatorname{Ent})$ (for $i = 1,2$) and γ has bounded deformation, from Proposition 4.18 also $\gamma_*\mu_i \in D(\operatorname{Ent})$, so an identical argument with μ_t replaced by $\gamma_*\mu_t$ shows that

$$\begin{aligned}&(1-t)\operatorname{Ent}_{\gamma_*\mu_t}(\gamma_*\mu_0) + t\operatorname{Ent}_{\gamma_*\mu_t}(\gamma_*\mu_1)\\ &= (1-t)\operatorname{Ent}_m(\gamma_*\mu_0) + t\operatorname{Ent}_m(\gamma_*\mu_1) - \operatorname{Ent}_m(\gamma_*\mu_t).\end{aligned}$$

By Proposition 4.16 we have $\operatorname{Ent}_{\gamma_*\mu_t}(\gamma_*\mu_i) \leq \operatorname{Ent}_{\mu_t}(\mu_i)$ for $i = 1,2$, therefore

$$\begin{aligned}&(1-t)\operatorname{Ent}_m(\gamma_*\mu_0) + t\operatorname{Ent}_m(\gamma_*\mu_1) - \operatorname{Ent}_m(\gamma_*\mu_t)\\ &\leq (1-t)\operatorname{Ent}_m(\mu_0) + t\operatorname{Ent}_m(\mu_1) - \operatorname{Ent}_m(\mu_t).\end{aligned}$$

Rearranging the terms we finally obtain

$$\begin{aligned}&\operatorname{Ent}_m(\mu_t) - \operatorname{Ent}_m(\gamma_*\mu_t)\\ &\le (1-t)\operatorname{Ent}_m(\mu_0) + t\operatorname{Ent}_m(\mu_1) - (1-t)\operatorname{Ent}_m(\gamma_*\mu_0) - t\operatorname{Ent}_m(\gamma_*\mu_1).\end{aligned}$$

□

4.4.2. Approximability in Entropy and distance

The following is a technical result which allows us to control the Entropy of a perturbation of a measure.

For $\gamma \in \mathcal{P}(X^2)$, define the *transportation cost*

$$C(\gamma) := \int \mathfrak{d}^2(x,y)\,\gamma(dx,dy).$$

LEMMA 4.20. *If $\mu,\nu \in D(\operatorname{Ent}_m)$, there exists a sequence $(\gamma^n)_n$ of plans with bounded deformation such that $\operatorname{Ent}_m(\gamma^n_*\mu) \to \operatorname{Ent}_m(\nu)$ and $C(\gamma^n_\mu) \to W_2^2(\mu,\nu)$ as $n \to \infty$.*

Proof. Let $f, g \in L^1(m)$ be non-negative such that $\mu = fm, \nu = gm$. Pick $\gamma \in \operatorname{Adm}(\mu,\nu)$ s.t.

$$\int \mathfrak{d}^2\,\gamma = \inf_{\gamma' \in \operatorname{Adm}(\mu,\nu)} \int \mathfrak{d}^2\,\gamma'$$

and $\forall n \in \mathbb{N}$ define

$$\begin{aligned}A'_n &:= \Big\{(x,y) \in X^2 : f(x) + g(y) \le n\Big\},\\ A_n &:= \Big\{(x,y) \in A'_n : \gamma_x(A'_n) > \tfrac{1}{2}\Big\},\\ \gamma^n(dx,dy) &:= c_n\left(\gamma_{|A_n}(dx,dy) + \frac{1}{n}(id,id)_* m(dx,dy)\right),\end{aligned}$$

with c_n the normalization constant, i.e. $c_n = \dfrac{1}{\gamma(A_n) + \frac{1}{n}}$. Disintegrating γ we obtain

$$\begin{aligned}\gamma(dx,dy) &= \int_X \big[\gamma_x(dy)\big]\,\mu(dx) = \int_X \big[\gamma_y(dx)\big]\,\nu(dy),\\ \gamma_{|A_n}(dx,dy) &= \int_X \big[\gamma_{x\,|A_n}(dy)\big]\,\mu(dx) = \int_X \big[\gamma_{y\,|A_n}(dx)\big]\,\nu(dy).\end{aligned}$$

Therefore

$$\begin{aligned}\frac{\gamma^n}{c_n} &= \left(\int_X \big[\gamma_{x\,|A_n}(dy)\big]\,\mu(dx)\right) + \frac{1}{n}(id,id)_* m(dx,dy)\\ &= \int_X \left[f(x)\gamma_{x\,|A_n}(dy) + \frac{1}{n}\delta_x(dy)\right]\,m(dx),\end{aligned}$$

and analogously for ν,

$$\frac{\gamma^n}{c_n} = \int_X \left[g(y)\gamma_{y\,|A_n}(dx) + \frac{1}{n}\delta_y(dx) \right] \; m(dy).$$

Then the marginals of γ^n are

$$\pi^1_*\gamma^n = c_n \left(\gamma_x(A_n) f(x) + \frac{1}{n} \right) m(dx),$$
$$\pi^2_*\gamma^n = c_n \left(\gamma_y(A_n) g(y) + \frac{1}{n} \right) m(dy).$$

Since $0 \le f, g \le n$ and $0 \le \gamma_x(A_n), \gamma_y(A_n) \le 1$,

$$\frac{c_n}{n} m \le \pi^1_*\gamma^n, \pi^2_*\gamma^n \le \left(nc_n + \frac{c_n}{n} \right) \; m,$$

i.e. γ^n has bounded deformation for every n.

By Radon-Nikodym Theorem

$$f_n(x) := \frac{d\mu}{d\pi^1_*\gamma^n} = \frac{f(x)}{c_n(\gamma_x(A_n)f(x) + \frac{1}{n})}, \tag{16}$$

so by definition of push-forward of a measure

$$\begin{aligned} \gamma^n_\mu(dx, dy) &= \frac{d\mu}{d\pi^1_*\gamma^n}\gamma^n = f_n(x)\gamma^n(dx, dy) \\ &= \int_X c_n \left[f_n(x) g(y) \gamma_{y\,|A_n}(dx) + \frac{f_n(x)}{n}\delta_y(dx) \right] \; m(dy), \end{aligned}$$

and thus

$$\gamma^n_*\mu = \pi^2_*\gamma^n_\mu(dy) = \int_X c_n \left(g(y) \int_{A_n} f_n(x) \; \gamma_y(dx) + \frac{f_n(y)}{n} \right) \; m(dy).$$

Defining

$$h_n(y) := g(y) \left(\int_{A_n} f_n(x) \; \gamma_y(dx) \right) + \frac{f_n(y)}{n},$$

we can write

$$\mathrm{Ent}_m(\gamma^n_*\mu) = \int_X c_n h_n(y) \log \left[c_n h_n(y) \right] \; m(dy).$$

We notice that since $c_n \to 1$, $c_n \ge 2/3$ definitely, so

$$c_n\gamma_x(A_n) + \frac{1}{nf(x)} \ge \frac{1}{2}c_n + \frac{1}{n^2} \ge \frac{1}{3},$$

thus by definition (16), $f_n \leq 3$ definitely.
Therefore, defined

$$q_n(y) := 3\left(g(y) + \frac{1}{n}\right),$$

we have that $0 \leq h_n \leq q_n(y)$ definitely.

A calculation shows that $q_n(y)\log q_n(y) \in L^1(m)$ definitely; thus by Dominated Convergence Theorem

$$\begin{aligned}\lim_n \mathrm{Ent}_m(\gamma^n_* \mu) &= \int_X \lim_n \Big(c_n h_n(y) \log\big(c_n h_n(y)\big)\Big)\, m(dy) \\ &= \int_X g(y)\log g(y)\, m(dy) = \mathrm{Ent}_m(\nu),\end{aligned}$$

since $\lim_n c_n = 1$ and $\lim_n h_n(y) = g(y)$.

We pass to show the convergence of the cost. We can rewrite the cost of γ^n_μ as

$$\begin{aligned}C(\gamma^n_\mu) = \int_{X^2} \mathfrak{d}^2(x,y)\, \gamma^n_\mu(dx,dy) &= \int_{X^2} \mathfrak{d}^2(x,y) f_n\, \gamma^n(dx,dy) \\ = \int_{X^2} \mathfrak{d}^2 c_n f_n \chi_{A_n}\, \gamma &+ \frac{1}{n}\int_{X^2} \mathfrak{d}^2 f_n\, (id,id)_* m.\end{aligned}$$

By hypothesis X is compact and we have that $c_n, f_n, \chi_{A_n} \to 1$, therefore there exists $k > 1$ such that $\mathfrak{d}^2 c_n f_n \chi_{A_n} \leq k$ definitely. In conclusion, by Dominated Convergence

$$C(\gamma^n_\mu) \to \int_{X^2} \mathfrak{d}^2\, \gamma = W_2^2(\mu,\nu).$$

□

4.4.3. Convexity of the squared descending slope

If $(X, \mathfrak{d}, m)$ has Ricci curvature bounded from below by K, from (4) we know that

$$|D^- \mathrm{Ent}_m|(\mu) = \sup_{\nu \in \mathcal{P}(X),\, \nu \neq \mu} \frac{\left(\mathrm{Ent}_m(\mu) - \mathrm{Ent}_m(\nu) + \frac{K}{2} W_2^2(\mu,\nu)\right)^+}{W_2(\mu,\nu)}.$$

We give yet another characterization of $|D^- \mathrm{Ent}_m|$, which relies only on plans with bounded deformation and which we will use in the proof of Proposition 4.22.

LEMMA 4.21. *If $(X, \mathfrak{d}, m)$ has Ricci curvature bounded from below by K then*

$$|D^- \operatorname{Ent}_m|(\mu) = \sup_\gamma \frac{\left(\operatorname{Ent}_m(\mu) - \operatorname{Ent}_m(\gamma_*\mu) + \frac{K}{2}C(\gamma_\mu)\right)^+}{(C(\gamma_\mu))^{1/2}},$$

where the supremum is taken among all $\gamma \in \operatorname{Adm}(\mu, \nu)$ with bounded deformation, and if $C(\gamma_\mu) = 0$ the right hand side is taken 0 by definition.

Proof. We show both inequalities.

" $\geq$ ". We can assume $C(\gamma_\mu) > 0$, $\nu = \gamma_*\mu$ and $K < 0$ (thanks to Remark 3.8).
The following inequality is easily proven: if $a, b, c \in \mathbb{R}$ and $0 < b \leq c$, then

$$\frac{(a-b)^+}{\sqrt{b}} \geq \frac{(a-c)^+}{\sqrt{c}}.$$

Substituting

$$\begin{aligned} a :=& \operatorname{Ent}_m(\mu) - \operatorname{Ent}_m(\gamma_*\mu), \\ b :=& -\frac{K}{2}W_2^2(\mu, \gamma_*\mu), \\ c :=& -\frac{K}{2}C(\gamma_\mu), \end{aligned}$$

proves the thesis.

" $\leq$ ". It comes directly from Lemma 4.20.

□

A key ingredient in proving the uniqueness of the flow of the Entropy is the convexity of the squared descending slope of the Entropy, which we now show.

PROPOSITION 4.22. *If (X, d, m) has Ricci curvature bounded from below by K, then the map*

$$\mu \in D(\operatorname{Ent}_m) \mapsto |D^- \operatorname{Ent}_m|^2(\mu)$$

is convex with respect to linear interpolation of measures.

Proof. Recalling that the supremum of convex maps is still convex, and considering Lemma 4.21, we are done if we prove that the map

$$\mu \mapsto \frac{\left(\left(\operatorname{Ent}_m(\mu) - \operatorname{Ent}_m(\gamma_*\mu) + \frac{K^-}{2}C(\gamma_\mu)\right)^+\right)^2}{C(\gamma_\mu)} \tag{17}$$

is convex.
The map

$$\mu \in D(\mathrm{Ent}_m) \mapsto C(\gamma_\mu) = \int_{X\times X} d^2(x,y)\, d\gamma_\mu$$

is linear. Hence, together with the fact that $\mu \mapsto \mathrm{Ent}_m(\mu) - \mathrm{Ent}_m(\gamma_*\mu)$ is convex (Proposition 4.19), also

$$\mu \mapsto \mathrm{Ent}_m(\mu) - \mathrm{Ent}_m(\gamma_*\mu) - \frac{K^-}{2} C(\gamma_\mu)$$

is convex. Taking its positive part we still have a convex function.

Now take

$$a(\mu) := \Big(\mathrm{Ent}_m(\mu) - \mathrm{Ent}_m(\gamma_*\mu) - \frac{K^-}{2} C(\gamma_\mu)\Big)^+,\ b(\mu) := C(\gamma_\mu),$$

and define also $\psi : [0,\infty) \times [0,\infty) \to \mathbb{R} \cup \{\infty\}$ as

$$\psi(a,b) := \begin{cases} \dfrac{a^2}{b} & \text{if } b > 0, \\ \infty & \text{if } b = 0,\ a > 0, \\ 0 & \text{if } a = b = 0. \end{cases}$$

It is immediately shown that ψ is convex and it is non-decreasing with respect to a. Therefore we obtain

$$\begin{aligned} &\psi\Big(a\big((1-t)\mu_0 + t\mu_1\big), b\big((1-t)\mu_0 + t\mu_1\big)\Big) \\ &\le \psi\Big((1-t)a(\mu_0) + ta(\mu_1), (1-t)b(\mu_0) + tb(\mu_1)\Big) \\ &\le (1-t)\psi\Big(a(\mu_0), b(\mu_0)\Big) + t\psi\Big(a(\mu_1), b(\mu_1)\Big), \end{aligned}$$

thus the convexity of (17). □

We finally have all the tools to prove the uniqueness of the gradient flow generated by the Entropy.

THEOREM 4.23 (Uniqueness). *Let (X,d,m) have Ricci curvature bounded from below by K and let $\tilde\mu \in D(\mathrm{Ent}_m)$; then there exists a unique gradient flow of* Ent_m *in $(\mathcal{P}(X), W_2)$ starting from $\tilde\mu$.*

Proof. Let μ_t, ν_t be gradient flows of Ent_m starting both from $\tilde\mu$. Then

$$\eta_t := \frac{\mu_t + \nu_t}{2}$$

is an absolutely continuous curve (by Remark 4.3) starting from $\tilde{\mu}$. From the definition of gradient flow,

$$\mathrm{Ent}_m(\tilde{\mu}) = \mathrm{Ent}_m(\mu_t) + \frac{1}{2}\int_0^t |\dot{\mu}_s|^2\, ds + \frac{1}{2}\int_0^t |D^- \mathrm{Ent}_m|^2(\mu_s)\, ds,$$

$$\mathrm{Ent}_m(\tilde{\mu}) = \mathrm{Ent}_m(\nu_t) + \frac{1}{2}\int_0^t |\dot{\nu}_s|^2\, ds + \frac{1}{2}\int_0^t |D^- \mathrm{Ent}_m|^2(\nu_s)\, ds,$$

for every $t \geq 0$.

Adding up these two equalities, by the squared slope convexity (Proposition 4.22), the squared metric speed convexity (Remark 4.3) and the strict convexity of the relative Entropy (Proposition 4.7), we obtain

$$\mathrm{Ent}_m(\tilde{\mu}) > \mathrm{Ent}_m(\eta_t) + \frac{1}{2}\int_0^t |\dot{\eta}_s|^2\, ds + \frac{1}{2}\int_0^t |D^- \mathrm{Ent}_m|^2(\eta_s)\, ds$$

for every t where $\mu_t \neq \nu_t$. But this contradicts (6); therefore it must be $\mu_t \equiv \nu_t$. □

References

[1] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, 2000.

[2] L. Ambrosio and N. Gigli, *A users guide to optimal transport*, Lecture notes of C.I.M.E. summer school, 2009.

[3] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, second ed., Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008.

[4] L. Ambrosio, N. Gigli, and G. Savaré, *Heat flow and calculus on metric measure spaces with Ricci curvature bounded below—the compact case*, Boll. Unione Mat. Ital. (9) **5** (2012), no. 3, 575–629.

[5] E. De Giorgi, A. Marino, and M. Tosques, *Problems of evolution in metric spaces and maximal decreasing curve*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8) **68** (1980), no. 3, 180–187.

[6] D. H. Fremlin, *Measure theory*, vol. 4, Torres Fremlin, 2000.

[7] N. Gigli, *On the heat flow on metric measure spaces: existence, uniqueness and stability*, Calc. Var. Partial Differential Equations **39** (2010), no. 1-2, 101–120.

[8] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal. **29** (1998), no. 1, 1–17.

[9] J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. (2) **169** (2009), no. 3, 903–991.

[10] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations **26** (2001), no. 1-2, 101–174.

[11] K.-T. Sturm, *On the geometry of metric measure spaces. I-II*, Acta Math. **196** (2006), no. 1, 65–177.

Authors' addresses:

Stefano Bianchini and Alexander Dabrowski
SISSA
Via Bonomea 265
34136 Trieste, Italy
E-mail: bianchin@sissa.it, adabrow@sissa.it

Received March 19, 2014

Rend. Istit. Mat. Univ. Trieste
Volume 46 (2014), 71–99

Verbal functions of a group

Daniele Toller

Abstract. *The aim of this paper is the study of elementary algebraic subsets of a group G, first defined by Markov in 1944 as the solution-set of a one-variable equation over G. We introduce the group of words over G, and the notion of verbal function of G in order to better describe the family of elementary algebraic subsets. The intersections of finite unions of elementary algebraic subsets are called algebraic subsets of G, and form the family of closed sets of the Zariski topology $\mathfrak{Z}_G$ on G. Considering only some elementary algebraic subsets, one can similarly introduce easier-to-deal-with topologies $\mathfrak{T} \subseteq \mathfrak{Z}_G$, that nicely approximate $\mathfrak{Z}_G$ and often coincide with it.*

Keywords: group of words, universal word, verbal function, (elementary, additively) algebraic subset, (partial) Zariski topology, centralizer topology, quasi-topological group topology.
MS Classification 2010: primary 54E15, 20F70; secondary: 57M07, 20K45.

1. Introduction

In 1944, Markov [20] introduced three special families of subsets of a group G, calling a subset $X \subseteq G$:

(a) *elementary algebraic* if there exist an integer $n > 0$, elements $g_1, \dots, g_n \in G$ and $\varepsilon_1, \dots, \varepsilon_n \in \{-1, 1\}$, such that

$$X = \{x \in G : g_1 x^{\varepsilon_1} g_2 x^{\varepsilon_2} \cdots g_n x^{\varepsilon_n} = e_G\}; \tag{1}$$

(b) *additively algebraic* if X is a finite union of elementary algebraic subsets of G;

(c) *algebraic* if X is an intersection of additively algebraic subsets of G.

If G is a group, take x as a symbol for a variable, and denote by $G[x] = G * \langle x \rangle$ the free product of G and the infinite cyclic group $\langle x \rangle$ generated by x. We call $G[x]$ *the group of words with coefficients in* G, and its elements w are called *words in* G. An element $w \in G[x]$ has the form

$$w = g_1 x^{\varepsilon_1} g_2 x^{\varepsilon_2} \cdots g_n x^{\varepsilon_n}, \tag{2}$$

for an integer $n \geq 0$, elements $g_1, \ldots, g_n \in G$ and $\varepsilon_1, \ldots, \varepsilon_n \in \{-1, 1\}$. The group $G[x]$ is defined also via a universal property in Fact 2.1, and explicitly described in §2.2. Then an elementary algebraic subset X of G as in (1) will be denoted by E_w^G (or simply, E_w), where $w \in G[x]$ as in (2) is its defining word considered as an element of $G[x]$.

In particular, a word $w \in G[x]$ determines its associated evaluation function $f_w \colon G \to G$, mapping $g \mapsto w(g)$, where $w(g) \in G$ is obtained replacing x with g in (2) and taking products in G. We call *verbal* a function $G \to G$ of the form f_w. Let us immediately show that some very natural functions $G \to G$ are verbal.

Example 1.1. 1. If $g \in G$, then one can consider the word $w = g \in G[x]$, so that f_w is the constant function g on G.

2. The identity map of G is the function $f_x \colon g \mapsto g$.
3. The inversion function of G is $f_{x^{-1}} \colon g \mapsto g^{-1}$.
4. More generally, for every integer $n \in \mathbb{Z}$, the word $x^n \in G[x]$ determines the verbal function $f_{x^n} \colon g \mapsto g^n$.
5. The left translation in G by an element $a \in G$ is the function $f_{ax} \colon g \mapsto ag$, and the right translation is the function $f_{xa} \colon g \mapsto ga$.
6. For an element $a \in G$, the word $w = axa^{-1}$ determines the conjugation by a, as $f_w \colon g \mapsto aga^{-1}$.

In §3.2 we equip the set $\mathscr{F}(G)$ of verbal functions of G with the pointwise product operation, making $\mathscr{F}(G)$ a group. The surjection $G[x] \to \mathscr{F}(G)$, mapping $w \mapsto f_w$, shows that $\mathscr{F}(G)$ is isomorphic to a quotient of $G[x]$.

We dedicate §3.3 to *monomials*, namely words of the form $w = gx^m \in G[x]$, for $g \in G$ and $m \in \mathbb{Z}$. In the final §3.4 we consider abelian groups. We note that if G is abelian, and $f \in \mathscr{F}(G)$, then $f = f_w$ for a monomial $w \in G[x]$. Then we describe $\mathscr{F}(G)$ in Proposition 3.7.

In §4 we study the elementary algebraic subsets, which we redefine using verbal functions (Definition 4.1) as the subsets of the form

$$E_w = f_w^{-1}(\{e_G\}),$$

for $w \in G[x]$. In this sense one can consider E_w as the solution-set of the equation $w(x) = e_G$ in G.

Example 1.2. 1. For an element $g \in G$, let $w = g^{-1}x \in G[x]$. Then $f_w \colon G \to G$ is the left translation by g^{-1} by Example 1.1, item 5, and $E_w = \{g\}$.

This shows that every singleton is an elementary algebraic subset of G.

2. If $g \in G$, then the centralizer
$$C_G(g) = \{h \in G \mid gh = hg\}$$
coincides with E_w, where $w = gxg^{-1}x^{-1} \in G[x]$ (see also Example 3.2). Hence the centralizer $C_G(g)$ is an elementary algebraic subset of G. Therefore, the centralizer $C_G(S) = \bigcap_{g\in S} C_G(g)$ of any subset S of G is an algebraic subset. In particular, the center $Z(G) = C_G(G)$ is an algebraic subset.

3. By Example 1.1, item 4, for every $n \in \mathbb{N}$ the word $x^n \in G[x]$ determines the verbal function $f_{x^n} \colon g \mapsto g^n$. Hence, $E_w = G[n]$ by definition, where
$$G[n] = \{g \in G \mid g^n = e_G\}.$$
If G is abelian, then $G[n]$ is a subgroup of G, called the n-socle of G. In the abelian case, these subsets (together with their cosets, of course) are all the non-empty elementary algebraic subsets of G (see (11)).

Then we see that the family of elementary algebraic subsets of G is stable under taking inverse image under verbal functions (Lemma 4.4). As a consequence, the translate of an elementary algebraic subset is an elementary algebraic subset (Example 4.5).

The family of algebraic subsets is stable under taking intersections and finite unions, and contains every finite subset (by Example 1.2, item 1), so is the family of closed sets of a T_1 topology $\mathfrak{Z}_G$ on G, the *Zariski topology of* G. As a matter of fact, Markov did not explicitly introduce such a topology, that was first explicitly defined by Bryant in [7], as the *verbal topology of* G. Here we keep the name Zariski topology, and the notation $\mathfrak{Z}_G$, for this topology, already used [3, 11, 12, 13, 15, 16].

In §5.2 and §5.3.1 we briefly also consider two other topologies on G, the *Markov topology* $\mathfrak{M}_G$ and the *precompact Markov topology* $\mathfrak{P}_G$, introduced respectively in [12] and [13] as the intersections
$$\mathfrak{M}_G = \bigcap\{\tau \mid \tau \text{ Hausdorff group topology on G}\},$$
$$\mathfrak{P}_G = \bigcap\{\tau \mid \tau \textit{ precompact } \text{Hausdorff group topology on G}\}.$$

The topology $\mathfrak{M}_G$ was only implicitly introduced by Markov in the same paper [20], via the notion of *unconditionally closed subset* of G, namely a subset of G that is closed with respect to every Hausdorff group topology on G. Of course, the family of unconditionally closed subsets of G is the family of the closed sets of $\mathfrak{M}_G$. It can be directly verified from the definitions that
$$\mathfrak{Z}_G \subseteq \mathfrak{M}_G \subseteq \mathfrak{P}_G. \tag{3}$$

Now we recall the definition of a *quasi-topological group.*

Definition 1.3. *Let G be a group, and τ a topology on G. The pair (G,τ) is called* quasi-topological group *if for every $a, b \in G$ the function $(G,\tau) \to (G,\tau)$, mapping $x \mapsto ax^{-1}b$, is continuous.*

Obviously, (G,τ) is a quasi-topological group if and only if the inversion function $f_{x^{-1}}$ and both the translations f_{gx} and f_{xg} are continuous, for every $g \in G$.

We dedicate §5 to quasi-topological groups, proving first some general results in §5.1. Then we use verbal functions in §5.2 to prove (3) and that all the pairs $(G,\mathfrak{Z}_G)$, $(G,\mathfrak{M}_G)$, $(G,\mathfrak{P}_G)$ are quasi-topological groups (Corollary 5.7).

By Example 1.2, item 2, the center $Z(G)$ of G is $\mathfrak{Z}_G$-closed. We extend this result proving that for every positive integer n also the n-th center $Z_n(G)$ is $\mathfrak{Z}_G$-closed in Corollary 5.10, as a consequence of Theorem 5.9.

In §5.3, inspired by [3, 4, 5], we introduce the notion of a *partial Zariski topology* $\mathfrak{T} \subseteq \mathfrak{Z}_G$ on G, namely a topology having *some* elementary algebraic subsets as a subbase for its closed sets. The aim of this definition is to study the cases when indeed the equality $\mathfrak{T} = \mathfrak{Z}_G$ holds for some partial Zariski topology $\mathfrak{T}$ on G, in order to have easier-to-deal-with subbases of $\mathfrak{Z}_G$.

For example, we introduce the *monomial topology* $\mathfrak{T}_{mon}$ on a group G in Definition 5.19, whose closed sets are generated by the subsets E_w, for monomials $w \in G[x]$. We note that $\mathfrak{T}_{mon} = \mathfrak{Z}_G$ when G is abelian (Example 5.20), and we prove that $\mathfrak{T}_{mon}$ is the cofinite topology when G is nilpotent, torsion free (Corollary 5.21).

In §5.3.1 we recall a recent result from [3]. If X is an infinite set, we denote by $S(X)$ the *symmetric group of* X, consisting of the permutations of X. If $\phi \in S(X)$, its *support* is the subset $\mathrm{supp}(\phi) = \{x \in X \mid \phi(x) \neq x\} \subseteq X$. We denote by $S_\omega(X)$ the subgroup of $S(X)$ consisting of the permutations having finite support. If G is a group with $S_\omega(X) \leq G \leq S(X)$, let $\tau_p(G)$ denote the *point-wise convergence topology of* G. The authors of [3] have introduced a partial Zariski topology $\mathfrak{Z}'_G$ on G and proved that $\mathfrak{Z}'_G = \mathfrak{Z}_G = \mathfrak{M}_G = \tau_p(G)$ (see Theorem 5.26).

We dedicate §5.4 to two others partial Zariski topologies, the *centralizer topologies* $\mathfrak{C}_G$ and $\mathfrak{C}'_G$, defined as follows. Let $\mathcal{C} = \{gC_G(a) \mid a, g \in G\}$, and note that its members are elementary algebraic subsets by Example 1.2, item 2, and Example 4.5. Then $\mathfrak{C}_G$ is the topology having $\mathcal{C}$ as a subbase for its closed sets (Definition 5.15). If $\mathcal{C}' = \mathcal{C} \cup \{\{g\} \mid g \in G\}$, we similarly introduce $\mathfrak{C}'_G$ as the topology having $\mathcal{C}'$ as a subbase for its closed sets (Definition 5.28). See Proposition 5.30 for the first few properties of $\mathfrak{C}_G$ and $\mathfrak{C}'_G$.

Finally, we see in §5.4.1 that every free non-abelian group F satisfies $\mathfrak{C}_F = \mathfrak{C}'_F = \mathfrak{Z}_F$. On the other hand, we consider a class of matrix groups H in §5.4.2, satisfying $\mathfrak{C}_H \neq \mathfrak{C}'_H = \mathfrak{Z}_H$.

This paper is a part of articles dedicated to the study of the Zariski topology of a group G, using the group of words $G[x]$ and the group of verbal functions

$\mathscr{F}(G)$ of G as main tools, see [14, 15, 16, 17]. In particular, we develop the basic theory here.

We denote by $\mathbb{Z}$ the group of integers, by $\mathbb{N}_+$ the set of positive integers, and by $\mathbb{N}$ the set of naturals. If $n \in \mathbb{Z}$, the cyclic subgroup it generates is $n\mathbb{Z}$, while the quotient group $\mathbb{Z}/n\mathbb{Z}$ will be denoted by $\mathbb{Z}_n$.

If X is a set, and $\mathcal{B} \subseteq \mathcal{P}(X)$ is a family of subsets of X, we denote by $\mathcal{B}^\cup \subseteq \mathcal{P}(X)$ the family of finite unions of members of $\mathcal{B}$.

2. The group of words $G[x]$

2.1. The categorical aspect of $G[x]$

The group $G[x]$ is determined by the universal property stated below.

Fact 2.1. *Let G be a group. Then there exist a unique (up to isomorphism) group $G[x]$, together with an injective group homomorphism $i_G\colon G \to G[x]$, satisfying the following universal property:*
for every group Γ, for every group homomorphism $\phi\colon G \to \Gamma$, and for every $\gamma \in \Gamma$, there exists a unique group homomorphism $\widetilde{\phi}\colon G[x] \to \Gamma$ such that $\widetilde{\phi} \circ i_G = \phi$ and $\widetilde{\phi}(x) = \gamma$.

$$\begin{tikzcd} & G[x] \ni x \arrow[d, "\widetilde{\phi}"] \arrow[d, mapsto, shift left=6] \\ G \arrow[ur, hook, "i_G"] \arrow[r, "\phi"] & \Gamma \ni \gamma \end{tikzcd}$$

From now on, we will identify G with $i_G(G) \leq G[x]$.

In the following example we illustrate a few particular cases when Fact 2.1 can be applied.

Example 2.2. 1. Consider the identity map $\mathrm{id}_G\colon G \to G$. By Fact 2.1, for every $g \in G$ there exists a unique map $\mathrm{ev}_g\colon G[x] \to G$, with $\mathrm{ev}_g \restriction_G = \mathrm{id}_G$ and $\mathrm{ev}_g(x) = g$, that we call *evaluation map*. Then we define $w(g) = \mathrm{ev}_g(w)$ for every $w \in G[x]$.

$$\begin{tikzcd} & G[x] \ni x \arrow[d, "\mathrm{ev}_g"] \arrow[d, mapsto, shift left=6] \\ G \arrow[ur, "i_G"] \arrow[r, "\mathrm{id}_G"] & G \ni g \end{tikzcd}$$

2. A *G-endomorphism of* $G[x]$ is a group homomorphism $\phi\colon G[x] \to G[x]$ such that $\phi \circ i_G = i_G$, i.e. $\phi \restriction_G = \mathrm{id}_G$, so that the following diagram

commutes:

$$\begin{tikzcd}
& G[x] \arrow[d, "\phi"] \\
G \arrow[ur, "i_G"] \arrow[r, "i_G"] & G[x] \ .
\end{tikzcd}$$

Then ϕ is uniquely determined by the element $w = \phi(x) \in G[x]$, and now we show that every choice of $w \in G[x]$ can be made, thus classifying the G-endomorphisms of $G[x]$. To this end, consider the map $i_G \colon G \to G[x]$. By Fact 2.1, for every $w \in G[x]$ there exists a unique G-endomorphism $\xi_w \colon G[x] \to G[x]$, with $x \mapsto w$.

$$\begin{tikzcd}
& G[x] \arrow[d, "\xi_w"] & \ni x \arrow[d, mapsto] \\
G \arrow[ur, "i_G"] \arrow[r, "i_G"] & G[x] & \ni w
\end{tikzcd}$$

Proposition 2.3. *Let $\phi \colon G_1 \to G_2$ be a group homomorphism. Then there exists a unique group homomorphism $F \colon G_1[x] \to G_2[x]$ such that $F \restriction_{G_1} = \phi$, $F(x) = x$. In particular, if ϕ is surjective (resp., injective), then F is surjective (resp., injective).*

Moreover, the following hold.

1. *If $H \leq G$ is a subgroup of G, then $H[x] \leq G[x]$.*
2. *If $H \trianglelefteq G$ is a normal subgroup of G, and $\overline{G} = G/H$, then $\overline{G}[x]$ is a quotient of $G[x]$.*

Proof. Composing $\phi \colon G_1 \to G_2$ and the map $i_{G_2} \colon G_2 \to G_2[x]$, we obtain $\psi = i_{G_2} \circ \phi \colon G_1 \to G_2[x]$. Then apply Fact 2.1 and use the universal property of $G_1[x]$ to get $F = \widetilde{\psi} \colon G_1[x] \to G_2[x]$ such that $F(x) = x$ and $F \circ i_{G_1} = i_{G_2} \circ \phi$, i.e. $F \restriction_{G_1} = \phi$.

$$\begin{tikzcd}
& & G_1[x] \arrow[d, "\widetilde{\psi}=F"] & \ni x \arrow[d, mapsto] \\
G_1 \arrow[urr, bend left, "i_{G_1}"] \arrow[r, "\phi"] \arrow[rr, bend right, "\psi"'] & G_2 \arrow[r, "i_{G_2}"] & G_2[x] & \ni x
\end{tikzcd}$$

If ϕ is surjective, then F is surjective too, as $F(G_1[x])$ contains both x and $\phi(G_1) = G_2$, which generate $G_2[x]$.

In Remark 3.3, item 1, we will explicitly describe the map F, so that by (5) it will immediately follow that F is injective when ϕ is injective.

1. In this case, the injection $\phi\colon H \hookrightarrow G$ gives the injection $F\colon H[x] \hookrightarrow G[x]$.
2. The canonical projection $\phi\colon G \to \overline{G}$ gives the surjection $F\colon G[x] \to \overline{G}[x]$.

□

The following corollary immediately follows from Proposition 2.3.

COROLLARY 2.4. *The assignment $G \mapsto G[x]$, and the canonical embedding $G \xrightarrow{i_G} G[x]$, define a pointed endofunctor $\varpi\colon \mathbf{Gr} \to \mathbf{Gr}$ in the category of groups and group homomorphism. In other words, for every group homomorphism $\phi\colon G_1 \to G_2$, the following diagram commutes*

$$\begin{tikzcd} G_1 \arrow[r, "i_{G_1}"] \arrow[d, "\phi"'] & G_1[x] \arrow[d, "\varpi(\phi)"] \\ G_2 \arrow[r, "i_{G_2}"] & G_2[x]\,, \end{tikzcd}$$

where $\varpi(\phi) = F$ is the map given by Proposition 2.3.

2.2. The concrete form of $G[x]$

Here we recall the concrete definition of $G[x]$ in terms of products of the form (4) below that will be called *words*. In particular, if $g \in G$, then $w = g \in G[x]$ will be called *constant word*, and we define its *lenght* to be $\mathrm{l}(w) = 0 \in \mathbb{N}$. In the general case, for $w \in G[x]$ there exist $n \in \mathbb{N}$, elements $g_1, \ldots, g_n, g_0 \in G$ and $\varepsilon_1, \ldots, \varepsilon_n \in \{-1, 1\}$, such that

$$w = g_1 x^{\varepsilon_1} g_2 x^{\varepsilon_2} \cdots g_n x^{\varepsilon_n} g_0. \tag{4}$$

Notice that in Markov's definition (1) of elementary algebraic subset of G, he was assuming the defining word w not to be constant (see Remark 4.2 for more details).

If $g_i \neq e_G$ whenever $\varepsilon_{i-1} = -\varepsilon_i$ for $i = 2, \ldots, n$, we say that w is a *reduced word* in the free product $G[x] = G * \langle x \rangle$ and we define the *lenght* of w by $\mathrm{l}(w) = n$, where $n \in \mathbb{N}$ is the least natural such that w is as in (4).

DEFINITION 2.5. If $w \in G[x]$ is as in (4), we define the following notions.

- The *constant term of* w is $\mathrm{ct}(w) = w(e_G) = g_1 g_2 \cdots g_n g_0 \in G$;
- The *content* of w is $\epsilon(w) = \sum_{i=1}^{n} \varepsilon_i \in \mathbb{Z}$, which will also be denoted simply by ϵ when no confusion is possible.

If $w = g$, then we define $\epsilon(w) = 0$ and $\mathrm{ct}(w) = w(e_G) = g$. We call *singular* a word w such that $\epsilon(w) = 0$. All constant words are singular by definition.

Given two elements g, h of a group G, recall that their commutator element is $[g,h] = ghg^{-1}h^{-1} \in G$. Note that $[g,h] = e_G$ if and only if $gh = hg$, i.e. g and h commute. Then the *commutator subgroup* G' of G is the subgroup

$$G' = \langle [g,h] \mid g,h \in G \rangle$$

generated by all elements of G of the form $[g,h]$. It can easily verified that if $H \trianglelefteq G$ is a normal subgroup of G, then the quotient group G/H is abelian if and only if $G' \leq H$.

Both the functions $\mathrm{ct}\colon G[x] \to G$, mapping $w \mapsto \mathrm{ct}(w)$, and $\epsilon\colon G[x] \to \mathbb{Z}$, mapping $w \mapsto \epsilon(w)$, are surjective group homomorphisms. In particular, $\mathrm{ct}(G[x]') \leq G'$ and $\epsilon(G[x]') \leq \mathbb{Z}' = \{0\}$, so that $G[x]' \leq \mathrm{ct}^{-1}(G') \cap \ker(\epsilon)$. In the following theorem, we prove the reverse inclusion.

THEOREM 2.6. *For every group G, $G[x]' = \mathrm{ct}^{-1}(G') \cap \ker(\epsilon)$.*

Proof. Let $U = \mathrm{ct}^{-1}(G') \cap \ker(\epsilon) = \{w \in G[x] \mid \mathrm{ct}(w) \in G', \epsilon(w) = 0\}$. Then $G[x]' \subseteq U$ as we have noted above, and we prove the other inclusion by induction on $\mathrm{l}(w)$ for a word $w \in U$.

Let $w \in G[x]$ and assume $w \in U$. We first consider the case when $\mathrm{l}(w) = 0$, i.e. $w = \mathrm{ct}(w) \in G$ is a constant word, so that $w \in G' \leq G[x]'$ and there is nothing to prove. So now let $w \in U$ be as in (4), and note that $\epsilon(w) = 0$ implies that $n = \mathrm{l}(w) > 0$ is even, so that for the base case we have to consider $n = 2$. Then w has the form $w = g_1 x^{\varepsilon} g_2 x^{-\varepsilon} (g_1 g_2)^{-1} c$, with $c = \mathrm{ct}(w) \in G'$. Let $g = g_1 g_2$, and $w_0 = [g_2^{-1}, x^{\varepsilon}] \in G[x]'$, so that $w = g w_0 g^{-1} c = [g, w_0] w_0 c \in G[x]'$.

Now assume $n > 2$. As $\epsilon(w) = 0$, we have $\varepsilon_{i+1} = -\varepsilon_i$ for some $1 \leq i \leq n-1$. Then $w = w_1 w_2 w_3$ for the words

$$\begin{aligned} w_1 &= g_1 x^{\varepsilon_1} g_2 x^{\varepsilon_2} \cdots g_{i-1} x^{\varepsilon_{i-1}}, \\ w_2 &= g_i x^{\varepsilon_i} g_{i+1} x^{\varepsilon_{i+1}} (g_i g_{i+1})^{-1}, \\ w_3 &= (g_i g_{i+1}) g_{i+2} x^{\varepsilon_{i+2}} \cdots g_n x^{\varepsilon_n} g_0. \end{aligned}$$

As $w_2 \in G[x]'$ by the base case, and $w = [w_1, w_2] w_2 w_1 w_3$, we only have to show that $w_1 w_3 \in G[x]'$. As

$$\mathrm{ct}(w) = \mathrm{ct}(w_1)\,\mathrm{ct}(w_2)\,\mathrm{ct}(w_3) = \mathrm{ct}(w_1) e_G\,\mathrm{ct}(w_3) = \mathrm{ct}(w_1 w_3),$$

we have $\mathrm{ct}(w_1 w_3) \in G'$, and similarly $\epsilon(w_1 w_3) = 0$. Then $w_1 w_3 \in G[x]'$ by the inductive hypothesis. $\square$

3. Verbal functions

3.1. Definition and examples

DEFINITION 3.1. A word $w \in G[x]$ determines the associated evaluation function $f_w^G\colon G \to G$. We will often write f_w for f_w^G. We call *verbal function* of G

a function $G \to G$ of the form f_w, and we denote by $\mathscr{F}(G)$ the set of verbal functions on G.

If $w \in G[x]$ and $g \in G$, sometimes we also write $w(g)$ for the element $f_w(g) \in G$. So *a priori*, if f is a verbal function, then $f = f_w$ for a word $w \in G[x]$ as in (4).

Besides the basic examples already given in Example 1.1, we will also consider verbal functions of the following form.

EXAMPLE 3.2. If $\varepsilon \in \{\pm 1\}$, and $a \in G$, the word $w = [a, x^\varepsilon] = ax^\varepsilon a^{-1}x^{-\varepsilon} \in G[x]$ determines the verbal function $f_w \colon g \mapsto [a, g^\varepsilon]$. We will call *commutator verbal function* a function of this form.

Note that $f_w \colon G \to G$ is the only map such that $f_w \circ \mathrm{ev}_g = \mathrm{ev}_g \circ \xi_w$ for every $g \in G$, i.e. making the following diagram commute:

$$\begin{tikzcd} G[x] \arrow[r, "\xi_w"] \arrow[d, "\mathrm{ev}_g"'] & G[x] \arrow[d, "\mathrm{ev}_g"] \\ G \arrow[r, "f_w"] & G. \end{tikzcd}$$

REMARK 3.3. 1. Let $\phi \colon G_1 \to G_2$ be a group homomorphism, and

$$F = \varpi(\phi) \colon G_1[x] \to G_2[x]$$

be as in Proposition 2.3. If $w \in G[x]$ is as in (4), then

$$F \colon w \mapsto F(w) = \phi(g_1)x^{\varepsilon_1}\phi(g_2)x^{\varepsilon_2} \cdots \phi(g_n)x^{\varepsilon_n}\phi(g_0). \tag{5}$$

By (5), it immediately follows that F is injective when ϕ is injective.

Moreover, one can easily see that $\phi \circ f_w = f_{F(w)} \circ \phi$, i.e. the following diagram commutes:

$$\begin{tikzcd} G_1 \arrow[r, "f_w"] \arrow[d, "\phi"'] & G_1 \arrow[d, "\phi"] \\ G_2 \arrow[r, "f_{F(w)}"] & G_2\ . \end{tikzcd} \tag{6}$$

2. In particular, we will often consider the case when ϕ is the canonical projection $\pi \colon G \to G/N$, if N is a normal subgroup of G. In this case, let $\overline{G} = G/N$ be the quotient group, and for an element $g \in G$, let $\overline{g} = \pi(g) \in \overline{G}$. If $w = g_1x^{\varepsilon_1}g_2x^{\varepsilon_2} \cdots g_nx^{\varepsilon_n}g_0 \in G[x]$, let also $\overline{w} = F(w) = \overline{g}_1x^{\varepsilon_1}\overline{g}_2x^{\varepsilon_2} \cdots \overline{g}_nx^{\varepsilon_n}\overline{g}_0 \in \overline{G}[x]$. Then (6) (with $\phi = \pi$) gives $\pi \circ f_w = f_{\overline{w}} \circ \pi$.

3.2. Universal words

The group operation on $G[x]$ induces a group operation on $\mathscr{F}(G)$ as follows. If $w_1, w_2 \in G[x]$, let $w = w_1 w_2 \in G[x]$ be their product, and consider the verbal functions $f_{w_1}, f_{w_2}, f_w \in \mathscr{F}(G)$. Obviously, f_w is the pointwise product $f_{w_1} \cdot f_{w_2}$ of f_{w_1} and f_{w_2}, namely the map $f_w \colon g \mapsto f_{w_1}(g) f_{w_2}(g) = f_w(g)$. With this operation, $(\mathscr{F}(G), \cdot)$ is a group, with identity element the constant function $e_{\mathscr{F}(G)} \colon g \mapsto e_G$ for every $g \in G$. If w^{-1} is the inverse of $w \in G[x]$, then the inverse of $f_w \in \mathscr{F}(G)$ is $f_{w^{-1}}$, and will be denoted by $(f_w)^{-1}$.

For $S \subseteq G$, we denote by

$$(f_w)^{-1}(S) = \{(f_w)^{-1}(s) \mid s \in S\} = \{f_{w^{-1}}(s) \mid s \in S\} \subseteq G$$

the *image* of S under $(f_w)^{-1} = f_{w^{-1}}$, while $f_w^{-1}(S) = \{g \in G \mid f_w(g) \in S\}$ will denote the preimage of S under f_w.

Consider the surjective group homomorphism $\Phi_G \colon G[x] \to \mathscr{F}(G)$, $w \mapsto f_w$. Then $\mathscr{F}(G) \cong G[x]/\mathcal{U}_G$, where $\mathcal{U}_G$ is the kernel

$$\mathcal{U}_G = \ker(\Phi_G) = \{w \in G[x] \mid \forall g \in G \;\; f_w(g) = e_G\} \leq G[x]. \tag{7}$$

Definition 3.4. *If G is a group, and $w \in G[x]$, we say that w is a* universal word for G *if $w \in \mathcal{U}_G$.*

Note that a word $w \in G[x]$ is universal exactly when $E_w = G$.

Recall that the *exponent* $\exp(G)$ of a group G is the least common multiple, if it exists, of the orders of the elements of G. In this case, $\exp(G) > 0$. Otherwise, we conventionally define $\exp(G) = 0$. For example, every finite group G has positive exponent, and $\exp(G)$ divides $|G|$.

Example 3.5.

1. If $w \in \mathcal{U}_G$, then obviously $\mathrm{ct}(w) = f_w(e_G) = e_G$.
2. If G has $k = \exp(G) > 0$, then $w = x^k \in G[x]$ is a non-singular universal word for G, i.e. $f_w \equiv e_G$ is the constant function.

The singular universal words will play a prominent role, so we set

$$\mathcal{U}_G^{sing} = \{w \in \mathcal{U}_G : \epsilon(w) = 0\} = \mathcal{U}_G \cap \ker \epsilon \leq G[x].$$

In particular, also $\mathcal{U}_G^{sing}$ is a normal subgroup of $G[x]$.

Remark 3.6. Let $[G, \langle x \rangle] = \langle [g, x^i] \mid g \in G, i \in \mathbb{Z} \rangle \leq G[x]$ be the subgroup of $G[x]$ generated by all commutators $[g, x^i] \in G[x]$, for $g \in G$ and $i \in \mathbb{Z}$. It can be easily verified that $[G, \langle x \rangle] \subseteq \ker(\mathrm{ct}) \cap \ker(\epsilon)$. The other inclusion can be proved by induction on $\mathrm{l}(w)$ of the words $w \in \ker(\mathrm{ct}) \cap \ker(\epsilon)$, similarly to what we did in the proof of Theorem 2.6. Then

$$[G, \langle x \rangle] = \ker(\mathrm{ct}) \cap \ker(\epsilon).$$

In particular, $[G, \langle x \rangle]$ is a normal subgroup of $G[x]$, being the kernel of the natural surjective homomorphism $G[x] \to G \times \langle x \rangle$ mapping $w \mapsto (\mathrm{ct}(w), x^{\epsilon(w)})$.

Then, we have the following map of relevant subgroups of $G[x]$ considered so far.

$$\begin{tikzcd}
 & G[x] \arrow[dl, dash] \arrow[dr, dash, "G[x]/\ker(\epsilon) \cong \mathbb{Z}"] & \\
\mathrm{ct}^{-1}(G') \arrow[d, dash] \arrow[dr, dash] & & \ker(\epsilon) \arrow[dl, dash] \\
\ker(\mathrm{ct}) \arrow[d, dash] \arrow[dr, dash] & \mathrm{ct}^{-1}(G') \cap \ker(\epsilon) = G[x]' \arrow[d, dash] & \\
\mathcal{U}_G \arrow[dr, dash] & \ker(\mathrm{ct}) \cap \ker(\epsilon) = [G, \langle x \rangle] \arrow[d, dash] & \\
 & \mathcal{U}_G^{sing} = \mathcal{U}_G \cap \ker(\epsilon) = \mathcal{U}_G \cap G[x]' &
\end{tikzcd}$$

Using the normal subgroup $\mathcal{U}_G$ of $G[x]$, we can define a congruence relation $\approx$ on $G[x]$ as follows: for a pair of words $w_1, w_2 \in G[x]$, we define $w_1 \approx w_2$ if $w_1 \mathcal{U}_G = w_2 \mathcal{U}_G$. Then

$$w_1 \approx w_2 \ \text{ if and only if } \ \Phi_G(w_1) = \Phi_G(w_2), \text{ i.e., } f_{w_1} = f_{w_2}.$$

In particular, a word w is universal when $w \approx e_{G[x]}$, i.e. f_w is the constant function e_G on G. Note that the quotient group is $G[x]/_{\approx} = G[x]/\mathcal{U}_G \cong \mathscr{F}(G)$.

A second monoid operation in $\mathscr{F}(G)$ can be introduced as follows. If w is as in (4), and $w_1 \in G[x]$, one can consider the word

$$\xi_{w_1}(w) = g_1 w_1^{\varepsilon_1} g_2 w_1^{\varepsilon_2} \cdots g_n w_1^{\varepsilon_n} g_0$$

obtained substituting w_1 to x in w and taking products in $G[x]$. We shall also denote by $w \circ w_1$ the word $\xi_{w_1}(w)$. On the other hand, one can consider the usual composition of the associated verbal functions $f_w, f_{w_1} \in \mathscr{F}(G)$. Then this composition of words is compatible with the composition of functions, in the sense that

$$f_w \circ f_{w_1} = f_{w \circ w_1} \in \mathscr{F}(G).$$

With this operation, $(\mathscr{F}(G), \circ)$ is a monoid, with identity element the identity function $\mathrm{id}_G = f_x$ of G, mapping $\mathrm{id}_G \colon g \mapsto g$ for every $g \in G$. Obviously, $(\mathscr{F}(G), \circ)$ is a submonoid of the monoid $(G^G, \circ)$ of all self-maps $G \to G$.

3.3. Monomials

Even if a group G has a quite simple structure (for example, is abelian), the group of words $G[x]$ may be more difficult to study (for example, $G[x]$ is never abelian, unless G is trivial). As we are more interested in its quotient group of verbal function $\mathscr{F}(G)$, it will be useful to consider some subset $W \subseteq G[x]$ such that $G[x] = W \cdot \mathcal{U}_G$, i.e. $\Phi_G(W) = \{f_w \mid w \in W\} = \mathscr{F}(G)$, i.e. $\mathscr{F}(G) = W/_{\approx}$. In the following §3.4 we will present such an appropriate subset $W \subseteq G[x]$ in the case when G is abelian.

A word of the form $w = gx^m$, for $g \in G$ and $m \in \mathbb{Z}$, is called a *monomial*. One can associate a monomial to an arbitrary word $w = g_1 x^{\varepsilon_1} g_2 x^{\varepsilon_2} \cdots g_n x^{\varepsilon_n} g_0 \in G[x]$ as follows, letting

$$w_{ab} = \mathrm{ct}(w) x^{\epsilon(w)} = g_1 g_2 \cdots g_n g_0 x^{\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_n} \in G[x]. \tag{8}$$

The monomials in $G[x]$ do not form a subgroup unless G is trivial. Nevertheless, one can "force" them to form a group, by taking an appropriate *quotient* of $G[x]$. Indeed, recall the surjective homomorphism $G[x] \to G \times \langle x \rangle$ mapping $w \mapsto (\mathrm{ct}(w), x^{\epsilon(w)})$ considered in Remark 3.6. Then the group $G \times \langle x \rangle$ "parametrizes" in the obvious way all monomials of $G[x]$ (although the group operation is not the one from $G[x]$).

3.4. A leading example: the abelian case

A case when $\mathscr{F}(G)$ has a very transparent description is that of abelian groups. Let $(G, +, 0_G)$ be an abelian group. While $G[x]$ is not abelian in any case, its quotient $\mathscr{F}(G)$ becomes indeed abelian, and so we keep additive notation also to denote words $w \in G[x]$. Remind that we really are interested only in the evaluation function $f_w \in \mathscr{F}(G)$ associated to w, and to its preimage $E_w^G = f_w^{-1}(\{0_G\}) = \{g \in G \mid f_w(g) = 0_G\}$ (see Definition 4.1).

Then, $w \approx w_{ab} = \mathrm{ct}(w) + \epsilon(w)x$ for every word $w \in G[x]$, and in particular, letting

$$W = \{w_{ab} \mid w \in G[x]\} = \{g + nx \mid g \in G, n \in \mathbb{Z}\} \subseteq G[x],$$

we have $W/_{\approx} = G[x]/_{\approx}$, so that

$$\mathscr{F}(G) = \{f_{g+nx} \mid g \in G, n \in \mathbb{Z}\}.$$

For these reasons, when G is abelian, we will only consider the monomials $w \in W$. These observations are heavily used in computing $\mathscr{F}(G)$ for an abelian group G (hence also $\mathbb{E}_G$, see Example 4.3).

Note that the surjective homomorphism $\Psi_G : G[x] \to G \times \mathbb{Z}$, mapping $w \mapsto (\mathrm{ct}(w), \epsilon(w))$, has kernel $\ker(\Psi_G) = \ker(\mathrm{ct}) \cap \ker(\epsilon) = G[x]'$ by Theorem 2.6, so that $G[x]/G[x]' \cong G \times \mathbb{Z}$. So, if one considers the quotient $G[x]/G[x]'$, the canonical projection $G[x] \to G[x]/G[x]'$ is exactly $w \mapsto w_{ab} = \mathrm{ct}(w) + \epsilon(w)x$. Moreover, being $\mathscr{F}(G) \cong G[x]/\mathcal{U}_G$ abelian, we have that $\mathcal{U}_G \geq G[x]'$, so $\mathscr{F}(G) \cong \frac{G[x]/G[x]'}{G[x]'/\mathcal{U}_G}$ is a quotient of $G \times \mathbb{Z}$.

Here we give an explicit description of the group $\mathscr{F}(G)$.

Proposition 3.7. *If G is an abelian group, then:*

$$\mathscr{F}(G) \cong \begin{cases} G \times \mathbb{Z} & \textit{if } \exp(G) = 0, \\ G \times \mathbb{Z}_n & \textit{if } \exp(G) = n > 0. \end{cases}$$

Proof. Let $n = \exp(G) \in \mathbb{N}$. Note that $\Psi'_G : G \times \mathbb{Z} \to \mathscr{F}(G)$, mapping $(g, k) \to f_{g+kx}$, is a surjective group homomorphism, and that $(g, k) \in \ker(\Psi'_G)$ if and only if $w = g + kx \in \mathcal{U}_G$.

In this case, $g = \mathrm{ct}(w) = 0_G$ by Example 3.5, item 1, so that $w = kx$. If $n = 0$, then $k = 0$. If $n > 0$, then either $k = 0$, or $k \neq 0$ and $n \mid k$. In any case, $k \in n\mathbb{Z}$.

This proves $\ker(\Psi'_G) = \{0_G\} \times n\mathbb{Z}$. □

4. Elementary algebraic subsets

This section is focused on the family $\mathbb{E}_G \subseteq \mathcal{P}(G)$, consisting of preimages $f_w^{-1}(\{e_G\})$, rather than on the group $G[x]$, consisting of words w, or its quotient $\mathscr{F}(G)$, consisting of verbal functions f_w. We begin recalling Markov's definition of elementary algebraic subset of a group, using the terminology of verbal functions.

Definition 4.1. *If $w \in G[x]$, we let*

$$E_w^G = f_w^{-1}(\{e_G\}) = \{g \in G \mid f_w(g) = e_G\} \subseteq G,$$

we call E_w^G elementary algebraic subset of G, *and we denote it simply by E_w when no confusion is possible. We denote by $\mathbb{E}_G = \{E_w \mid w \in G[x]\} \subseteq \mathcal{P}(G)$ the family of elementary algebraic subsets of G.*

According to Markov's definition on page 71, if $X \subseteq G$, we call it:

- additively algebraic *if X is a finite union of elementary algebraic subsets of G, i.e. if $X \in \mathbb{E}_G^{\cup}$;*
- algebraic *if X is an intersection of additively algebraic subsets of G.*

Then the algebraic subsets form the family of $\mathfrak{Z}_G$-closed sets, and $\mathbb{E}_G$ is a subbase for the $\mathfrak{Z}_G$-closed sets; while the additively algebraic subsets are exactly the members of $\mathbb{E}_G^{\cup}$, and are a base for the $\mathfrak{Z}_G$-closed sets.

Remark 4.2. If $w = g \in G$ is a constant word, then either $E_w = G$ or $E_w = \emptyset$ (depending on whether $g = e_G$ or $g \neq e_G$). For this reason, in studying the family $\mathbb{E}_G$ and the topology $\mathfrak{Z}_G$, there is no harm in assuming w not to be constant, i.e. to be as in Markov's definition (1) of elementary algebraic subset (see also §4.1).

Let us fix a group G. We will now consider the iterated images of G under ϖ^n, for $n \in \mathbb{N}_+$, and to this end we need to introduce a countable set of variables $\{x_n \mid n \in \mathbb{N}_+\}$. Then applying ϖ we obtain the following diagram:

$$G \xrightarrow{\varpi} G[x_1] \xrightarrow{\varpi} (G[x_1])[x_2] \xrightarrow{\varpi} \big((G[x_1])[x_2]\big)[x_3] \xrightarrow{\varpi} \dots \tag{9}$$

If $n \in \mathbb{N}_+$, we let $G_n = G[x_1, \dots, x_n] = \varpi^n(G)$, and it can be proved that if $\sigma \in S_n$, then

$$G_n \cong G[x_{\sigma(1)}, \dots, x_{\sigma(n)}] \,.$$

Every $w = w(x_1, \dots, x_n) \in G_n$ determines the associated evaluation function of n variables over G, that we denote by $f_w \colon G^n \to G$, in analogy with Definition 3.1.

Finally, one can define $E_w \subseteq G^n$ as the preimage $E_w = f_w^{-1}(\{e_G\})$, and consider the family $\{E_w \mid w \in G_n\}$ as a subbase for the closed sets of a topology on G^n.

These observations are the basis of a theory of algebraic geometry over groups, recently started with [6] and developed in a series of subsequent papers. In this work, we focus on the case when $n = 1$, considering only verbal functions $f_w \colon G \to G$ of one variable, and elementary algebraic subsets $E_w \subseteq G$.

Example 4.3 (**A leading example: the abelian case II**). Let G be an abelian group (see §3.4). Then the elementary algebraic subset of G determined by f_{g+nx} is

$$E_{g+nx} = \begin{cases} \emptyset & \text{if } g + nx = 0_G \text{ has no solution in } G, \\ G[n] + x_0 & \text{if } x_0 \text{ is a solution of } g + nx = 0_G. \end{cases} \tag{10}$$

On the other hand, if $n \in \mathbb{Z}$, and $g \in G$, then $G[n] + g = E_{nx-ng}$. So the non-empty elementary algebraic subsets of G are exactly the cosets of the n-socles of G:

$$\mathbb{E}_G \setminus \{\emptyset\} = \{G[n] + g \mid n \in \mathbb{N}, g \in G\}. \tag{11}$$

Then $\mathbb{E}_G^{\cup}$ is the family of all the $\mathfrak{Z}_G$-closed subsets of an abelian group G. In other words, *every algebraic subset of G is additively algebraic.*

It follows from Remark 4.2 and (10) that if G is abelian, and $w \in G[x]$ is singular, then either $E_w = G$ or $E_w = \emptyset$.

There are easy examples showing that in general none of the elementary algebraic subsets $E_{x^n}^G = G[n]$ need to be a coset of a subgroup. See for example [17], where we show a class of groups G such that the subgroup generated by $G[n]$ is the whole group G, for every $n \in \mathbb{N}_+$.

Now we prove that the inverse image of an elementary algebraic subset under a verbal function is still an elementary algebraic subset.

LEMMA 4.4. *For every group G, the family $\mathbb{E}_G$ is stable under taking inverse image under verbal functions.*

Proof. For every pair $w, w' \in G[x]$, consider the verbal function f_w and the elementary algebraic subset $E_{w'}$. Then

$$f_w^{-1}E_{w'} = f_w^{-1}f_{w'}^{-1}(\{e_G\}) = (f_{w'} \circ f_w)^{-1}(\{e_G\}) = f_{w' \circ w}^{-1}(\{e_G\}) = E_{w' \circ w}, \quad (12)$$

so that $f_w^{-1}E_{w'} \in \mathbb{E}_G$. □

As a first application of Lemma 4.4, we see that the translate of an elementary algebraic subset is still an elementary algebraic subset.

EXAMPLE 4.5.

1. By Example 1.1, item 5, the left translation in G by an element $g \in G$ is the verbal function f_{gx}, and so $gS = f_{g^{-1}x}^{-1}(S)$ for every subset $S \subseteq G$. In particular, by (12) we have

$$gE_w = f_{g^{-1}x}^{-1}(E_w) = E_{w \circ g^{-1}x}. \quad (13)$$

Similarly, $E_w g = E_{w \circ xg^{-1}}$. Note that $\epsilon(w \circ g^{-1}x) = \epsilon(w) = \epsilon(w \circ xg^{-1})$.

2. If $a \in G$, then $C_G(a) = E_w$, for the word $w = axa^{-1}x^{-1} \in G[x]$ by Example 1.2, item 2. By (13), its left coset determined by an element $g \in G$ is $gC_G(a) = E_{w_1}$ for

$$w_1 = w \circ (g^{-1}x) = a(g^{-1}x)a^{-1}(g^{-1}x)^{-1} = ag^{-1}xa^{-1}x^{-1}g.$$

Note also that, for $w_2 = gw_1g^{-1} = (gag^{-1})xa^{-1}x^{-1}$, we have $E_{w_2} = E_{w_1} = gC_G(a)$.

On the other hand, $C_G(a)g = gg^{-1}C_G(a)g = gC_G(g^{-1}ag)$, so that

$$\mathcal{C} = \{gC_G(a) \mid a, g \in G\}$$

is the family of all cosets of one-element centralizers in G. By the above observations,

$$\mathcal{C} = \{E_w \mid \exists a, g \in G\ w = (gag^{-1})xa^{-1}x^{-1}\} \subseteq \mathbb{E}_G.$$

4.1. Further reductions

As already noted above, to study $\mathscr{F}(G)$ it is sufficient to consider a subset $W \subseteq G[x]$ such that $\mathscr{F}(G) = \Phi_G(W) = W/_{\approx}$. Since our effort is really devoted to the study of the Zariski topology $\mathfrak{Z}_G$ on a group G, hence to the family $\mathbb{E}_G$, a further reduction is also possible as follows.

As an example to introduce this reduction, consider the abelian group $G = \mathbb{Z} \times \mathbb{Z}_2$, and the verbal functions $f_w, f_{w'} \in \mathscr{F}(G)$, associated to $w = 2x, w' = 4x \in G[x]$. Then $f_w \neq f_{w'}$, and yet $E_w = f_w^{-1}(\{0_G\}) = \{0_\mathbb{Z}\} \times \mathbb{Z}_2 = f_{w'}^{-1}(\{0_G\}) = E_{w'}$.

Another example of a more general property could be the following: consider a word $w \in G[x]$, and its inverse $w^{-1} \in G[x]$. Obviously $f_{w^{-1}} = (f_w)^{-1} \neq f_w$ in general, but for an element $g \in G$ we have $f_{w^{-1}}(g) = e_G$ if and only if $f_w(g) = e_G$. In particular,

$$E_{w^{-1}} = f_{w^{-1}}^{-1}(\{e_G\}) = f_w^{-1}(\{e_G\}) = E_w.$$

So $E_w = E_{w^{-1}}$ in every group G, and in Remark 4.6 below we slightly generalize this result.

So we introduce another equivalence relation $\sim$ on $G[x]$ defined as follows: for a pair of words $w_1, w_2 \in G[x]$, we define $w_1 \sim w_2$ if $E_{w_1} = E_{w_2}$. Obviously, $w \approx w'$ implies $w \sim w'$.

For example, as noted above $w \sim w^{-1}$ for every $w \in G[x]$.

Remark 4.6. Let $w \in G[x]$, and $s \in \mathbb{Z}$. Consider the element $w^s \in G[x]$, and note that $\epsilon(w^s) = s\epsilon(w)$ and $w^s(g) = (w(g))^s$ for every $g \in G$. Hence, $E_{w^s} = \{g \in G \mid (w(g))^s = e_G\} = f_w^{-1}(G[s])$ is the preimage of $G[s]$ under f_w.

In particular, if $G[s] = \{e_G\}$, then $E_{w^s} = E_w$, i.e. $w \sim w^s$.

Then, in describing $\mathbb{E}_G$, we can restrict ourselves to a subset $W \subseteq G[x]$ of representants with respect to the equivalence $\sim$, that is such that the quotient set $W/_{\sim} = G[x]/_{\sim}$. For example, if $W \subseteq G[x]$ satisfies $\mathscr{F}(G) = \{f_w \mid w \in W\}$, that is $G[x]/_{\approx} = W/_{\approx}$, then $G[x]/_{\sim} = W/_{\sim}$. As we have seen in §3.4, in the *abelian* case the set $W = \{gx^n \in G[x] \mid n \in \mathbb{N},\ g \in G\}$ satisfies $G[x]/_{\approx} = W/_{\approx}$, so that this W will do.

Finally, note that $f_w(g) = e_G$ if and only if $f_{awa^{-1}}(g) = e_G$ holds for every $a \in G$, so that $w \sim awa^{-1}$, and $\epsilon(awa^{-1}) = \epsilon(w)$. Then, in describing $\mathbb{E}_G$, there is no harm in assuming that a word $w = g_1x^{\varepsilon_1}g_2x^{\varepsilon_2}\cdots g_nx^{\varepsilon_n}g_0 \in G[x]$ has $g_0 = e_G$ (or $g_1 = e_G$); indeed, *from now on, we will often consider exclusively words w of the form*

$$w = g_1x^{\varepsilon_1}g_2x^{\varepsilon_2}\cdots g_nx^{\varepsilon_n} \in G[x]. \tag{14}$$

Lemma 4.7. *Let $v \in G[x]$. Then $v \sim w$ for a word $w \in G[x]$ as in (14), with $\epsilon(w) = |\epsilon(v)| \geq 0$.*

Proof. By Remark 4.6, we have that $v \sim v^{-1}$, and $\epsilon(v^{-1}) = -\epsilon(v)$, so that we can assume $\epsilon(v) \geq 0$.

Then, by the above discussion, $v \sim w$ for a word w as in (14), and with $\epsilon(w) = \epsilon(v)$. □

5. Quasi-topological group topologies

Let X be a set, and λ be an infinite cardinal number. We denote by $[X]^{<\lambda}$ the family of subsets of X having size strictly smaller than λ.

As the family $\mathcal{B} = [X]^{<\lambda} \cup \{X\}$ is stable under taking finite unions and arbitrary intersections, it is the family of closed sets of a topology on X, denoted by *co*-λ_X. For example, taking $\lambda = \omega$, one obtains the *cofinite topology*

$$cof_X = co\text{-}\omega_X.$$

For example, for every infinite cardinal number λ, the space $(G, co\text{-}\lambda_G)$ is a T_1 quasi-topological group. In particular, if G is infinite, (G, cof_G) is a T_1, non-Hausdorff (being Noetherian) quasi-topological group. So if G is infinite, then (G, cof_G) is not a topological group. We will use the topologies *co*-λ_G on G as a source of counter-examples in Example 5.5.

5.1. General results

In what follows we give some general results for quasi-topological groups. For a reference on this topic, see for example [2].

THEOREM 5.1. *Let (G, τ) be a quasi-topological group.*

(a) If $S \subseteq G$, then the τ-closure of S is

$$\overline{S} = \bigcap_{U \in \mathcal{V}_\tau(e_G)} U \cdot S = \bigcap_{V \in \mathcal{V}_\tau(e_G)} S \cdot V.$$

(b) If H is a subgroup with non-empty interior, then H is open.

(c) A finite-index closed subgroup of G is open.

(d) The closure of a (normal) subgroup is a (normal) subgroup.

Proof. To prove (a), (b) and (c) one only needs the inversion and shifts to be continuous, so proceed as in the case of topological groups.

(d) Let H be a subgroup of G, and $\overline{H}$ be its τ-closure. We have to show that $\overline{H}$ is a subgroup, that is: $\overline{H}^{\,-1} \subseteq \overline{H}$ and $\overline{H} \cdot \overline{H} \subseteq \overline{H}$.

The hypothesis that the inversion function is τ-continuous guarantees that $\overline{H}^{\,-1} \subseteq \overline{H^{-1}} = \overline{H}$.

In the same way, for every $h \in H$, the left traslation by h in G is τ-continuous, so $h \cdot \overline{H} \subseteq \overline{h \cdot H} = \overline{H}$; as this holds for every $h \in H$, we get $H \cdot \overline{H} \subseteq \overline{H}$.

Now consider the right traslation in G by an element $c \in \overline{H}$. It is τ-continuous, so

$$\overline{H} \cdot c \subseteq \overline{H \cdot c} \subseteq \overline{H \cdot \overline{H}} \subseteq \overline{\overline{H}} = \overline{H}.$$

From the above inclusion, we finally deduce $\overline{H} \cdot \overline{H} \subseteq \overline{H}$.

Composing translations we obtain that also conjugations are τ-continuous; so if H is a normal subgroup and $g \in G$, then

$$g \cdot \overline{H} \cdot g^{-1} \subseteq \overline{g \cdot H \cdot g^{-1}} = \overline{H}.$$

□

Let (G, τ) be a quasi-topological group, and N be a normal subgroup of G. Consider the quotient group $\overline{G} = G/N$ and the canonical map $\pi\colon (G, \tau) \to \overline{G}$. The quotient topology $\overline{\tau}$ of τ on $\overline{G}$ is the final topology of π, namely $\overline{\tau} = \{A \subseteq \overline{G} \mid \pi^{-1}(A) \in \tau\}$. Then, the following results hold.

Proposition 5.2. *If (G, τ) is a quasi-topological group, then $(\overline{G}, \overline{\tau})$ is a quasi-topological group, and the map $\pi\colon (G, \tau) \to (\overline{G}, \overline{\tau})$ is continuous and open. In particular, $\overline{\tau} = \{\pi(X) \subseteq \overline{G} \mid X \in \tau\}$.*

Proof. Proceed as in the case of topological groups to verify that $(\overline{G}, \overline{\tau})$ is a quasi-topological group.

We prove that π is open. Let $A \in \tau$, and note that $\pi(A) \in \overline{\tau}$ if and only if $\pi^{-1}\pi(A) \in \tau$. As

$$\pi^{-1}\pi(A) = A \cdot N = \bigcup_{n \in N} A \cdot n,$$

and $(\overline{G}, \overline{\tau})$ is a quasi-topological group, we are done. □

Proposition 5.3. *If (G, τ) is a quasi-topological group, then the following are equivalent.*

(1) N is τ-closed;

(2) $\{e_{\overline{G}}\}$ is $\overline{\tau}$-closed;

(3) $\overline{\tau}$ is a T_1 topology.

Proof. (1) ⇒ (2). Let N be τ-closed. We are going to prove that $A = \overline{G} \setminus \{e_{\overline{G}}\}$ is $\overline{\tau}$-open, and note that this holds if and only if $\pi^{-1}(A)$ is τ-open. As $\pi^{-1}(A) = G \setminus N$ is τ-open by assumption, we are done.

(2) ⇒ (3) holds as $(\overline{G}, \overline{\tau})$ is a quasi-topological group by Proposition 5.2.

(3) ⇒ (1). If $\overline{\tau}$ is a T_1 topology, in particular $\{e_{\overline{G}}\}$ is $\overline{\tau}$-closed, so that $N = \pi^{-1}(\{e_{\overline{G}}\})$ is τ-closed, being π continuous. □

5.2. Verbal functions

We will now characterize which topologies on a group make it a quasi-topological group, in term of continuity of an appropriate family of verbal functions.

LEMMA 5.4. *Let G be a group, and τ a topology on G. Then (G, τ) is a quasi-topological group if and only if $f_w : (G, \tau) \to (G, \tau)$ is continuous for every word of the form $w = gx^\varepsilon \in G[x]$, with $g \in G$ and $\varepsilon = \pm 1$.*

In particular, if a topology σ on a group G makes continuous every verbal function, then (G, σ) is a quasi-topological group. If σ is also T_1, then $\mathfrak{Z}_G \subseteq \sigma$.

Proof. Let ι denote the inversion function of G. If (G, τ) is a quasi-topological group, then every verbal function of the form f_{gx}, being a left translation, is τ-continuous. Then also every $f_{gx^{-1}} = f_{gx} \circ \iota$ is τ-continuous.

For the converse, let τ be a topology on G such that $f_w : (G, \tau) \to (G, \tau)$ is continuous for every word $w = gx^\varepsilon$, with $g \in G$ and $\varepsilon = \pm 1$. Then items 3 and 5 in Example 1.1 show that the inversion and the left translations are verbal functions of this form, hence are τ-continuous. Finally, the right translation by an element g is $f_{xg} = f_{x^{-1}} \circ f_{g^{-1}x^{-1}}$.

For the last part, note that if σ is T_1, then $\{e_G\}$ is σ-closed. If moreover every f_w is σ-continuous, then also every $E_w = f_w^{-1}(\{e_G\})$ is σ-closed. As $\mathbb{E}_G$ is a subbase for the $\mathfrak{Z}_G$-closed sets, we conclude $\mathfrak{Z}_G \subseteq \sigma$. □

EXAMPLE 5.5. Let (G, τ) be a T_1 quasi-topological group. By Lemma 5.4, every verbal function in $\{f_{gx^\varepsilon} \mid g \in G, \varepsilon = \pm 1\}$ is τ-continuous. We shall see that not every verbal function need to be τ-continuous.

To this end, recall that the space $(G, co\text{-}\lambda_G)$ is a T_1 quasi-topological group for every infinite cardinal number λ.

So let $\omega \leq \lambda < \kappa = |G|$, note that $co\text{-}\lambda_G$ is not the discrete topology on G, and consider $\tau = co\text{-}\lambda_G$.

1. Let G be a group having a non-central element a such that $|C_G(a)| \geq \lambda$ (for example, the group $G = \oplus_\kappa S_3$ will do). Then let $w = [a, x] \in G[x]$, and consider the commutator verbal function $f_w \in \mathscr{F}(G)$. As $f_w^{-1}(\{e_G\}) = C_G(a) \neq G$, we have that f_w is not τ-continuous.

2. Let G be a non-abelian group such that $|G[2]| \geq \lambda$ (also in this case the group $G = \oplus_\kappa S_3$ considered above will do). Then let $w = x^2 \in G[x]$, and consider the verbal function $f_w \in \mathscr{F}(G)$. As $f_w^{-1}(\{e_G\}) = G[2] \neq G$, we have that f_w is not τ-continuous.

In the following results we prove that $(G, \mathfrak{Z}_G)$, $(G, \mathfrak{M}_G)$ and $(G, \mathfrak{P}_G)$ are quasi-topological groups.

PROPOSITION 5.6. *For every group G, the following hold.*

1. *Every verbal function is* $\mathfrak{Z}_G$*-continuous.*

2. *The pair* $(G, \mathfrak{Z}_G)$ *is a quasi-topological group.*

3. $\mathfrak{Z}_G$ *is the initial topology of the family of all verbal functions* $\{f \colon G \to (G, \mathfrak{Z}_G) \mid f \in \mathscr{F}(G)\}$.

Proof. 1. Follows from the fact that $\mathbb{E}_G$ is a subbase for the $\mathfrak{Z}_G$-closed subsets of G, and from Lemma 4.4.

2. Immediately follows by Lemma 5.4 and item 1.

3. Also follows by item 1.

□

Corollary 5.7. *Every group topology on a group G makes continuous every verbal function of* G*. In particular* $\mathfrak{M}_G$ *and* $\mathfrak{P}_G$ *make continuous every verbal function of* G*, so* $\mathfrak{Z}_G \subseteq \mathfrak{M}_G \subseteq \mathfrak{P}_G$*, and all the three are quasi-topological group topologies.*

Proof. As a verbal function is a composition of products and inversions, it is continuous with respect to every group topology. The same is true for $\mathfrak{M}_G$ and $\mathfrak{P}_G$, which are intersections of group topologies, then Lemma 5.4 applies. □

If N is a normal subgroup of a group G, and $\overline{G} = G/N$ is the quotient group, consider the quotient topology $\overline{\mathfrak{Z}}_G$ on $\overline{G}$. In the following proposition we prove that every verbal function $(\overline{G}, \overline{\mathfrak{Z}}_G) \to (\overline{G}, \overline{\mathfrak{Z}}_G)$ of $\overline{G}$ is continuous.

Proposition 5.8. *Let* N *be a normal subgroup of a group* G*, and let* $\overline{G} = G/N$*. Then the quotient topology* $\overline{\mathfrak{Z}}_G$ *makes continuous every verbal function of* $\overline{G}$*.*

Proof. Let $v = \overline{g}_1 x^{\varepsilon_1} \overline{g}_2 x^{\varepsilon_2} \cdots \overline{g}_n x^{\varepsilon_n} \in \overline{G}[x]$, and we have to prove that

$$f_v \colon (\overline{G}, \overline{\mathfrak{Z}}_G) \to (\overline{G}, \overline{\mathfrak{Z}}_G)$$

is continuous. If $w = g_1 x^{\varepsilon_1} g_2 x^{\varepsilon_2} \cdots g_n x^{\varepsilon_n} \in G[x]$, then $v = \overline{w}$, and in the notation of Remark 3.3, item 2, the following diagram commutes.

$$\begin{array}{ccc} (G, \mathfrak{Z}_G) & \xrightarrow{f_w} & (G, \mathfrak{Z}_G) \\ \downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi} \\ (\overline{G}, \overline{\mathfrak{Z}}_G) & \xrightarrow{f_{\overline{w}}} & (\overline{G}, \overline{\mathfrak{Z}}_G). \end{array} \tag{15}$$

As f_w is continuous and $\overline{\mathfrak{Z}}_G$ is the final topology of the canonical projection $\pi \colon (G, \mathfrak{Z}_G) \to \overline{G}$, also $f_{\overline{w}}$ is continuous. □

The main result of this subsection is the following theorem characterizing the normal subgroups N of G such that the canonical projection $\pi\colon (G, \mathfrak{Z}_G) \to (\overline{G}, \mathfrak{Z}_{\overline{G}})$ is continuous, where $\overline{G} = G/N$.

Theorem 5.9. *Let N be a normal subgroup of a group G, and let $\overline{G} = G/N$. Then the following conditions are equivalent:*

(1) N is $\mathfrak{Z}_G$-closed;

(2) $\overline{\mathfrak{Z}}_G$ is a T_1 topology;

(3) $\mathfrak{Z}_{\overline{G}} \subseteq \overline{\mathfrak{Z}}_G$;

(4) the canonical map $\pi\colon (G, \mathfrak{Z}_G) \to (\overline{G}, \mathfrak{Z}_{\overline{G}})$ is continuous.

Proof. (1) $\Leftrightarrow$ (2) follows by Proposition 5.3 and Proposition 5.6, item 2.

(2) $\Rightarrow$ (3) follows by Proposition 5.8 and Lemma 5.4.

(3) $\Rightarrow$ (4). In this case, the map $\mathrm{id}\colon (\overline{G}, \overline{\mathfrak{Z}}_G) \to (\overline{G}, \mathfrak{Z}_{\overline{G}})$ is continuous, and so also the composition

$$(G, \mathfrak{Z}_G) \xrightarrow{\pi} (\overline{G}, \overline{\mathfrak{Z}}_G) \xrightarrow{\mathrm{id}} (\overline{G}, \mathfrak{Z}_{\overline{G}}).$$

(4) $\Rightarrow$ (1) holds as $\{e_{\overline{G}}\}$ is $\mathfrak{Z}_{\overline{G}}$-closed and $N = \pi^{-1}(\{e_{\overline{G}}\})$. $\square$

As an application of Theorem 5.9, we prove in the following corollary that every n-th center $Z_n(G)$ of G is $\mathfrak{Z}_G$-closed, where $Z_n(G) \leq G$ is defined inductively as follows, for $n \in \mathbb{N}_+$. Let $Z_1(G) = Z(G)$. Consider the quotient group $G/Z(G)$, its center $Z\big(G/Z(G)\big)$, and its preimage $Z_2(G) \leq G$ under the canonical projection $\pi\colon G \to G/Z(G)$. Proceed by induction to define an ascending chain of characteristic subgroups $Z_n(G)$.

Corollary 5.10. *For every group G, and every positive integer n, the subgroup $Z_n(G)$ is $\mathfrak{Z}_G$-closed.*

Proof. The center $Z(G) = Z_1(G)$ is $\mathfrak{Z}_G$-closed by Example 1.2, item 2. If $\overline{G} = G/Z(G)$, then the projection $\pi\colon (G, \mathfrak{Z}_G) \to (\overline{G}, \mathfrak{Z}_{\overline{G}})$ is continuous by Theorem 5.9. As $Z(\overline{G})$ is $\mathfrak{Z}_{\overline{G}}$-closed, we have that $Z_2(G) = \pi^{-1}(Z(\overline{G}))$ is $\mathfrak{Z}_G$-closed.

Proceed by induction to get the thesis. $\square$

Remark 5.11. Corollary 5.10 can also be proved observing that it is possible to define by induction $Z_1(G) = Z(G)$ and, for an integer $i \geq 1$, and $x \in G$, note that $x \in Z_{i+1}(G)$ if and only if $[g, x] = gxg^{-1}x^{-1} \in Z_i(G)$ for every $g \in G$. Equivalently, if $w_g = [g, x] \in G[x]$, then

$$Z_{i+1}(G) = \bigcap_{g \in G} \{x \in G \mid [g, x] \in Z_i(G)\} = \bigcap_{g \in G} f_{w_g}^{-1}(Z_i(G)).$$

As f_{w_g} is $\mathfrak{Z}_G$-continuous by Proposition 5.6, item 1, for every $g \in G$, and $Z_i(G)$ is $\mathfrak{Z}_G$-closed by inductive hypothesis, we deduce that $Z_{i+1}(G)$ is $\mathfrak{Z}_G$-closed.

5.3. Partial Zariski topologies

Given a subset $W \subseteq G[x]$, we consider the family $\mathcal{E}(W) = \{E_w^G \mid w \in W\} \subseteq \mathbb{E}_G$ of elementary algebraic subsets of G determined by the words $w \in W$. Then, following [4] and [5], we consider the topology $\mathfrak{T}_W$ having $\mathcal{E}(W)$ as a subbase for its closed sets.

Example 5.12. 1. Note that $\mathcal{E}(G[x]) = \mathbb{E}_G$, so $\mathfrak{T}_{G[x]} = \mathfrak{Z}_G$.

2. Taking $W = \{gx \mid g \in G\}$, one obtains that $\mathcal{E}(W) = \{\{g\} \mid g \in G\}$, so that $\mathfrak{T}_W = cof_G$.

Lemma 5.13. *Let $W \subseteq G[x]$, and assume that $gw \in W$, for every $w \in W$ and every $g \in G$. Then $\mathfrak{T}_W$ is the initial topology of the family of verbal functions $\{f_w : G \to (G, cof_G) \mid w \in W\}$.*

Proof. If τ is such initial topology, then $\mathcal{F} = \big\{f_w^{-1}(\{g\}) = f_{g^{-1}w}^{-1}(\{e_G\}) \mid w \in W, g \in G\big\}$ is a subbase for the τ-closed sets.

By assumption, $\mathcal{F}$ coincides with $\mathcal{E}(W) = \big\{E_w = f_w^{-1}(\{e_G\}) \mid w \in W\big\}$, so that $\tau = \mathfrak{T}_W$. □

In particular, $\mathfrak{Z}_G$ can be equivalently defined as the initial topology of the family of all verbal functions $\{f : G \to (G, cof_G) \mid f \in \mathscr{F}(G)\}$.

Example 5.14. Let $a, b \in G$, and $w = bxax^{-1} = ba[a^{-1}, x] \in G[x]$. Note that $E_w \neq \emptyset$ if and only if there exists an element $g \in G$ such that $b = ga^{-1}g^{-1}$, i.e. b and a^{-1} are conjugated elements in G. In this case, $w = (ga^{-1}g^{-1})xax^{-1}$.

In particular, letting $V = \{bxax^{-1} \mid a, b \in G\} \subseteq G[x]$ and

$$W_{\mathfrak{C}} = \{[g, a][a, x] = (gag^{-1})xa^{-1}x^{-1} \mid a, g \in G\} \subseteq V, \tag{16}$$

we obtain that $\mathcal{E}(V) \setminus \emptyset = \mathcal{E}(W_{\mathfrak{C}}) \subseteq \mathcal{E}(V)$, so that $\mathfrak{T}_V = \mathfrak{T}_{W_{\mathfrak{C}}}$. Moreover, by Example 4.5, item 2, we have

$$\mathcal{E}(W_{\mathfrak{C}}) = \{gC_G(a) \mid a, g \in G\}. \tag{17}$$

Definition 5.15. *Given a group G, we denote by $\mathfrak{C}_G$ the topology $\mathfrak{T}_{W_{\mathfrak{C}}}$, for $W_{\mathfrak{C}} \subseteq G[x]$ as in (16). We call $\mathfrak{C}_G$ the* centralizer topology *of G.*

We will study $\mathfrak{C}_G$ in more details in §5.4.

By definition, the family $\mathcal{C} = \{gC_G(a) \mid a, g \in G\}$ is a subbase for the $\mathfrak{C}_G$-closed subsets of G. On the other hand, one can consider the topology $\mathcal{C}$ generates taking its members as open sets, i.e. the coarsest topology $\mathcal{T}_G$ on G such that $gC_G(a)$ is $\mathcal{T}_G$-open, for every $a, g \in G$. The topology $\mathcal{T}_G$ has been introduced by Taĭmanov in [22] and is now called the *Taĭmanov topology* of G. See for example [10] for a recent work on this topic.

Definition 5.16. *The* Taĭmanov topology $\mathcal{T}_G$ *on a group* G *is the topology having the family of the centralizers of the elements of* G *as a subbase of the filter of the neighborhoods of* e_G.

It is easy to check that $\mathcal{T}_G$ is a group topology, and for every element $g \in G$ the subgroup $C_G(g)$ is a $\mathcal{T}_G$-open (hence, closed) subset of G. So note that $\mathfrak{C}_G \subseteq \mathcal{T}_G$ in general (see Lemma 5.33 for a sufficient condition on G to have $\mathfrak{C}_G = \mathcal{T}_G$).

Note that $\overline{\{e_G\}}^{\mathcal{T}_G} = Z(G)$, so $\mathcal{T}_G$ need not be Hausdorff.

Lemma 5.17 ([10, Lemma 4.1]). *If* G *is a group, then the following hold for* $\mathcal{T}_G$.

1. $\mathcal{T}_G$ *is Hausdorff if and only if* G *is center-free.*
2. $\mathcal{T}_G$ *is indiscrete if and only if* G *is abelian.*

Remark 5.18. If $S \subseteq G$, let

$$C(S) = \{[g,a][a,x] = (gag^{-1})xa^{-1}x^{-1} \mid g \in G, a \in S\} \subseteq G[x],$$
$$D(S) = \{[xcx^{-1}, b] \mid b, c \in S\} \subseteq G[x].$$

For example, $C(G) = W_{\mathfrak{C}}$ as in (16), so that $\mathfrak{T}_{C(G)} = \mathfrak{C}_G$.

In [3], the authors introduced two *restricted Zariski topologies* $\mathfrak{Z}'_G$, $\mathfrak{Z}''_G$ on a group G, that in our notation are respectively $\mathfrak{Z}'_G = \mathfrak{T}_{C(G[2]) \cup D(G[2])}$, and $\mathfrak{Z}''_G = \mathfrak{T}_{C(G[2])}$. Obviously, $\mathfrak{Z}''_G \subseteq \mathfrak{Z}'_G \subseteq \mathfrak{Z}_G$ and $\mathfrak{Z}''_G \subseteq \mathfrak{C}_G$ hold for every group G. See also Theorem 5.26.

In the following definition we introduce the partial Zariski topology $\mathfrak{T}_{mon}$ determined by the monomials. Note that by Lemma 4.7 there is no harm in considering only the monomials with non-negative content. Moreover, by Remark 4.2 we can indeed consider only positive-content monomials.

Definition 5.19. *If* $M = \{gx^n \mid g \in G,\ n \in \mathbb{N}_+\} \subseteq G[x]$ *is the family of the monomials with positive content, then we denote by* $\mathfrak{T}_{mon}$ *the topology having* $\mathcal{E}(M)$ *as a subbase for its closed sets, and we call it the* monomial topology *of* G.

Note that $gx \in M$ for every $g \in G$, so that $\mathfrak{T}_{mon}$ is T_1 topology.

Example 5.20. Let G be abelian. We have seen in §3.4 that $w \approx w_{ab}$ for every $w \in G[x]$. As in studying E_w we can assume $\epsilon(w) \geq 0$ by Lemma 4.7, this shows that $\mathfrak{T}_{mon} = \mathfrak{Z}_G$.

In [21], Ol'shanskij built the first example of a countable group G with $\mathfrak{Z}_G = \delta_G$, so that $\mathfrak{Z}_G = \mathfrak{M}_G = \delta_G$. A closer look at his proof reveals that really also $\mathfrak{T}_{mon} = \delta_G$ for such a group G.

Recall that a group G is said to *satisfy the cancellation law* if $x^n = y^n$ implies $x = y$, for every $n \in \mathbb{N}_+$ and $x, y \in G$. Here we recall also that a group G is called:

- *nilpotent* if $Z_n(G) = G$ for some $n \in \mathbb{N}_+$,
- *torsion-free* if every element has infinite order.

It is a classical result due to Chernikov, that if G is a nilpotent, torsion-free group, then G satisfies the cancellation law.

Corollary 5.21. *If G satisfies the cancellation law, then $\mathfrak{T}_{mon} = cof_G$. In particular, $\mathfrak{T}_{mon} = cof_G$ for every nilpotent, torsion-free group G.*

Proof. It suffices to prove that E_w has at most one element, for every monomial $w = gx^m \in G[x]$ with $m > 0$. So let $w = gx^m$, and assume $a \in E_w$, so that $a^m = g^{-1}$. Then $E_w = \{p \in G \mid p^m = a^m\}$, so that $E_w = \{a\}$.

If G is a nilpotent, torsion-free group, then Chernikov's result applies. □

5.3.1. Permutation groups

In what follows, X is an infinite set. For a subgroup $G \leq S(X)$ of the permutation group of X, recall that $\tau_p(G)$ denote the *point-wise convergence topology of* G. Then $\tau_p(G)$ is a Hausdorff group topology, so that $\mathfrak{Z}_G \subseteq \mathfrak{M}_G \subseteq \tau_p(G)$ for every group $G \leq S(X)$. The following classic result was proved by Gaughan in 1967.

Theorem 5.22 ([18]). *Let $G = S(X)$. Then $\tau_p(G)$ is contained in every Hausdorff group topology on G.*

In particular, it follows from Theorem 5.22 that $\mathfrak{M}_{S(X)} = \tau_p(S(X))$ is itself a Hausdorff group topology.

Ten years after Gaughan's Theorem 5.22, Dierolf and Schwanengel (unaware of his result) proved the following:

Theorem 5.23 ([9]). *Let $S_\omega(X) \leq G \leq S(X)$. Then $\tau_p(G)$ is a minimal Hausdorff group topology.*

Although Theorem 5.23 provides new results for groups $S_\omega(X) \leq G \lneq S(X)$, Theorem 5.22 gives a much stronger result for the whole group $S(X)$. That is why Dikranjan conjectured the following.

Conjecture 5.24 ([19]). *Let $S_\omega(X) \leq G \leq S(X)$. Then $\mathfrak{M}_G = \tau_p(G)$.*

The following question was raised by Dikranjan and Shakhmatov (see Theorem 5.22).

QUESTION 5.25 ([11]). *Does* $\mathfrak{M}_{S(X)}$ *coincide with* $\mathfrak{Z}_{S(X)}$*?*

It has recently turned out that Dikranjan's conjecture is true, and Dikranjan-Shakhmatov's question has a positive answer. It has been proved in [3] that $\mathfrak{Z}_G = \mathfrak{M}_G$ is the pointwise convergence topology for all subgroups G of infinite permutation groups $S(X)$, that contain the subgroup $S_\omega(X)$ of all permutations of finite support.

THEOREM 5.26 ([3]). *If* $S_\omega(X) \leq G \leq S(X)$*, then* $\mathfrak{Z}''_G \subsetneq \mathfrak{Z}'_G = \mathfrak{Z}_G = \mathfrak{M}_G = \tau_p(G)$.

As a corollary of Theorem 5.26, the same authors have obtained the following answer to another question posed by Dikranjan and Shakhmatov.

COROLLARY 5.27 ([3]). *The class of groups G satisfying $\mathfrak{Z}_G = \mathfrak{M}_G$ is not closed under taking subgroups.*

Proof. Let H be a group such that $\mathfrak{Z}_H \neq \mathfrak{M}_H$, embed it in $G = S(H)$, and apply Theorem 5.26 to conclude that $\mathfrak{Z}_G = \mathfrak{M}_G$. $\square$

5.4. Centralizer topologies

In this subsection, we study two partial Zariski topologies. The first one is the topology $\mathfrak{C}_G$ introduced in Definition 5.15. As we shall see in Proposition 5.30, item 3, the topology $\mathfrak{C}_G$ is not T_1 in general, so $\mathfrak{C}_G \neq \mathfrak{Z}_G$ in general. As $\mathfrak{C}_G \subseteq \mathfrak{Z}_G$, we can still consider the coarsest T_1 topology $\mathfrak{C}'_G$ on G such that

$$\mathfrak{C}_G \subseteq \mathfrak{C}'_G \subseteq \mathfrak{Z}_G.$$

The cofinite topology cof_G being the coarsest T_1 topology on G, in the following definition we introduce the T_1-refinement topology $\mathfrak{C}'_G$ of $\mathfrak{C}_G$.

DEFINITION 5.28. *The* T_1 centralizer topology $\mathfrak{C}'_G$ *on a group G is the supremum (in the lattice of all topologies on G)*

$$\mathfrak{C}'_G = \mathfrak{C}_G \vee cof_G.$$

REMARK 5.29. Then $\mathfrak{C}'_G$ is T_1, and $\mathfrak{C}_G \subseteq \mathfrak{C}'_G \subseteq \mathfrak{Z}_G$, so that $\mathfrak{C}_G = \mathfrak{C}'_G$ if and only if $\mathfrak{C}_G$ is T_1.

Let $W = W_{\mathfrak{C}} \cup \{gx \mid g \in G\}$. Then $\mathfrak{C}'_G = \mathfrak{T}_W$, as

$$\mathcal{E}(W) = \mathcal{E}(W_{\mathfrak{C}}) \cup \{\{g\} \mid g \in G\} = \{gC_G(a) \mid a, g \in G\} \cup \{\{g\} \mid g \in G\},$$

where the second equality follows from (17). Obviously, $\mathfrak{C}'_G = \mathfrak{T}_{W'}$ also for $W' = \{axbx^{-1} \mid a, b \in G\} \cup \{xg \mid g \in G\}$.

In what follows, we denote by ι_G the indiscrete topology on G, namely

$$\iota_G = \{\emptyset, G\} \subseteq \mathcal{P}(G).$$

Here follows some easy-to-establish properties of the centralizer topologies $\mathfrak{C}_G$ and $\mathfrak{C}'_G$.

Proposition 5.30. *Let G be a group. Then the following hold.*

1. *Both the pair $(G, \mathfrak{C}_G)$ and $(G, \mathfrak{C}'_G)$ are quasi-topological groups.*
2. $\overline{\{e_G\}}^{\mathfrak{C}_G} = Z(G)$.
3. *$\mathfrak{C}_G$ is T_1 (so $\mathfrak{C}_G = \mathfrak{C}'_G$) if and only if $Z(G) = \{e_G\}$, while $\mathfrak{C}_G = \iota_G$ if and only if $G = Z(G)$ is abelian.*
4. *If $H \leq G$, then $\mathfrak{C}_H \subseteq \mathfrak{C}_G \restriction_H$ and $\mathfrak{C}'_H \subseteq \mathfrak{C}'_G \restriction_H$.*

Proof. (1) is straightforward.

(2). As $Z(G) = \bigcap_{g \in G} C_G(g)$ is $\mathfrak{C}_G$-closed, one only has to verify that every $\mathfrak{C}_G$-closed subset containing e_G must also contain $Z(G)$.

(3). Immediately follows from items (1) and (2).

(4). To prove that $\mathfrak{C}_H \subseteq \mathfrak{C}_G \restriction_H$, it suffices to note that for every element $h \in H$ we have that $C_H(h) = C_G(h) \cap H$ is a $\mathfrak{C}_G \restriction_H$-closed subset of H.

To prove the inclusion $\mathfrak{C}'_H \subseteq \mathfrak{C}'_G \restriction_H$, note that $cof_H = cof_G \restriction_H$, so that

$$\mathfrak{C}'_H = \mathfrak{C}_H \vee cof_H \subseteq \mathfrak{C}_G \restriction_H \vee cof_G \restriction_H \subseteq (\mathfrak{C}_G \vee cof_G) \restriction_H = \mathfrak{C}'_G \restriction_H .$$

□

We shall see in §5.4.1 that every free non-abelian group F satisfies $\mathfrak{C}_F = \mathfrak{C}'_F = \mathfrak{Z}_F$. On the other hand, we consider a class of matrix groups H in §5.4.2, satisfying $\mathfrak{C}_H \neq \mathfrak{C}'_H = \mathfrak{Z}_H$.

Example 5.31. Let us show that the inclusion $\mathfrak{C}_H \subseteq \mathfrak{C}_G \restriction_H$ in Proposition 5.30, item 4, may be proper. To this end, it will suffice to consider a group G having an abelian, non-central subgroup H, so that

$$\iota_H = \mathfrak{C}_H \subsetneq \mathfrak{C}_G \restriction_H .$$

Indeed, $\iota_H = \mathfrak{C}_H$ holds by Proposition 5.30, item 3, as H is abelian, while $\emptyset \neq Z(G) \cap H \subsetneq H$ is a $\mathfrak{C}_G \restriction_H$-closed subset of H.

Proposition 5.32. *Let G be a group, $\overline{G} = G/Z(G)$, and τ be the initial topology on G of the canonical projection map*

$$\pi \colon G \to (\overline{G}, cof_{\overline{G}}). \tag{18}$$

Then $\tau \subseteq \mathfrak{C}_G$.

Moreover, $\mathfrak{C}_G = \tau$ if and only if for every $g \in G \setminus Z(G)$ the index $[C_G(g) : Z(G)]$ is finite.

Proof. As the family of singletons of $\overline{G}$ is a subbase for the $cof_{\overline{G}}$-closed sets, and $\pi^{-1}(\{gZ(G)\}) = gZ(G)$ is $\mathfrak{C}_G$-closed for every $g \in G$ by Proposition 5.30, items 1 and 2, we immediately obtain $\tau \subseteq \mathfrak{C}_G$.

For the reverse inclusion, we have that $\mathfrak{C}_G \subseteq \tau$ if and only if $C_G(g)$ is τ-closed for every $g \in G$ by Proposition 5.30, item 1. As $C_G(g) = G$ is certanly τ-closed for an element $g \in Z(G)$, it is sufficient to consider the case when $C_G(g)$ is τ-closed for every $g \in G \setminus Z(G)$.

Finally note that, if $g \in G \setminus Z(G)$, then $G \gneq C_G(g) \geq Z(G)$. So $C_G(g) = \pi^{-1}\big(\pi(C_G(g))\big)$ is τ-closed exactly when $\pi(C_G(g)) = C_G(g)/Z(G)$ is finite. □

Recall that a group G is called an *FC-group* if the index $[G : C_G(F)]$ is finite for every $F \in [G]^{<\omega}$, or equivalently if $[G : C_G(g)]$ is finite for every $g \in G$.

Now we prove that the centralizer topology $\mathfrak{C}_G$ and the Taĭmanov topology $\mathcal{T}_G$ coincide on an FC-group G.

LEMMA 5.33. *If G is an FC-group, then $\mathfrak{C}_G = \mathcal{T}_G$.*

Proof. The inclusion $\mathfrak{C}_G \subseteq \mathcal{T}_G$ holds for every group, so we prove the reverse one. To this end, it suffices to prove that $C_G(F)$ is a $\mathfrak{C}_G$-neighborhood of e_G, for every $F \in [G]^{<\omega}$. So let $F \in [G]^{<\omega}$, and note that $C_G(F)$ is a finite-index subgroup as G is an FC-group. As $(G, \mathfrak{C}_G)$ is a quasi-topological group by Proposition 5.30, item 1, we can apply Theorem 5.1 (c) to conclude that $C_G(F)$ is $\mathfrak{C}_G$-open. □

5.4.1. The Zariski topology of free non-abelian groups

Let F be a free non-abelian group, and let

$$\mathcal{B} = \{\{f\}, fC_F(g) \mid f, g \in F\} \subseteq \mathbb{E}_F.$$

By Remark 5.29, the family $\mathcal{B}$ is a subbase for the $\mathfrak{C}'_F$-closed subsets.

PROPOSITION 5.34 ([8, Theorem 5.3]). *Arbitrary intersections of proper elementary algebraic subsets of F are elements of $\mathcal{B}^\cup$.*

In the original statement of Proposition 5.34, the authors used the family

$$\{\{f\}, fC_F(g)h \mid f, g, h \in F\}$$

instead of $\mathcal{B}$. Recall that $fC_F(g)h = fhC_F(h^{-1}gh)$, so that really the two families coincide.

THEOREM 5.35. *If F is a free non-abelian group, then $\mathfrak{C}_F = \mathfrak{C}'_F = \mathfrak{Z}_F$.*

Proof. It trivially follows from Proposition 5.34 that $\mathbb{E}_F \subseteq \mathcal{B}^\cup$, so that $\mathcal{B} \subseteq \mathbb{E}_G$ yields $\mathbb{E}_F^\cup = \mathcal{B}^\cup$. Then $\mathfrak{Z}_F = \mathfrak{C}'_F$, while $\mathfrak{C}_F = \mathfrak{C}'_F$ holds by Proposition 5.30, item 3. □

5.4.2. The Zariski topology of Heisenberg groups

If n is a positive integer, and K is an infinite field, the *n-th Heisenberg group with coefficients in K* is the matrix group

$$H = H(n,K) = \\ = \left\{ \begin{pmatrix} 1 & x_1 & \cdots & x_n & y \\ & 1 & & 0 & z_1 \\ 0 & & \ddots & & \vdots \\ & 0 & & 1 & z_n \\ 0 & & 0 & & 1 \end{pmatrix} \in \mathrm{GL}_{n+2}(K) \mid x_1, \dots, x_n, z_1, \dots, z_n, y \in K \right\}.$$

As $Z(H) \cong K$ is not trivial, Proposition 5.30, item 3, implies $\mathfrak{C}_H \neq \mathfrak{C}'_H$.

In [16], we have computed the Zariski topology of $H(1,K)$. It follows from [16, Remark 6.9] that $\mathfrak{Z}_{H(1,K)} = \mathfrak{C}'_{H(1,K)}$ when $\operatorname{char} K \neq 2$.

If $\operatorname{char} K \neq 2$, then it can also be proved using the same techniques that $\mathfrak{C}'_H = \mathfrak{Z}_H$, so that

$$\mathfrak{C}_H \neq \mathfrak{C}'_H = \mathfrak{Z}_H$$

for every $n \in \mathbb{N}_+$.

Aknowledgements

We gratefully thank Dikran Dikranjan for the valuable corrections and suggestions on the first draft of this paper. We also thank the referee for the careful reading and the useful comments.

References

[1] J. F. Adams, *Lectures on Lie groups*, W. A. Benjamin Inc., New York-Amsterdam (1969), xii+182.

[2] A. Arhangel′skiĭ and M. Tkachenko, *Topological groups and related structures*, Atlantis Studies in Mathematics, 1. Atlantis Press, Paris; World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ (2008), xiv+781.

[3] T. Banakh, I. Guran, and I. Protasov, *Algebraically determined topologies on permutation groups*, Topology Appl. **159** (2012), 2258–2268.

[4] T. Banakh and I. Protasov, *Zariski topologies on groups*, preprint (http://arxiv.org/abs/1001.0601).

[5] T. Banakh, I. Protasov, and O. Sipacheva, *Topologization of sets endowed with an action of a monoid*, Topology Appl., http://dx.doi.org/10.1016/j.topol.2014.02.041.

[6] G. Baumslag, A. Myasnikov, and V. Remeslennikov, *Algebraic geometry over groups. I: Algebraic sets and ideal theory*, J. Algebra **219** (1999), 16–79.

[7] R.M. Bryant, *The verbal topology of a group*, J. Algebra **48** (1977), 340–346.

[8] I.M. Chiswell and V. Remeslennikov, *Equations in free groups with one variable. I*, J. Group Theory **3** (2000), 445–466.

[9] S. Dierolf and U. Schwanengel, *Un exemple d'un groupe topologique q-minimal mais non precompact*, Bull. Sci. Math. (2) **101:3** (1977), 265–269.

[10] D. Dikranjan and A. Giordano Bruno, *Arnautov's problems on semitopological isomorphisms*, Applied General Topology **10** (2009), 85–119.

[11] D. Dikranjan and D. Shakhmatov, *Selected topics from the structure theory of topological groups*, Open problems in topology II, E. Perl. Ed., Elsevier (2007), 389–406.

[12] D. Dikranjan and D. Shakhmatov, *Reflection principle characterizing groups in which unconditionally closed sets are algebraic*, J. Group Theory **11** (2008), 421–442.

[13] D. Dikranjan and D. Shakhmatov, *The Markov-Zariski topology of an abelian group*, J. Algebra **324** (2010), 1125–1158.

[14] D. Dikranjan and D. Toller, *The universal exponents of a group*, work in progress.

[15] D. Dikranjan and D. Toller, *Markov's problems through the looking glass of Zariski and Markov topologies*, Ischia Group Theory 2010, Proc. of the Conference, World Scientific Publ. Singapore (2012), 87–130.

[16] D. Dikranjan and D. Toller, *The Markov and Zariski topologies of some linear groups*, Topology Appl. **159** (2012), 2951–2972.

[17] D. Dikranjan and D. Toller, *Productivity of the Zariski topology on groups*, Comment. Math. Univ. Carolin. **54** (2013), 219–237.

[18] E.D. Gaughan, *Topological group structures of infinite symmetric groups*, Proc. Nat. Acad. Sci. U.S.A. **58** (1967), 907–910.

[19] G. Lukács, *Report on the open problems session*, Topology Appl. **159** (2012), 2476–2482.

[20] A.A. Markov, *On unconditionally closed sets*, Comptes Rendus Doklady AN SSSR (N.S.) **44** (1944), 180–181 (in Russian).

[21] A.Yu. Ol′shanskij, *A note on countable non-topologizable groups*, Vestnik Mosk. Gos. Univ. Mat. Mekh. **3** (1980), 103.

[22] A.D. Taĭmanov, *Topologizable groups. II.*, Sibirsk. Mat. Zh. **19** (1978), 1201–1203, 1216 (in Russian; English translation: Siberian Math. J. **19** (1978), 848–850).

Author's address:

Daniele Toller
Dipartimento di Matematica e Informatica,
Università di Udine,
Via delle Scienze, 206
33100 Udine, Italy
E-mail: `toller.daniele@gmail.com`

Received April 12, 2013
Revised April 15, 2014

Rend. Istit. Mat. Univ. Trieste
Volume 46 (2014), 101–121

Dolcher fixed point theorem and its connections with recent developments on compressive/expansive maps [1]

Elisa Sovrano and Fabio Zanolin

"In memory of Professor Mario Dolcher (1920-1997)"

Abstract. *In 1948 Mario Dolcher proposed an expansive version of the Brouwer fixed point theorem for planar maps. In this article we reconsider Dolcher's result in connection with some properties, such as covering relations, which appear in the study of chaotic dynamics.*

Keywords: continuous maps, fixed points, chaotic dynamics, covering relations, snap-back repellers.
MS Classification 2010: 37D45, 37B10, 47H10, 55M20

1. Introduction

The discovery of the so-called complex or chaotic dynamics, about the coexistence of periodic and non-periodic trajectories and sensitive dependence on the initial conditions, is usually attributed to Henri Poincaré [42, 43] (see also [4, 7, 35]) who found very complicated dynamics in his studies of the three body problem. According to Robert May [37, 38], the term *chaos* was introduced in a mathematical context by Li and Yorke in their famous article "Period three implies chaos" [30]. After this paper, thanks also to the preceding work by Stephen Smale on the *horseshoe* [46, 47] (see also [48]) and without forgetting the many other contributions about the so-called *strange attractors* (by Lorenz, Ruelle and Takens, Ueda, just to quote a few names), various formal definitions of chaotic dynamics were proposed (see, for instance [6, page 183], [10, page 127], [15, page 50] and [51, page 57]). A detailed analysis of these definitions, as well as a comparison about different points of view can be found in [5, 8, 9, 11, 21, 24, 34, 45, 50]. In any case, since there are thousands of references about chaotic dynamics as well as many different and interesting points of view on this topic, our discussion is clearly not exhaustive. Although some concepts of chaos may look very different from each other, it is interesting

[1] Paper written under the auspices of INdAM-GNAMPA.

to observe that the notion of chaos according to Li and Yorke is implied by several other definitions. In particular, for a broad class of spaces and mappings, it follows from Devaney's definition and it holds for maps with positive topological entropy.

In 1978 Frederick R. Marotto [31] extended Li–Yorke's approach to higher dimensions, by introducing the concept of *snap-back repeller* for a given map f. One of the key assumptions in this method is the existence of a repulsive fixed point for f. This in particular implies the existence of a neighborhood U of such a point where the expansive property

$$f(U) \supseteq U \tag{1}$$

holds. Maps which are expansive (at least on some parts of their domain) are typical in the context of chaotic dynamics. Conditions of the form (1) or their generalizations (see Section 3) are usually named *covering relations*. They are, in some sense, dual with respect to the assumption $f(U) \subseteq U$ (for U a compact set homeomorphic to a closed ball) of the Brouwer fixed point theorem. It can be interesting to recall that in 1948 Mario Dolcher already considered an expansive version of the Brouwer theorem for the search of fixed points of planar maps [17].

The aim of this paper is to reconsider Dolcher's result in the context of the covering relations. In Section 2 we recall some classical facts related to Brouwer fixed point theorem and we give a comparison with their dual aspect concerning expansive properties. We also show, by means of a counterexample, that the assumptions in Dolcher's theorem are sharp. In Section 3 we survey some results about maps which are compressive/expansive only on some components of the space. With this respect, Dolcher's approach, if applied only to some components of the map, can find its interpretation in the context of the Markov partitions as presented in [52]. The content of this section is based on results obtained in [2, 3, 41, 52]. It is also strictly related to a recent paper by Jean Mawhin [36] where various fixed points theorems for such maps are unified in a generalized setting.
This paper is partially based on two lectures delivered by the authors at the University of Trieste and on the thesis [49].

2. Fixed points and periodic points

2.1. Results related to Brouwer fixed point theorem

Let $||\cdot||$ be a fixed norm in $\mathbb{R}^N$ and let $B_r := \{x \in \mathbb{R}^N : ||x|| \leq r\}$ be the closed ball of center the origin and radius $r > 0$ in $\mathbb{R}^N$. The Brouwer fixed point theorem is one of the most classical and known results about the existence of fixed points for continuous maps in finite dimensional spaces. It can be formally expressed as follows:

Theorem 2.1 (Brouwer). *For any continuous map* $\phi : B_r \to B_r$ *there exists* $\tilde{x} \in B_r$ *such that* $\phi(\tilde{x}) = \tilde{x}$.

Usually, the presentation of this theorem is accompanied by some related results as the following.

Theorem 2.2 (Rothe). *For any continuous map* $\phi : B_r \to \mathbb{R}^N$ *such that* $\phi(\partial B_r) \subseteq B_r$, *there exists* $\tilde{x} \in B_r$ *such that* $\phi(\tilde{x}) = \tilde{x}$.

Theorem 2.3 (Poincaré–Bohl). *For any continuous map* $\phi : B_r \to \mathbb{R}^N$ *such that*

$$\phi(x) \neq \mu x, \quad \forall\, x \in \partial B_r \text{ and } \mu > 1,$$

there exists $\tilde{x} \in B_r$ *such that* $\phi(\tilde{x}) = \tilde{x}$.

Theorem 2.2 and Theorem 2.3, although they are apparently more general than Theorem 2.1, can be easily proven by Brouwer's theorem. Indeed, one can apply it to the continuous map

$$\psi : B_r \to B_r\,, \quad \psi(x) := P_r(\phi(x)),$$

where

$$P_r(y) := \begin{cases} y & y \in B_r \\ \\ r\,\dfrac{y}{||y||} & y \notin B_r \end{cases}$$

is the radial projection of $\mathbb{R}^N$ onto B_r. One can easily check that if $\tilde{x} \in B_r$ is a fixed point of ψ, then the assumptions of Theorem 2.2 or Theorem 2.3 prevent the possibility that $\phi(\tilde{x}) \notin B_r$. Indeed, if $||\phi(\tilde{x})|| > r$, then, from $\psi(\tilde{x}) = \tilde{x}$, it follows that $\tilde{x} \in \partial B_r$ (that contradicts the condition of Rothe theorem) and, moreover, $\phi(\tilde{x}) = \mu\tilde{x}$ with $\mu = ||\phi(\tilde{x})||/r$ (that contradicts the assumption of Poincaré–Bohl theorem). Hence, in any case $\psi(\tilde{x}) \in B_r$ and so $\tilde{x} = \psi(\tilde{x}) = \phi(\tilde{x})$.

An equivalent manner to express the Brouwer fixed point theorem is that of saying that a closed ball in a finite dimensional normed space has the *fixed point property* (FPP). In general, we say that a topological space X has the FPP if any continuous map $f : X \to X$ has at least a fixed point. The FPP is preserved by homeomorphisms, thus, if we define a *m-dimensional cell* as a topological space which is homeomorphic to a closed ball of $\mathbb{R}^m$ (according to [40, page 4]), we can express the Brouwer fixed point theorem as follows.

Theorem 2.4. *For any continuous map* $\phi : C \to C$, *where* C *is a* m*-dimensional cell, there exists* $\tilde{x} \in C$ *such that* $\phi(\tilde{x}) = \tilde{x}$.

All the above versions of the Brouwer theorem, from a geometrical point of view, describe a situation in which the image of a ball (or its boundary) is contained in the ball itself. A dual result would be naturally expected, namely the existence of fixed points when the image of a ball covers it. This is indeed true for *homeomorphisms* in finite dimensional spaces. More precisely, the following result holds.

THEOREM 2.5. *Let $C \subseteq \mathbb{R}^N$ be a m-dimensional cell and let $\phi : C \to \phi(C) \subseteq \mathbb{R}^N$ be a homeomorphism such that*

$$\phi(C) \subseteq C \quad \textit{or} \quad \phi(C) \supseteq C. \tag{2}$$

Then there exists $\tilde{x} \in C$ such that $\phi(\tilde{x}) = \tilde{x}$.

Clearly, we are precisely in the setting of Theorem 2.4 when $\phi(C) \subseteq C$. On the other hand, when $\phi(C) \supseteq C$, we can enter again in the setting of Theorem 2.4 by observing that $C' := \phi(C)$ is a m-dimensional cell and $\phi^{-1} : C' \to C \subseteq C'$ is continuous. Hence there exists $\tilde{x} \in C'$ with $\phi^{-1}(\tilde{x}) = \tilde{x}$, so that $\tilde{x} \in C$ is also a fixed point for ϕ.

2.2. Covering relations for continuous maps

A more interesting problem arises if ϕ is only continuous and not necessarily a homeomorphism. In the one-dimensional case ($N = 1$), using the Bolzano intermediate value theorem we can provide an affirmative answer as follows.

THEOREM 2.6. *Let $I \subseteq \mathbb{R}$ be a compact interval and let $\phi : I \to \mathbb{R}$ be a continuous map such that $\phi(I) \supseteq I$. Then there exists $\tilde{x} \in I$ such that $\phi(\tilde{x}) = \tilde{x}$.*

We can find an application of this result in the classical paper of Li and Yorke "Period three implies chaos" (see [30, Lemma 2]). A second result which plays a crucial role in that paper is the following (see [30, Lemma 1]).

THEOREM 2.7. *Let $f : J \to J$ be a continuous map (where $J \subseteq \mathbb{R}$ is an interval) and let $(I_n)_n$ be a sequence of compact intervals with $I_n \subseteq J$ and $f(I_n) \supseteq I_{n+1}$ for all $n \in \mathbb{N}$. Then there is a sequence of compact intervals Q_n such that $I_0 \supseteq Q_n \supseteq Q_{n+1}$ and $f^n(Q_n) = I_n$ for all $n \geq 0$. For any $x \in Q := \cap_{n=0}^{\infty} Q_n$ we have $f^n(x) \in I_n$ for all n.*

This last result can be extended to a general setting. For example, Marotto, extending Li–Yorke's approach to higher dimensions, used a version of Theorem 2.7 where $J = \mathbb{R}^N$ and $(I_n)_n$ is sequence of nonempty compact sets (see [31, Lemma 3.2]). The same situation has been considered by Kloeden in [23, Lemma 2], referring to Diamond [16, Lemma 1].

A more general version of Theorem 2.7 can be applied in order to prove the existence of arbitrary itineraries for a continuous map and then to obtain chaotic dynamics in the coin–tossing sense, according to [21]. More in detail, let X be a metric space and $f : X \to X$ be a continuous map. Let A_0, A_1 be two nonempty compact and disjoint sets. Following the terminology adopted in [19] we say that an *itinerary* in $\{A_0, A_1\}$ is a sequence of symbol sets $\mathcal{S} := (A_{s_0}, A_{s_1}, \ldots, A_{s_n}, \ldots)$ with $(s_n)_n \in \Sigma_2^+ := \{0,1\}^{\mathbb{N}}$. A point $x \in A_0 \cup A_1$ is said to *follow the itinerary* $\mathcal{S}$ iff$f^n(x) \in A_{s_n}$ for all n. If, moreover, A_0 and A_1 satisfy condition

$$f(A_i) \supseteq f(A_j),\ \forall\, i,j \in \{0,1\}, \tag{3}$$

then it holds that all itineraries in $\{A_0, A_1\}$ are followed. The meaning of this result can be explained as follows: given any prescribed sequence of two symbols, for instance a sequence of *Heads* $= 1$ and *Tails* $= 0$, there is a forward orbit $(x_n)_n$ for f, i.e. $x_{n+1} = f(x_n)$, such that $x_n \in A_1$ or $x_n \in A_0$ according to the fact that $s_n =$ *Head* or $s_n =$ *Tail*. In other words, the deterministic map f is able to reproduce any outcome of a general coin flipping experiment (see [48]). We can derive many consequences from (3) which are relevant for the theory of chaotic dynamics, like the existence of a compact invariant set on which the map (or some of its iterates) is semiconjugate to the Bernoulli shift, or the positive topological entropy of f, or also the existence of ergodic invariant measures (see [5, 11, 28, 29, 44]).

In several applications, for instance when dealing with dynamical systems induced by the Poincaré map associated to a system of differential equations, it would be quite interesting to provide information also with respect to the existence of fixed points or periodic points. In this context, given a *periodic itinerary* $\mathcal{S}$, a natural question which arises is whether there exists a *periodic point* for f which follows it. In the one-dimensional setting and for A_0, A_1 compact intervals, a positive answer can be provided by applying Theorem 2.6 together with Theorem 2.7. Indeed, already in [30], periodic points of every period were found in the one-dimensional case. Extensions to higher dimensions of the covering relation, in the spirit of Li and Yorke paper, in order to provide the existence of infinitely many periodic points have been obtained by Marotto [31, 32], Kloeden [23] (see also [22]). The extension of Theorem 2.6 in the appropriate setting is made possible by assuming that the map is a homeomorphism restricted to suitable regions of its domain. In view of the above discussion, the question whether the assumption of local homeomorphism can be relaxed to the only hypothesis of continuity seems to be of some interest. As a first step in this direction, we shall focus our attention to the search of fixed points for *continuous* maps which satisfy expansivity conditions or covering relations of some kind.

2.3. Dolcher's fixed point theorem

In [17] Dolcher proposed a result of fixed points for planar maps which satisfy a covering relation. More precisely, denoting by $i(P,C)$ the topological index of a closed curve C with respect to a point $P \notin C$, the following fixed point theorem holds.

THEOREM 2.8 (Dolcher). *For $N = 2$, let $\phi : B_r \to \mathbb{R}^2$ be a continuous map such that the curve $\phi(\partial B_r)$ is external to the disc B_r and, moreover,*

$$i(P, \phi(\partial B_r)) \neq 0$$

for the points $P \in B_r$. Then, there exists $\tilde{x} \in B_r$ such that $\phi(\tilde{x}) = \tilde{x}$.

A more general version of Theorem 2.8 consists in assuming that the curve $\phi(\partial B_r)$ has no points in the interior of the disc B_r and that $i(P, \phi(\partial B_r)) \neq 0$ for the points $P \in \mathrm{int}B_r$ (see [17, Teorema II]). Theorem 2.8, as well as its variant, is expressed in the planar setting, however, as already observed by the author in the introduction of [17], the result can be extended to any dimension. Indeed, using the Brouwer degree on the open ball

$$\Omega_r := \mathrm{int}B_r,$$

we can state the following.

THEOREM 2.9. *Let $\phi : B_r \to \mathbb{R}^N$ be a continuous map such that $\phi(\partial B_r) \subseteq \mathbb{R}^N \setminus \Omega_r$ and suppose that $\deg(\phi, \Omega_r, 0) \neq 0$. Then, there exists $\tilde{x} \in B_r$ such that $\phi(\tilde{x}) = \tilde{x}$.*

Proof. If there exists $\tilde{x} \in \partial B_r$ such that $\phi(\tilde{x}) = \tilde{x}$, we have the result. Therefore, we can suppose that $\phi - Id$ never vanishes on $\partial\Omega_r$. Hence, if we define the homotopy $h_\lambda(x) := \phi(x) - \lambda x$, for $x \in B_r$ and $\lambda \in [0,1]$, we easily find that $h_\lambda(x) \neq 0, \forall\, x \in \partial B_r$. In fact $||\phi(x)|| \geq r$ for all $x \in \partial B_r$, while $||\lambda x|| < r$ for all $x \in \partial B_r$ and $0 \leq \lambda < 1$. By the homotopic invariance of the topological degree, we have $\deg(\phi - Id, \Omega_r, 0) = \deg(\phi, \Omega_r, 0) \neq 0$ and thus we conclude that there exists $\tilde{x} \in \Omega_r$ such that $\phi(\tilde{x}) - \tilde{x} = 0$. Hence, in any case, there exists a fixed point for ϕ in B_r. $\square$

Notice that from the assumption $\phi(\partial B_r) \subseteq \mathbb{R}^N \setminus \Omega_r$ it follows that

$$\deg(\phi, \Omega_r, 0) = \deg(\phi, \Omega_r, P), \quad \forall\, P \in \Omega_r\,.$$

Indeed, for every $P \in \Omega_r$ the homotopy

$$h_\lambda(x) := \phi(x) - \lambda P, \quad \lambda \in [0,1]$$

is admissible (i.e., the sets of zeros of h_λ in B_r is contained in Ω_r). By the same argument we can also prove that the degree condition on ϕ implies

$$\phi(B_r) \supseteq B_r \tag{4}$$

and then we can say that Dolcher's theorem provides an example of a covering relation for continuous maps (not necessarily homeomorphisms) which is accompanied by the existence of fixed points. To check (4), let us consider an arbitrary point $P \in B_r$. If $P \in \Omega_r$, it holds that $\deg(\phi, \Omega_r, P) \neq 0$ and therefore $P \in \phi(\Omega_r)$. Hence, we can suppose $P \in \partial B_r$. If $P \in \phi(\partial B_r)$, we are done. Otherwise, the whole segment $[0, P] = \{\lambda P : 0 \leq \lambda \leq 1\}$ is disjoint from $\phi(\partial B_r)$ and we can conclude as before by the same homotopy h_λ.

2.4. Remarks on the nonexistence of fixed points

In this section we show, by means of some examples, that the hypotheses of Dolcher's theorem are sharp.

First of all, we observe that there is a clear asymmetry in the statements of Theorem 2.1 and Theorem 2.9 due to the additional degree condition in the latter result. One can provide simple cases of nonexistence of fixed points if the degree hypothesis is not satisfied while $\phi(\partial B_r) \cap \Omega_r = \emptyset$ holds. An example is already described in [36, Remark 3] (see also [2, Example 1] for a different context) and the function involved is any translation of the form $\phi : \mathbb{R}^N \to \mathbb{R}^N$, $\phi(x) := x + \vec{v}$, with $||\vec{v}|| \geq 2r$. In this situation it is evident that the degree condition fails. Moreover ϕ is a homeomorphism with $\phi(B_r) \cap \Omega_r = \emptyset$ and also the covering relation in Theorem 2.5 fails. So, it could be interesting to provide an example in which a covering relation as (4) is satisfied with zero degree.

A possible step in this direction can be described as follows. Let $\mathcal{A} := [-1, 1]^2$ be the unit square in $\mathbb{R}^2$ and let

$$\mathcal{B} := (\,[-6, 6] \times [-2, 10]\,) \setminus (\,]-2, 2[\, \times\,]2, 6[\,)$$

be an annular region containing $\mathcal{A}$ in its interior. We define a continuous map $\phi : \mathcal{A} \to \phi(\mathcal{A}) = \mathcal{B}$ by gluing together three homeomorphisms g_1, g_2, g_3 defined as follows. The (continuous) map g_1 is defined on the rectangle $\mathcal{A}_1 := [-1, -1/3] \times [-2, 2]$ and maps its domain homeomorphically onto the pluri-rectangle $\mathcal{B}_1 := (\,[0, 6] \times [-2, 2]\,) \cup (\,[2, 6] \times [-2, 10]\,)$. It is possible to define g_1 in such a way that the part of the boundary $\partial\mathcal{A} \cap \partial\mathcal{A}_1$ is transformed onto $\partial\mathcal{B}_1 \setminus (\{2\} \times]6, 10[\,)$. In a symmetric manner, we define the (continuous) map g_3 as a homeomorphism of the rectangle $\mathcal{A}_3 := [1/3, 1] \times [-2, 2]$ onto the pluri-rectangle $\mathcal{B}_3 := (\,[-6, 0] \times [-2, 2]\,) \cup (\,[-6, -2] \times [-2, 10]\,)$. Analogously, g_3 is such that the part of the boundary $\partial\mathcal{A} \cap \partial\mathcal{A}_3$ is transformed onto $\partial\mathcal{B}_3 \setminus (\{-2\} \times]6, 10[\,)$. Finally, the (continuous) map g_2 is defined on the rectangle $\mathcal{A}_2 := [-1/3, 1/3] \times [-2, 2]$ and maps its domain homeomorphically onto the square $\mathcal{B}_2 := [-2,2] \times [6,10]$. We define g_2 in such a way that the part of the boundary $\partial\mathcal{A} \cap \partial\mathcal{A}_2$ is transformed onto $\partial\mathcal{B} \cap \partial\mathcal{B}_2$. The geometric construction for the resulting piecewise homeomorphism is sketched in Figure 1.

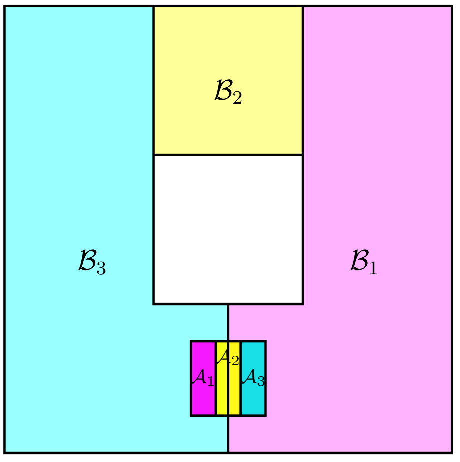


Figure 1: Action of the map ϕ defined on the square $\mathcal{A}$ onto the annular region $\mathcal{B}$.

If the g_i's are glued together correctly, a continuous function ϕ such that $\phi(\mathcal{A}) = \mathcal{B} \supseteq \mathcal{A}$ can be exhibited. Moreover $\phi(x) \neq x$, for each $x \in \mathcal{A}$, since $g_i(\mathcal{A}_i) \cap \mathcal{A}_i = \emptyset$ for $i \in \{1,2,3\}$. We also have $\deg(\phi, \mathrm{int}\mathcal{A}, 0) = 0$. Indeed, while a point P moves on the boundary of $\mathcal{A}$ in the counterclockwise sense, its image $\phi(P)$ makes a loop along $\partial B \cup \Gamma$, where $\Gamma := \{0\} \times [-2,2]$. In fact, the point $\phi(P)$ moves on the vertical segment twice in opposite directions.

The example above, even if it goes well with regard to the properties of covering, is not suitable in the context of Dolcher's theorem, because of the nonempty intersection $\phi(\partial\mathcal{A}) \cap \mathcal{A} = \Gamma$. Therefore, we provide a more elaborate construction of a continuous map which expands a smaller disk to a larger concentric one, it sends the boundary onto the boundary, but it has no fixed points. In this way, the following result holds.

Proposition 2.10. *For $N \geq 2$ and given $0 < r < R$, there exists a continuous map $\phi : B_r \to B_R$ satisfying*

$$\phi(B_r) = B_R\,, \quad \phi(\partial B_r) = \partial B_R \tag{5}$$

and without fixed points.

Proof. We start by proving the claim for $N = 2$. The trick of the proof is to find two rectangles

$$\mathcal{R} := [-a_0, a_0] \times [-b_0, b_0], \quad \mathcal{R}' := [-a_1, a_1] \times [-b_1, b_1], \tag{6}$$

with

$$0 < b_0 < a_0 \quad \text{and} \quad a_1 = ka_0\,, b_1 = kb_0\,, \quad \text{for some } k > 1 \tag{7}$$

and define a continuous map $f : \mathcal{R} \to \mathbb{R}^2$ (actually a piecewise homeomorphism) such that

- $f(\mathcal{R}) = \mathcal{R}'$,
- $f(p) \neq p$, $\forall\, p \in \mathcal{R}$,
- $f(\partial\mathcal{R}) \subseteq \mathbb{R}^2 \setminus \mathcal{R}$,
- $\{t\vec{v} : t > 0\} \cap f(\partial\mathcal{R}) \neq \emptyset$, $\forall\, \vec{v} \in S^1$.

The rectangle $\mathcal{R}$ is the closed unit disc for the norm

$$|||(x, y)||| = \max\{|x|/a_0, |y|/b_0\}$$

and, consequently, $\mathcal{R}' = \{z \in \mathbb{R}^2 : |||z||| \leq k\}$. Once we have introduced the map f, we can define $\phi_0 : B_1 \to B_k$ as

$$\phi_0(z) := h^{-1}(f(h(z))), \quad \forall\, z \in B_1\,,$$

where $h : \mathbb{R}^2 \to \mathbb{R}^2$ is the homeomorphism such that

$$h(z) := \begin{cases} \frac{||z||}{|||z|||}\, z & \text{for } z \neq 0, \\ 0 & \text{for } z = 0. \end{cases}$$

The continuous function ϕ_0 maps B_1 onto B_k , with $\phi_0(\partial B_1) \subseteq \mathbb{R}^2 \setminus B_1$ and, moreover, any half-ray from the origin meets $\phi_0(\partial B_1)$. Now, we take $\delta \in]1, k[$ such that $\phi_0(\partial B_1) \subseteq \mathbb{R}^2 \setminus B_\delta$ and define

$$\phi_1 := P_\delta \circ \phi_0\,,$$

where P_δ is the projection of $\mathbb{R}^2$ onto B_δ . In this manner, we have a fixed point free continuous map ϕ_1 of B_1 onto B_δ such that $\phi_1(\partial B_1) = \partial B_\delta$. The definition of the map $\phi : B_r \to B_R$ immediately follows from that of $\phi_1 : B_1 \to B_\delta$, by a suitable rescaling.

With this in mind, our problem reduces to the construction contained in the following example.

Example 2.11. We define f with respect to the rectangles $\mathcal{R}$ and $\mathcal{R}'$ defined as in (6)-(7) with

$$a_0 := 6, \quad b_0 := 2, \quad k := 2.$$

We evenly divide the rectangle $\mathcal{R}$ into six closed sub-rectangles $\mathcal{R}_1, \ldots, \mathcal{R}_6$, as described in Figure 2.

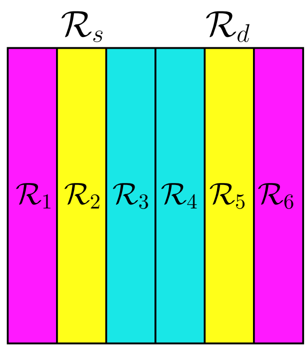


Figure 2: Subdivision of the rectangle $\mathcal{R}$. We have denoted by $\mathcal{R}_s := \mathcal{R}_1 \cup \mathcal{R}_2 \cup \mathcal{R}_3 = [-6,0] \times [-2,2]$ the left half of $\mathcal{R}$ and, analogously, by $\mathcal{R}_d$ its right hand half.

On each rectangle $\mathcal{R}_i$ we define a continuous map g_i, which is indeed a homeomorphism of $\mathcal{R}_i$ onto $g_i(\mathcal{R}_i) =: \mathcal{R}'_i$.

First of all, we introduce the map $g_s : \mathcal{R}_s \to \mathcal{R}'_1 := [-2, 12] \times [-4, 4]$ which is obtained by the gluing of the continuous maps g_1, g_2, g_3. In a symmetric manner, one defines a continuous map $g_d : \mathcal{R}_d \to \mathcal{R}'_6 := [-12, 2] \times [-4, 4]$ and, finally f is the result of the pasting lemma applied to g_s and g_d. At each step, we carefully avoid the presence of fixed points.

The map $g_1 : [-6,-4] \times [-2,2] =: \mathcal{R}_1 \to g_1(\mathcal{R}_1) = \mathcal{R}'_1$ is defined as

$$g_1(x,y) := (-7x - 30, 2y).$$

The sides $\ell_1, \ell_2, \ell_3, \ell_4$ of $\mathcal{R}_1$ are mapped by g_1 onto the sides $\ell'_1, \ell'_2, \ell'_3, \ell'_4$ of $\mathcal{R}'_1$ as in Figure 3. By construction, $g_1(p) \neq p$, for each $p \in \mathcal{R}_1$.

The map g_2 transforms the rectangle $\mathcal{R}_2 := [-4,-2] \times [-2,2]$ onto a Jordan domain $\mathcal{R}'_2$ which is bounded by a simple closed curve that we describe as follows.

We denote by $\ell_5, \ell_6, \ell_7, \ell_8$ the sides of $\mathcal{R}_2$ (traversed counterclockwise) and by $\ell'_i := g_2(\ell_i)$ their images (which are traversed in a counterclockwise manner, too). With this notation, first of all, we impose that

$$g_2(z) = g_1(z), \quad \forall\, z \in \ell_3 = \ell_5 = \mathcal{R}_1 \cap \mathcal{R}_2 .$$

In order to properly define g_2, it is important to notice that a point which moves in the up down direction along the segment ℓ_5 is transformed by g_1 to a point which moves in the same direction along the segment ℓ'_5.

Next, we take as ℓ'_7 the segment $\{11\} \times [5/2, 3]$ and, finally, we take as ℓ'_6 and ℓ'_8 two simple arcs contained in $\mathcal{R}'_1 \setminus \mathcal{R}$. A possible choice of ℓ'_6 is given by an arc of cubic with $(-2,-4)$ and $(11, 5/2)$ as endpoints and passing through $(0, 7/2)$

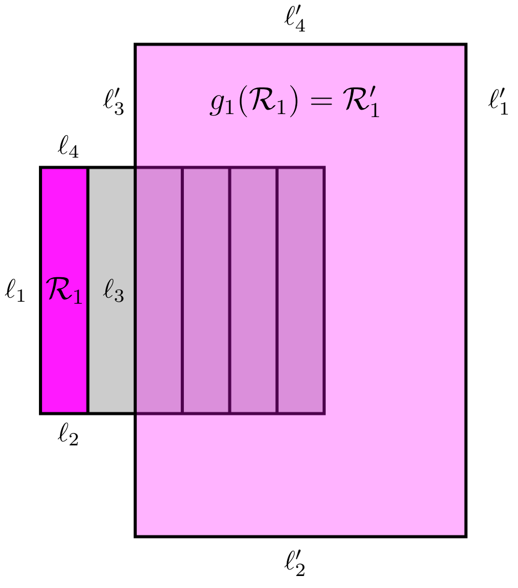


Figure 3: Transformation of the sub-rectangle $\mathcal{R}_1$ onto $\mathcal{R}_1'$ by g_1. The part of the boundary of $\mathcal{R}_1$ given by $\ell_4\ell_1\ell_2$ (traversed counterclockwise) is transformed into $\ell_4'\ell_1'\ell_2'$ (in the clockwise sense).

and $(13/2, -2)$. For ℓ_8' we have taken an arc of hyperbola with endpoints $(11, 3)$ and $(-2, 4)$ and passing through $(6, 7/2)$. The specific definition of g_2 is given by

$$g_2(x, y) := \left(\tfrac{13}{2}x + 24, \tfrac{2-y}{4} h_1\left(\tfrac{13}{2}x + 24\right) + \tfrac{2+y}{4} h_2\left(\tfrac{13}{2}x + 24\right)\right),$$

where

$$h_1(x) := -\frac{7}{2} + \frac{1685}{9724}x - \frac{265}{9724}x^2 + \frac{27}{4862}x^3,$$
$$h_2(x) := \frac{\sqrt{31678 - 859x - 29x^2}}{4\sqrt{130}}.$$

In Figure 4 we show $\mathcal{R}_2$ and its image $\mathcal{R}_2'$. The map g_2 is a homeomorphism. Moreover, $\mathcal{R}_2 \cap \mathcal{R}_2' = \ell_7$ and $g_2(\ell_7) = \ell_7' \cap \ell_7 = \emptyset$. Hence, also g_2 is fixed point free.

The map $g_3 : [-2, 0] \times [-2, 2] =: \mathcal{R}_3 \to g_3(\mathcal{R}_3) = \mathcal{R}_3' := [0, 11] \times [5/2, 3]$ is defined as

$$g_3(x, y) := \left(-\tfrac{11}{2}x, \tfrac{1}{8}y + \tfrac{11}{4}\right).$$

The sides $\ell_9, \ell_{10}, \ell_{11}, \ell_{12}$ of $\mathcal{R}_3$ are mapped by g_3 onto the sides $\ell_9', \ell_{10}', \ell_{11}', \ell_{12}'$

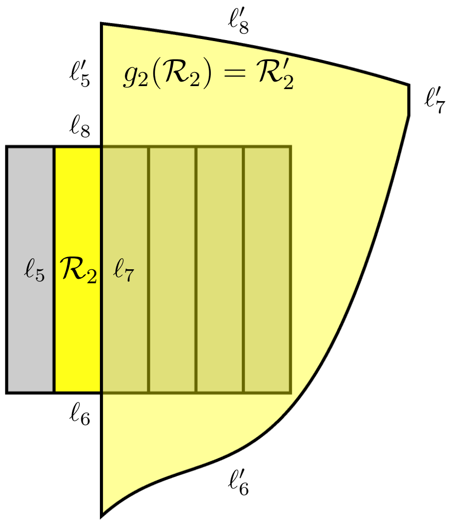


Figure 4: Transformation of the sub-rectangle $\mathcal{R}_2$ onto $\mathcal{R}_2'$ by g_2.

of $\mathcal{R}_3'$ as in Figure 5. By construction,

$$g_2(z) = g_3(z), \quad \forall\, z \in \ell_7 = \ell_9 = \mathcal{R}_2 \cap \mathcal{R}_3$$

and, moreover, $g_3(p) \neq p$, for each $p \in \mathcal{R}_3$.

Gluing together $g_1\,, g_2\,, g_3$ we obtain the map

$$g_s(x,y) := \begin{cases} g_1(x,y) & \text{for } -6 \leq x < -4, \\ g_2(x,y) & \text{for } -4 \leq x < -2, \\ g_3(x,y) & \text{for } -2 \leq x \leq 0 \end{cases}$$

which maps the rectangle $\mathcal{R}_s = [-6,0] \times [-2,2]$ onto $\mathcal{R}_1'$.

Using the symmetry $S_x : (x,y) \mapsto (-x,y)$, we can define the continuous map $g_d : \mathcal{R}_d \to \mathcal{R}_6'$ as

$$g_d(z) := S_x(g_s(S_x(z))), \quad \forall\, z \in \mathcal{R}_d = [0,6] \times [-2,2]$$

and then

$$f(z) := \begin{cases} g_s(z) & \text{for } z \in \mathcal{R}_s\,, \\ g_d(z) & \text{for } z \in \mathcal{R}_d\,. \end{cases}$$

See Figure 6 for a visualization of deformation of $\mathcal{R}$ through the final map.

At this point it is easy to check that f maps continuously $\mathcal{R}$ onto $\mathcal{R}'$, without fixed points and with $f(\partial\mathcal{R})$ a loop external to $\mathcal{R}$ which intersects any ray starting from the origin. $\triangleleft$

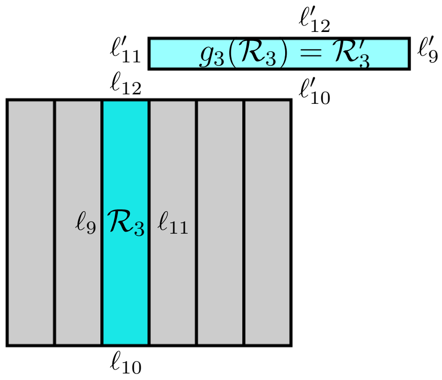


Figure 5: Transformation of the sub-rectangle $\mathcal{R}_3$ onto $\mathcal{R}'_3$ by g_3. The boundary of $\mathcal{R}_3$ given by $\ell_9\ell_{10}\ell_{11}\ell_{12}$ (traversed counterclockwise) is transformed into $\ell'_9\ell'_{10}\ell'_{11}\ell'_{12}$ (in the clockwise sense).

This example concludes the proof for $N = 2$ and now we consider an arbitrary dimension $N \geq 3$.

Applying Proposition 2.10 for $N = 2$ and the $||\cdot||_\infty$-norm, there exists a continuous and surjective planar map

$$\psi : [-r, r]^2 \to [-R, R]^2 \quad \text{with} \;\; \psi(x_1, x_2) = (\psi_1(x_1, x_2), \psi_2(x_1, x_2)),$$

without fixed points and such that $\psi(\partial[-r, r]^2) = \partial[-R, R]^2$. Next, we define $\Psi : [-r, r]^N \to [-R, R]^N$ by

$$\Psi(x_1, x_2, x_3, \dots, x_N) := \big(\psi_1(x_1, x_2), \psi_2(x_1, x_2), \tfrac{R}{r}x_3, \dots, \tfrac{R}{r}x_N\big)\,.$$

Clearly, also Ψ is continuous, surjective, without fixed points and maps the boundary onto the boundary. Finally, we set

$$\phi(z) := h^{-1}(\Psi(h(z))), \quad \forall\, z \in B_r\,,$$

where $h : \mathbb{R}^N \to \mathbb{R}^N$ is the homeomorphism defined by

$$h(z) := \begin{cases} \frac{||z||}{||z||_\infty}\, z & \text{for } z \neq 0, \\ 0 & \text{for } z = 0. \end{cases}$$

This concludes our proof. □

Remark 2.12. *With the same approach, one can obtain a version of Proposition 2.10 in the case $r = R$, for a continuous map ϕ satisfying* (5) *and such*

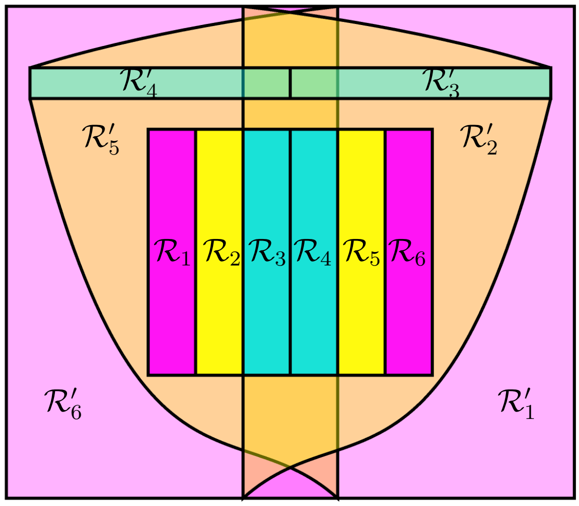


Figure 6: Action of the map f on $\mathcal{R}$.

that $\phi(z) \neq z$ for each $z \in \Omega_r$. With this respect, it may be interesting to recall a result by Brown and Greene [12, Theorem 1] where it is proved (for $N = 2$ and the Euclidean norm) that given a continuous map $f : \partial B_r \to \partial B_r$ with at least one fixed point, there exists a continuous map $\phi : B_r \to B_r$ which extends f and without fixed points on Ω_r. One can also see that if f is surjective, then ϕ is surjective, too.

In the two-dimensional case (which was the original setting in [17]) our construction can be used to provide a counterexample to Theorem 2.5 for maps which are only continuous when the second instance in (2) holds.

Proposition 2.13. *Let $C \subseteq \mathbb{R}^2$ be a 2-dimensional cell. There exists a continuous map $\psi : C \to \psi(C) =: D$, where $D \subseteq \mathbb{R}^2$ is a 2-dimensional cell such that $C \subseteq \mathrm{int} D$ and $\psi(\partial C) = \partial D$, with $\psi(z) \neq z$, $\forall\, z \in C$.*

Proof. Let $C \subseteq \mathbb{R}^2$ be a 2-dimensional cell and let $h : C \to h(C) = B_1$ be a homeomorphism, which exists by definition and is such that $h(\partial C) = \partial B_1$. Using the Schoenflies theorem (see [39]) we extend h to a homeomorphism $\tilde{h}$ of the whole plane. Let $\phi : B_1 \to B_2$ be a continuous map as in Proposition 2.10 for $r = 1$ and $R = 2$. Taking $\psi := \tilde{h}^{-1} \circ \phi \circ \tilde{h}$ and $D := \tilde{h}^{-1}(B_2)$, we achieve the thesis. □

3. Compressive/expansive maps

The search of fixed points for continuous maps which have, at the same time, a compressive and an expansive property is a field of research which has gradually attracted increasing interest in recent years. Besides that there be an intrinsic interest for this type of problems (see [18, 26, 27, 33]), motivations come from the study of Markov partitions and their generalizations (see [52]), as well as from the researches about topological horseshoes [13, 19, 20]. Another natural motivation comes from the search of periodic solutions for nonautonomous differential systems which are dissipative only with respect to some components of the phase space (see [2] and the references therein, as well as [1, 3, 25]).

The study of this type of maps leads to consider, as in [52], some generalized rectangles which are homeomorphic to the product of two closed balls whose dimensions correspond to the dimensions of the compressive and expansive directions, respectively. More formally, we consider two nonnegative integers $u = u(N)$ and $s = s(N)$ with $u + s = N$ and decompose the vector space $\mathbb{R}^N$ as $\mathbb{R}^u \times \mathbb{R}^s$, with the canonical projections

$$p_u : \mathbb{R}^N \to \mathbb{R}^u,\ p_u(x,y) = x,\ \ p_s : \mathbb{R}^N \to \mathbb{R}^s,\ p_s(x,y) = y$$

with $x \in \mathbb{R}^u$ and $y \in \mathbb{R}^s$. We denote by $||\cdot||_u$ and $||\cdot||_s$ two norms in $\mathbb{R}^u$ and $\mathbb{R}^s$, respectively, which are inherited by a given norm $||\cdot||$ in $\mathbb{R}^N$. In other words, we set $||x||_u := ||(x,0)||$ and $||y||_s := ||(0,y)||$. In the sequel, we'll avoid to indicate the subscripts u, s when there is no possibility of misunderstanding. Then we introduce a compact set

$$\mathcal{R}[a,b] := B^u_a \times B^s_b = \{(x,y) \in \mathbb{R}^u \times \mathbb{R}^s \,:\, ||x||_u \leq a,\ ||y||_s \leq b\},$$

where $a, b > 0$ are fixed real numbers. Let $\phi : \mathcal{R}[a,b] \to \mathbb{R}^N$ be a continuous map. We want to think of the number u as a dimension for which the map is expansive and s as a dimension for which the map is compressive, in analogy to the case of the unstable and stable manifolds for the saddle points. For this reason, it is convenient to split ϕ as

$$\phi(x,y) = (\phi_u(x,y), \phi_s(x,y)), \quad \text{with } \ \phi_u : \mathcal{R}[a,b] \to \mathbb{R}^u,\ \phi_s : \mathcal{R}[a,b] \to \mathbb{R}^s,$$

defined as $\phi_u := p_u \circ \phi$ and $\phi_s := p_s \circ \phi$. At this point, a natural approach is to consider the conditions of the theorem of Brouwer (or the Rothe's one) in the compressive direction $s(N)$ and those of the theorem of Dolcher in the expansive direction $u(N)$. This leads to the following result previously considered in [36, Corollary 1], [41, Lemma 1.1], where we denote by $\Omega^u_a = \{x \in \mathbb{R}^u \,:\, ||x|| < a\}$ and $\Omega^s_b = \{y \in \mathbb{R}^s \,:\, ||y|| < b\}$ the interiors of the balls B^u_a and B^s_b, respectively.

Theorem 3.1. *Let $u \geq 1$ and let $\phi = (\phi_u, \phi_s) : \mathcal{R}[a,b] \to \mathbb{R}^N$ be a continuous map such that*

$$\phi_u(\partial B^u_a \times B^s_b) \subseteq \mathbb{R}^u \setminus \Omega^u_a\,, \quad \phi_s(B^u_a \times \partial B^s_b) \subseteq B^s_b\,. \tag{8}$$

Suppose moreover that $\deg(\phi_u(\cdot,0),\Omega^u_a,0)\neq 0$. *Then there exists* $\tilde{z}\in\mathcal{R}[a,b]$ *such that* $\phi(\tilde{z})=\tilde{z}$.

Proof. We combine the proof of Theorem 2.9 with a classical degree argument for the Brouwer fixed point theorem. If there is already a fixed point on the boundary of $\mathcal{R}[a,b]$, we are done. Hence, we suppose that $z\neq\phi(z)$ for all $z\in\partial\mathcal{R}[a,b]$, so that the degree $\deg(Id-\phi,\Omega^u_a\times\Omega^s_b,0)$ is well defined. Next, we consider the homotopy

$$(z,\lambda)\mapsto h_\lambda(z):=(\lambda x-\phi_u(x,\lambda y),y-\lambda\phi_s(x,y)),$$

for $z=(x,y)$ and $\lambda\in[0,1]$. In order to prove that the homotopy is admissible on $\mathcal{R}[a,b]$ (that is $h_\lambda(z)\neq 0$ for each $z\in\partial\mathcal{R}[a,b]$ and $\lambda\in[0,1]$) it is sufficient to check that there are no solutions of the system

$$\begin{cases}\phi_u(x,\lambda y)=\lambda x\\ y=\lambda\phi_s(x,y)\end{cases}\qquad \forall\,\lambda\in[0,1[\ \text{ and }\ z=(x,y)\in\partial\mathcal{R}[a,b].$$

If $(x,y)\in(\partial\Omega^u_a)\times B^s_b$, then also $(x,\lambda y)\in(\partial\Omega^u_a)\times B^s_b$ and hence $||\phi_u(x,\lambda y)||\geq a>||\lambda x||$, so that the first equation in the system has no solutions. If $(x,y)\in B^u_a\times(\partial\Omega^s_b)$, then $||\lambda\phi_s(x,y)||<b=||y||$, so that the second equation in the system has no solutions. By the homotopic invariance of the topological degree we obtain

$$\begin{aligned}\deg(Id-\phi,\Omega^u_a\times\Omega^s_b,0) &= \deg((-\phi_u(\cdot,0),Id|_{\mathbb{R}^s}),\Omega^u_a\times\Omega^s_b,0)\\ &= (-1)^u\deg(\phi_u(\cdot,0),\Omega^u_a,0)\neq 0\end{aligned}$$

and thus we conclude that there exists a $\tilde{z}\in\Omega^u_a\times\Omega^s_b=\mathrm{int}(\mathcal{R}[a,b])$ such that $\tilde{z}-\phi(\tilde{z})=0$. Hence, in any case, there exists a fixed point for ϕ in $\mathcal{R}[a,b]$. □

We have considered the case $u=u(N)\geq 1$, since for $u=0$ the result reduces to Rothe fixed point theorem. On the other hand, Theorem 3.1 reduces to Theorem 2.9 for $u=N$. We refer to [52] and [41] for variants of Theorem 3.1 and we also recommend [36] for recent extensions as well as connections with Poincaré–Miranda theorem.

It may be interesting to investigate whether the assumptions of Theorem 2.9 imply a covering relation analogous to (4) for the expansive component. In fact, we prove the following.

Proposition 3.2. *In the setting of Theorem 2.9 the inclusion*

$$\phi_u(B^u_a\times\{y\})\supseteq B^u_a\,,\quad\forall\,y\in B^s_b \tag{9}$$

holds.

Proof. Let $y \in B^s_b$ be fixed and let $P \in B^u_a$. We claim that there exists $x \in B^u_a$ such that the equation $\phi_u(x,y) = P$ has a solution. If there exists $\tilde{x} \in \partial B^u_a$ with $\phi_u(\tilde{x},y) = P$, we are done. Otherwise, $\phi_u(x,y) \neq P$, $\forall\, x \in \partial B^u_a$ and thus the degree $\deg(\phi_u(\cdot,y), \Omega^u_a, 0)$ is defined. For $\lambda \in [0,1[$ and $x \in \partial B^u_a$ we have (by the second condition in (8)) $||\phi_u(x,\lambda y)|| \geq a > \lambda a \geq ||\lambda P||$. Hence the homotopy $h_\lambda(x) := \phi_u(x,\lambda y) - \lambda P$, for $\lambda \in [0,1]$ and $x \in B^u_a$, is admissible. Therefore,

$$\begin{aligned}&\deg(\phi_u(\cdot,y) - P, \Omega^u_a, 0) = \deg(h_1, \Omega^u_a, 0)\\ &= \deg(h_0, \Omega^u_a, 0) = \deg(\phi_u(\cdot,0), \Omega^u_a, 0) \neq 0\end{aligned}$$

and thus the equation $\phi_u(x,y) = P$ has a solution with $x \in \Omega^u_a$. In any case, P is the image through $\phi_u(\cdot,y)$ of some point in B^u_a. □

The examples in Section 2.4 can be easily adapted to the context of Theorem 2.9. In particular, one can provide examples of continuous maps satisfying (8) and (9), but without fixed points in $\mathcal{R}[a,b]$.

An application of Theorem 3.1 to ordinary differential equations can be described as follows.

Let $F : \mathbb{R} \times \mathbb{R}^N \to \mathbb{R}^N$ be a continuous vector field such that, for some $T > 0$,

$$F(t+T,w) = F(t,w), \quad \forall\, t \in \mathbb{R},\ \forall\, w \in \mathbb{R}^N.$$

Let $\mathcal{D} \subseteq \mathbb{R}^N$ be a nonempty set such that for each $w \in \mathcal{D}$ there is a unique solution of the Cauchy problem

$$\begin{cases}\zeta' = F(t,\zeta)\\ \zeta(0) = w\end{cases}$$

which is defined on $[0,T]$. We denote such a solution by $\zeta(\cdot,w)$. The fundamental theory of ODEs ensures that the Poincaré map

$$\Psi : \mathcal{D} \to \mathbb{R}^N, \quad \Psi(w) := \zeta(T,w)$$

is continuous, actually a homeomorphism of $\mathcal{D}$ onto $\Psi(\mathcal{D})$. The existence of a T-periodic solution $\zeta(t)$ for system

$$\zeta' = F(t,\zeta), \tag{10}$$

with $\zeta(0) \in \mathcal{D}$ is equivalent to the existence of a fixed point (in $\mathcal{D}$) for the map Ψ. In this context, every time we are in the presence of a flow induced by (10) which possesses an expansive property on the first j $(1 \leq j \leq N)$ components and it has a compressive property on the remaining $k := N - j$ components, we can enter into the scheme of Theorem 3.1 with $u = u(N) = j$ and $s = s(N) = k$.

In particular, the condition on the degree may be replaced by a more simple one to control, by adequately exploiting the properties of Poincaré map. In this manner, we obtain the following result where, for notational convenience, we represent a solution $\zeta(t)$ of (10) as

$$\zeta(t) = (E(t), C(t)),$$

with

$$E(t) := (\zeta_1(t), \dots, \zeta_j(t)) \in \mathbb{R}^j, \; C(t) := (\zeta_{j+1}(t), \dots, \zeta_N(t)) \in \mathbb{R}^k.$$

Similarly, we write $\zeta(t, w)$ as $(E(t, w), C(t, w))$, with reference to an initial value problem with $\zeta(0) = w$.

COROLLARY 3.3. *Suppose there exists two positive real numbers a, b such that $\mathcal{R}[a, b] \subseteq \mathcal{D}$ and, moreover, for each $w = (x, y) \in \mathcal{R}[a, b]$, we have:*

(i_1) $\quad ||x|| = a, ||y|| \leq b \Longrightarrow ||E(T, w)|| \geq a,$

(i_2) $\quad ||x|| \leq a, ||y|| = b \Longrightarrow ||C(T, w)|| \leq b,$

(i_3) $\quad ||x|| = a, y = 0 \Longrightarrow ||E(t, w)|| > 0, \quad \forall\, t \in]0, T[\,.$

Then system (10) *has a T-periodic solution $\zeta(t)$ with $\zeta(0) \in \mathcal{R}[a, b]$.*

Proof. As previously observed, we consider the case $u = j$, $s = N - j$ and we apply Theorem 3.1 to the map $\phi := \Psi$, by the obvious decomposition $\phi_u(w) := E(T, w)$, $\phi_s(w) := C(T, w)$, for $w = (x, y)$. With this notation, it is immediate to check that the two assumptions in (8) follow from (i_2) and (i_1), respectively. Condition (i_3) implies $E(\lambda T, (x, 0)) \neq 0$ for every $x \in \partial\Omega^u_a$, for all $\lambda \in]0, 1[\,$. Moreover, $E(\lambda T, (x, 0)) \neq 0$ for $\lambda = 1$ (by (i_1)). For $x \in \partial\Omega^u_a$, also $E(\lambda T, (x, 0)) \neq 0$ for $\lambda = 0$. In fact, $\zeta(0, (x, 0)) = (x, 0)$ and hence $E(0, (x, 0)) = x$. We have thus verified that the homotopy $(x, \lambda) \mapsto h_\lambda(x) := E(\lambda T, (x, 0))$, is admissible on B^u_a. Hence

$$\deg(\phi_u(\cdot, 0), \Omega^u_a, 0) = \deg(h_1, \Omega^u_a, 0) = \deg(h_0, \Omega^u_a, 0) = \deg(Id|_{\mathbb{R}^j}, \Omega^u_a, 0) = 1$$

and so the degree condition in Theorem 3.1 is satisfied. We conclude that there exists a fixed point for Ψ in $\mathcal{R}[a, b]$ that is a T-periodic solution $\zeta(t)$ of (10) with $\zeta(0) \in \mathcal{R}[a, b]$. $\square$

REMARK 3.4. *Corollary 3.3 is contained (in a slightly different form) in [2], which, in turn, is based on the continuation theorems developed in [14]. The assumptions $(i_1), (i_2), (i_3)$ are meaningful independently from the fact that they refer to the components of a Poincaré map. From this point of view, an abstract version of this result concerning the search of fixed points for multivalued maps depending on a parameter $t \in [0, T]$ has been obtained in [3, Theorem 3] and [1, Lemma 3.3, Theorem 3.4]. Applications to differential inclusions have been proposed in [1, 3], as well.*

References

[1] J. Andres, G. Gabor, and L. Górniewicz, *Boundary value problems on infinite intervals*, Trans. Amer. Math. Soc. **351** (1999), no. 12, 4861–4903.

[2] J. Andres, M. Gaudenzi, and F. Zanolin, *A transformation theorem for periodic solutions of nondissipative systems*, Rend. Sem. Mat. Univ. Politec. Torino **48** (1990), no. 2, 171–186.

[3] J. Andres, L. Górniewicz, and M. Lewicka, *Partially dissipative periodic processes*, Topology in nonlinear analysis (Warsaw, 1994), Banach Center Publ., vol. 35, Polish Acad. Sci. Inst. Math., Warsaw, 1996, pp. 109–118.

[4] D. Aubin and A. Dahan Dalmedico, *Writing the history of dynamical systems and chaos: longue durée and revolution, disciplines and cultures*, Historia Math. **29** (2002), no. 3, 273–339.

[5] B. Aulbach and B. Kieninger, *On three definitions of chaos*, Nonlinear Dyn. Syst. Theory **1** (2001), no. 1, 23–37.

[6] J. Auslander and J. A. Yorke, *Interval maps, factors of maps, and chaos*, Tôhoku Math. J. (2) **32** (1980), no. 2, 177–188.

[7] J. Barrow-Green, *Poincaré and the three body problem*, History of Mathematics, vol. 11, American Mathematical Society, Providence, RI; London Mathematical Society, London, 1997.

[8] F. Blanchard, E. Glasner, S. Kolyada, and A. Maass, *On Li-Yorke pairs*, J. Reine Angew. Math. **547** (2002), 51–68.

[9] L. Block, J. Guckenheimer, M. Misiurewicz, and L. S. Young, *Periodic points and topological entropy of one-dimensional maps*, Global theory of dynamical systems (Proc. Internat. Conf., Northwestern Univ., Evanston, Ill., 1979), Lecture Notes in Math., vol. 819, Springer, Berlin, 1980, pp. 18–34.

[10] L. S. Block and W. A. Coppel, *Dynamics in one dimension*, Lecture Notes in Mathematics, vol. 1513, Springer-Verlag, Berlin, 1992.

[11] A. Blokh and E. Teoh, *How little is little enough?*, Discrete Contin. Dyn. Syst. **9** (2003), no. 4, 969–978.

[12] R. F. Brown and R. E. Greene, *An interior fixed point property of the disc*, Amer. Math. Monthly **101** (1994), no. 1, 39–47.

[13] K. Burns and H. Weiss, *A geometric criterion for positive topological entropy*, Comm. Math. Phys. **172** (1995), no. 1, 95–118.

[14] A. Capietto, J. Mawhin, and F. Zanolin, *Continuation theorems for periodic perturbations of autonomous systems*, Trans. Amer. Math. Soc. **329** (1992), no. 1, 41–72.

[15] R. L. Devaney, *An introduction to chaotic dynamical systems*, second ed., Addison-Wesley Studies in Nonlinearity, Addison-Wesley Publishing Company, Advanced Book Program, Redwood City, CA, 1989.

[16] P. Diamond, *Chaotic behaviour of systems of difference equations*, Internat. J. Systems Sci. **7** (1976), no. 8, 953–956.

[17] M. Dolcher, *Due teoremi sull'esistenza di punti uniti nelle trasformazioni piane continue*, Rend. Sem. Mat. Univ. Padova **17** (1948), 97–101.

[18] J. Kampen, *On fixed points of maps and iterated maps and applications*, Nonlinear Anal. **42** (2000), no. 3, Ser. A: Theory Methods, 509–532.

[19] J. Kennedy, S. Koçak, and J. A. Yorke, *A chaos lemma*, Amer. Math. Monthly **108** (2001), no. 5, 411–423.
[20] J. Kennedy and J. A. Yorke, *Topological horseshoes*, Trans. Amer. Math. Soc. **353** (2001), no. 6, 2513–2530.
[21] U. Kirchgraber and D. Stoffer, *On the definition of chaos*, Z. Angew. Math. Mech. **69** (1989), no. 7, 175–185.
[22] P. Kloeden and Z. Li, *Li-Yorke chaos in higher dimensions: a review*, J. Difference Equ. Appl. **12** (2006), no. 3-4, 247–269.
[23] P. E. Kloeden, *Chaotic difference equations in* $\mathbf{R}^n$, J. Austral. Math. Soc. Ser. A **31** (1981), no. 2, 217–225.
[24] S. F. Kolyada, *Li-Yorke sensitivity and other concepts of chaos*, Ukraïn. Mat. Zh. **56** (2004), no. 8, 1043–1061.
[25] T. Küpper, Y. Li, and B. Zhang, *Periodic solutions for dissipative-repulsive systems*, Tohoku Math. J. (2) **52** (2000), no. 3, 321–329.
[26] M. K. Kwong, *On Krasnoselskii's cone fixed point theorem*, Fixed Point Theory Appl. (2008), Art. ID 164537, 18.
[27] M. K. Kwong, *The topological nature of Krasnoselskii's cone fixed point theorem*, Nonlinear Anal. **69** (2008), no. 3, 891–897.
[28] A. Lasota and G. Pianigiani, *Invariant measures on topological spaces*, Boll. Un. Mat. Ital. B (5) **14** (1977), no. 2, 592–603.
[29] A. Lasota and J. A. Yorke, *On the existence of invariant measures for transformations with strictly turbulent trajectories*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. **25** (1977), no. 3, 233–238.
[30] T. Y. Li and J. A. Yorke, *Period three implies chaos*, Amer. Math. Monthly **82** (1975), no. 10, 985–992.
[31] F. R. Marotto, *Snap-back repellers imply chaos in* $\mathbf{R}^n$, J. Math. Anal. Appl. **63** (1978), no. 1, 199–223.
[32] F. R. Marotto, *On redefining a snap-back repeller*, Chaos Solitons Fractals **25** (2005), no. 1, 25–28.
[33] M. M. Marsh, *Fixed point theorems for partially outward mappings*, Topology Appl. **153** (2006), no. 18, 3546–3554.
[34] M. Martelli, M. Dang, and T. Seph, *Defining chaos*, Math. Mag. **71** (1998), no. 2, 112–122.
[35] J. Mawhin, *Henri Poincaré. A life in the service of science*, Notices Amer. Math. Soc. **52** (2005), no. 9, 1036–1044.
[36] J. Mawhin, *Variations on some finite-dimensional fixed-point theorems*, Ukrainian Math. J. **65** (2013), no. 2, 294–301.
[37] R. May, *Biological populations with nonoverlapping generations, stable points, stable cycles, and chaos*, Science **186** (1974), 645–647.
[38] R. May, *Simple mathematical models with very complicated dynamics*, Nature **261** (1976), 459–467.
[39] E.E. Moise, *Geometric topology in dimensions* 2 *and* 3, Springer-Verlag, New York-Heidelberg, 1977, Graduate Texts in Mathematics, Vol. 47.
[40] S. B. Nadler, Jr., *Continuum theory*, Monographs and Textbooks in Pure and Applied Mathematics, vol. 158, Marcel Dekker, Inc., New York, 1992, An introduction.

[41] M. Pireddu and F. Zanolin, *Fixed points for dissipative-repulsive systems and topological dynamics of mappings defined on N-dimensional cells*, Adv. Nonlinear Stud. **5** (2005), no. 3, 411–440.
[42] H. Poincaré, *Méthodes nouvelles de la mécanique céleste*, GauthierVillars, 1892–1899.
[43] H. Poincaré, *Science et méthode*, Flammarion, 1908.
[44] R. Rudnicki, *Ergodic measures on topological spaces*, Univ. Iagel. Acta Math. (1987), no. 26, 231–237.
[45] S. Silverman, *On maps with dense orbits and the definition of chaos*, Rocky Mountain J. Math. **22** (1992), no. 1, 353–375.
[46] S. Smale, *Diffeomorphisms with many periodic points*, Differential and Combinatorial Topology (A Symposium in Honor of Marston Morse), Princeton Univ. Press, Princeton, N.J., 1965, pp. 63–80.
[47] S. Smale, *Differentiable dynamical systems*, Bull. Amer. Math. Soc. **73** (1967), 747–817.
[48] S. Smale, *Finding a horseshoe on the beaches of Rio*, Math. Intelligencer **20** (1998), no. 1, 39–44.
[49] E. Sovrano, *Snap-back repellers*, Master's thesis, University of Udine, Italy, Academic Year 2012-2013.
[50] P. Touhey, *Yet another definition of chaos*, Amer. Math. Monthly **104** (1997), no. 5, 411–414.
[51] S. Wiggins, *Chaotic transport in dynamical systems*, Interdisciplinary Applied Mathematics, vol. 2, Springer-Verlag, New York, 1992.
[52] P. Zgliczyński and M. Gidea, *Covering relations for multidimensional dynamical systems*, J. Differential Equations **202** (2004), no. 1, 32–58.

Authors' addresses:

Elisa Sovrano
Department of Mathematics and Computer Science
University of Udine
Via delle Scienze 206
33100 Udine, Italy
E-mail: `sovrano.elisa@spes.uniud.it`

Fabio Zanolin
Department of Mathematics and Computer Science
University of Udine
Via delle Scienze 206
33100 Udine, Italy
E-mail: `fabio.zanolin@uniud.it`

Received June 17, 2014
Revised July 2, 2014

Rend. Istit. Mat. Univ. Trieste
Volume 46 (2014), 123–136

An analysis of the Stokes system with pressure dependent viscosity

Eduard Marušić-Paloka

Abstract. *In this paper we study the existence and uniqueness of the solution of the Stokes system, describing the flow of a viscous fluid, in case of pressure dependent viscosity.*

Keywords: incompressible viscous flow, Stokes equations, pressure dependent viscosity, existence of the solution
MS Classification 2010: 35Q30, 76A05, 76B03

1. Introduction

In his classical paper from 1848 [22] Stokes predicted that the viscosity of the fluid can depend on the pressure. Those effects for various liquids have been measured in many engineering papers, starting from the beginning of the 20-th century (see e.g. [1, 2, 10, 13, 15] etc.). That effect is usually neglected as it becomes important only in case of high pressure. Several models have been used to describe that relation since. The most popular is probably the exponential law

$$\mu = \mu_0 \exp(\alpha\, p) \tag{1}$$

usually called the Barus formula [1]. Here μ_0 and α are the constants depending on the fluid. The formula seems to be reasonable for mineral oil, unless the pressure is very high (larger then 0.5 MPa). The coefficient α typically ranges between 1 and $10^{-8}\ Pa^{-1}$. The lower end of the range corresponding to *paraffinic* and the upper end to the *nephtenic* oils (see Jones et al [14]). That formula is frequently used by engineers, sometimes combined with temperature dependence. In case of two above mentioned laws explicit solutions of the equations of motion, for some particular situations like unidirectional and plane-parallel flows, were found in [12]. Discussion on other possibilities for the viscosity-pressure formula and some historical remarks on the subject can be found in the same paper. Several engineering papers can be found discussing other possible laws and their consistency. We mention for instance [19] and [21].

From mathematical point of view supposing that the viscosity depends on the pressure makes the Navier-Stokes system much more complicated. Not only that it brings in additional nonlinearity to the momentum equation, but it changes the nature of the pressure as it cannot be eliminated from the system using Helmholtz decomposition and it cannot be seen as just a Lagrange multiplier. First important contribution was made by Renardy [20], where the viscosity function $p \mapsto \mu(p)$ is assumed to be sublinear at infinity and its derivative is assumed to be bounded on $\mathbf{R}$. In three interesting papers Gazzola and Gazzola and Secchi have proved the existence theorems for stationary and evolutional case under the assumption that the flow is governed by almost conservative force, has small initial velocity (in non-stationary case) and that μ is a smooth function globally bounded from below by some positive constant (hypothesis that rules out the Barus formula). The approach is based on the local inverse function theorem and relies on the fact that given data are small enough. Since our next goal is to do the asymptotic analysis of the system in thin domain, the small data assumption is not acceptable as such condition, in general, depends on the domain, which is not practical when the domain shrinks.

Interesting results were found about evolution Navier- Stokes system with pressure and shear dependent viscosity in two papers [11] and [16], but under certain technical assumptions on the viscosity that are not fulfilled by the Barus formula. More precisely they assume that $\mu = \mu(p, |D\,\mathbf{u}|^2)$, satisfies

$$\begin{aligned}
&C_1\left(1+|\mathbf{D}|^2\right)^{\frac{r-2}{2}}|\mathbf{B}|^2 \le \frac{\partial}{\partial \mathbf{D}}\mu\left(p,|\mathbf{D}|^2\right)(\mathbf{B}\otimes\mathbf{B}) \le C_2\left(1+|\mathbf{D}|^2\right)^{\frac{r-2}{2}}|\mathbf{B}|^2\\
&\frac{\partial}{\partial p}\mu\left(p,|\mathbf{D}|^2\right)|\mathbf{D}|^2 \le \gamma_0\left(1+|\mathbf{D}|^2\right)^{\frac{r-2}{4}} \le \gamma_0\\
&\forall\ \ \mathbf{D}\,,\mathbf{B}\in\mathbf{R}^{n\times n}\ \text{symmetric}\ ,\ p\in\mathbf{R}\ ,\ 1<r<2\ .
\end{aligned}$$

Those assumptions allow them to derive the uniform a priori estimates for the Galerkin approximation. Also, for simplicity, they take periodic boundary conditions which is not useful for applications that we have in mind.

Reynolds lubrication equation with pressure dependent viscosity was studied in [17]. It's well-posedness was proved as well as an apropriate maximum principle. Finally the nonlocal effects obtained by homogenization of the same equation in case of large Reynolds number were studied. An asymptotic model for flow of the fluid with pressure dependent viscosity through thin domain was derived in [18].

We use an approach here that does not need any small data assumption and very general assumption on μ satisfied by Barus formula and other empiric laws that can be found in the literature. Indeed, we assume that the function $p \mapsto \mu(p)$ is convex in vicinity of $-\infty$, that $\lim_{s\to-\infty}\mu'(s) = 0$ and $\int_{-\infty}^{0}\frac{ds}{\mu(s)} = +\infty$. Those assumption are irrelevant from the physical point of view since the dependence of the viscosity on the pressure becomes significant only for large positive pressure, while all our assumptions concern the behavior of the viscosity for large negative pressures. Typically, the viscosity is taken to be constant for negative pressures, and thus all our assumptions are trivially fulfilled. We also work with physically relevant Dirichlet boundary condition. However the approach works only for the stationary Stokes system. i.e. we do not handle neither the inertial term nor the time derivative.

2. Position of the problem

The mathematical model can be written in the following form: Let $\Omega \subset \mathbf{R}^d$, $d = 2,3$ be a bounded smooth domain. We assume that the boundary is at least of class C^2. The unknowns in the model are $\mathbf{u}$ - the velocity , p - the pressure . We recall that the stationary motion of the incompressible viscous laminar flow is governed by the Navier-Stokes equations. Thus we write the following system

$$\begin{cases} -2\mathrm{div}\,[\mu(p)\,\mathbf{D}\,\mathbf{u}\,] + (\mathbf{u}\,\cdot\,\nabla\,)\,\mathbf{u} + \nabla p = \mathbf{0}\,, \quad \mathrm{div}\,\mathbf{u} = 0 \ \text{ in } \ \Omega\,, \\ \mathbf{u} = \mathbf{g} \ \text{ on } \ \partial\Omega\,. \end{cases} \tag{2}$$

If the Reynolds number is not too large it is reasonable to neglect the inertial term and replace the Navier-Stokes by the Stokes system

$$\begin{cases} -2\mathrm{div}\,[\mu(p)\,\mathbf{D}\,\mathbf{u}\,] + \nabla p = \mathbf{0}\,, \quad \mathrm{div}\,\mathbf{u} = 0 \ \text{ in } \ \Omega\,, \\ \mathbf{u} = \mathbf{g} \ \text{ on } \ \partial\Omega\,. \end{cases} \tag{3}$$

We assume that the function $\mathbf{g}$ satisfies the following regularity and compatibility conditions

$$\mathbf{g} \in W^{2-1/\beta,\beta}(\Omega)^d\,, \qquad \text{for some } \ \beta > d \tag{4}$$

$$\int_{\partial\Omega} \mathbf{g}\,\cdot\,\mathbf{n} = 0\,. \tag{5}$$

For the dependence of the viscosity on the pressure we assume that $p \mapsto \mu(p)$ is a C^2 function and that it satisfies the following conditions:

$$\int_{-\infty}^{0} \frac{ds}{\mu(s)} = +\infty \tag{6}$$

$$\lim_{s\to-\infty} \mu'(s) = 0\,, \quad \text{and } \mu \text{ is convex in vicinity of } \ -\infty\,. \tag{7}$$

3. The idea

For the sake of simplicity, in this chapter, we assume that μ satisfies the Barus formula (1). Then

$$\begin{aligned} \nabla p \ &= 2\mathrm{div}\ (\mu(p)\ \mathbf{Du}) = \mu(p)\,\Delta\mathbf{u} + 2\mu'(p)\,\nabla p\,\mathbf{Du} \\ &= \mu_0\,e^{\alpha p}\ \ (\Delta\mathbf{u} + 2\alpha\,\nabla p\ \mathbf{Du})\ . \end{aligned}$$

Dividing by $\mu_0\, e^{\alpha p}$ we get

$$-\Delta \mathbf{u} + \frac{1}{\mu_0}\, e^{-\alpha p}\, \nabla p = 2\alpha\ \nabla p\ \mathbf{D u}\,.$$

We now look for the function q such that

$$\frac{1}{\mu_0}\, e^{-\alpha p}\, \nabla p = \nabla q\,.$$

Obviously there is a continuum of such functions given by

$$q = \frac{1}{\alpha \mu_0}\ \left(e^{-\alpha\sigma} - e^{-\alpha p}\right)\ , \tag{8}$$

where $\sigma \in \mathbf{R}$ is arbitrary . We can use that liberty in choice of σ to get what we want. Now

$$\nabla p = \mu_0\, e^{\alpha p}\ \nabla q = \frac{\mu_0}{e^{-\alpha\sigma} - \alpha\mu_0 q}\ \nabla q\,.$$

Thus our system becomes

$$-\Delta \mathbf{u} + \nabla q = 2\frac{\alpha\, \mu_0}{e^{-\alpha\sigma} - \alpha\mu_0 q}\ \nabla q\ \mathbf{D u}\,. \tag{9}$$

Obviously

$$\lim_{\sigma \to -\infty} \frac{\alpha\, \mu_0}{e^{-\alpha\sigma} - \alpha\mu_0 q} = 0$$

meaning that the right-hand side can be made as small as we need by picking large $|\sigma|$, $\sigma < 0$.

Equation (9) is to be complemented by

$$\operatorname{div} \mathbf{u} = 0\,, \qquad \mathbf{u} = \mathbf{g}\ \text{ on } \partial\Omega$$

and it makes a nonlinear Stokes-like system, but with nonlinearity that can be made as small as we want. Under the circumstances, it is not too complicated to prove that it has a solution. Does it mean that our original problem (3) has a solution? Well, it does if we can invert the transformation (8) and reconstruct p from q. It is easy to see that

$$p(x) = \frac{1}{\alpha}\ln\left(\frac{1}{e^{-\alpha\sigma} - \alpha\, \mu_0\, q(x)}\right)\,.$$

For that formula to make sense, we need to make sure that

$$q(x) < \frac{1}{\alpha\, \mu_0\, e^{\alpha\sigma}}\ ,\ \ \forall\ x \in\ \Omega\,.$$

Once again that condition can be met by choosing σ small enough, i.e. $\sigma < 0$ and $|\sigma|$ large enough.

The uniqueness of such solution can be proved since the transformed system is the Stokes system with small nonlinear perturbation, and if the transformed system has a unique solution, so does the original system. At least as long as we look only for regular solutions.

4. The main results

The statement of the main result is that our problem has a solution, which is unique, under some technical conditions. In standard Stokes (or Navier-Stokes) system the pressure is obviously determined only up to a constant, so it is not unique unless we impose some additional condition, like prescribing the value of its integral over Ω or prescribing the value of the pressure in some point of the domain Ω. That is less obvious here, since the pressure appears in the viscosity formula. However, it turns out that similar condition is needed here to fix the pressure.

4.1. Existence theorem

THEOREM 1. *Let* $\mathbf{g}$ *satisfy (4) and (5). Assume, in addition that (6) and (7) hold. Then the problem (3) has a solution* $(\mathbf{u}, p) \in \mathbf{X} = W^{2,\beta}(\Omega)^d \times W^{1,\beta}(\Omega)$.

4.2. Existence proof

4.2.1. Transformation of the system

We slightly generalize the idea presented in chapter 3, for general viscosity-pressure relation, and give all the technical details.

So, for some $\sigma \in \mathbf{R}$ we define two mappings

$$B(p,\sigma) = \int_\sigma^p \frac{ds}{\mu(s)} \tag{10}$$

As $p \mapsto \frac{1}{\mu(p)}$ is continuous and positive, B is of class C^1. Since

$$\frac{\partial B}{\partial p}(p,\sigma) = \frac{1}{\mu(p)} > 0$$

the function $B(\,\cdot\,,\sigma) : \mathbf{R} \mapsto \mathbf{R}$ is strictly increasing (and thus injective), for any parameter σ. Furthermore

$$Im\, B(\,\cdot\,,\sigma) = [M_\sigma^-, M_\sigma^+]\,,$$

where

$$M_\sigma^+ = \lim_{p\to+\infty} B(p,\sigma) = \int_\sigma^{+\infty} \frac{ds}{\mu(s)}\,, \quad M_\sigma^- = \lim_{p\to+\infty} B(p,\sigma) = -\int_{-\infty}^{\sigma} \frac{ds}{\mu(s)}\,. \tag{11}$$

If the above integrals are divergent we take their value to be $+\infty$. We can now define another function

$$H(\,\cdot\,,\sigma) = B^{-1}(\,\cdot\,,\sigma)\,. \tag{12}$$

Thus $q \mapsto H(q,\sigma)$ is an inverse of the function $p \mapsto B(q,\sigma)$, while σ is treated only as a parameter. Obviously $H(\,\cdot\,,\sigma) : [M_\sigma^-, M_\sigma^+] \to \mathbf{R}$ is well defined, strictly increasing and smooth. Furthermore

$$\frac{\partial H}{\partial q}(q,\sigma) = \mu(p) = \mu(H(q,\sigma))\,.$$

Next we define the function

$$b(q,\sigma) = 2\,\mu'(p) = 2\,\mu'(H(q,\sigma)) \tag{13}$$

that we need in the sequel. Function μ' is defined on $\mathbf{R}$, but $H(\,\cdot\,,\sigma)$ is defined only on $[M_\sigma^-, M_\sigma^+]$, thus the domain of $b(\,\cdot\,,\sigma)$ is $[M_\sigma^-, M_\sigma^+]$. As μ and H are smooth b is continuous. Using the assumptions on μ we prove the following important technical result:

LEMMA 1. *Let $\mu : \mathbf{R} \to \mathbf{R}$ satisfy the following hypothesis*

1.) $\displaystyle\int_{-\infty}^{0} \frac{ds}{\mu(s)} = +\infty$

2.) μ is convex in vicinity of $-\infty$ and $\displaystyle\lim_{s\to-\infty} \mu'(s) = 0$.

Then

$$M_\sigma^- = -\infty\,, \qquad \lim_{\sigma\to-\infty} M_\sigma^+ = +\infty \tag{14}$$

$$\lim_{\sigma\to-\infty} b(q,\sigma) = 0\,, \qquad \textit{for any } q \in \mathbf{R}\,. \tag{15}$$

Proof. First of all, 1.) means that $M_\sigma^- = -\infty$ and that

$$\lim_{\sigma\to-\infty} M_\sigma^+ = \int_{-\infty}^{+\infty} \frac{ds}{\mu(s)} = \infty\,.$$

Next, since $b(q,\sigma) = \mu'\left(H(q,\sigma)\right)$ we only need to see that assumption 1.) implies that

$$\lim_{\sigma\to-\infty} H(q,\sigma) = -\infty\,. \tag{16}$$

But that is clear, because the mapping $\sigma \mapsto H(q,\sigma)$ is strictly increasing and unbounded from below. Indeed

$$\frac{\partial}{\partial\sigma} H(q,\sigma) = \frac{\mu\left(H(q,\sigma)\right)}{\mu(\sigma)} > 0\,.$$

On the other hand the mapping $\sigma \mapsto H(q,\sigma)$ cannot be bounded from below. Supposing that there exists some $\gamma(q) = \inf_{\sigma\in\mathbf{R}} H(q,\sigma) \in \mathbf{R}$ implies that

$$q = \int_\sigma^{H(q,\sigma)} \frac{ds}{\mu(s)} \geq \int_\sigma^{\gamma(q)} \frac{ds}{\mu(s)}\,.$$

Now the assumption 1.) leads to

$$q = \lim_{\sigma\to-\infty} \inf \int_\sigma^{H(q,\sigma)} \frac{ds}{\mu(s)} \geq \lim_{\sigma\to-\infty} \int_\sigma^{\gamma(q)} \frac{ds}{\mu(s)} = \int_{-\infty}^{\gamma(q)} \frac{ds}{\mu(s)} = +\infty\,.$$

Thus $\sigma \mapsto H(q,\sigma)$ is strictly increasing and unbounded from below and consequently

$$\lim_{\sigma\to-\infty} H(q,\sigma) = -\infty\,.$$

Then, since μ' tends to zero at $-\infty$ we have

$$b(q,\sigma) = 2\,\mu'(H(q,\sigma)) \to 0\,, \quad \text{as } \sigma \to -\infty\,,$$

We now define the new unknown

$$q = B(p,\sigma) \tag{17}$$

We can now rewrite our system in terms of new unknown and it reads

$$-\Delta \mathbf{u} + \nabla q = b(q,\sigma)\,\mathbf{D}\,\mathbf{u}\;\nabla q\,, \quad \operatorname{div}\mathbf{u} = 0 \text{ in } \Omega\,, \tag{18}$$

$$\mathbf{u} = \mathbf{g} \text{ on } \partial\Omega\,. \tag{19}$$

Recall that, due to (15)

$$\lim_{\sigma\to-\infty} b(q,\sigma) = 0\,, \qquad \text{for any } q\,. \tag{20}$$

□

Example 1. *In case of Barus law $\mu(p) = \mu_0\, e^{\alpha\, p}$ we have*

$$B(p,\sigma) = \frac{1}{\alpha\,\mu_0}\left(e^{-\alpha\sigma} - e^{-\alpha p}\right)\,, \quad H(q,\sigma) = \frac{1}{\alpha}\, ln\, \frac{1}{e^{-\alpha\sigma} - \alpha\,\mu_0 q}\,.$$

Furthermore

$$b(q,\sigma) \equiv 2\,\mu'(H(q,\sigma)) = \frac{2\,\alpha\,\mu_0}{e^{-\alpha\sigma} - \alpha\,\mu_0 q}\,, \quad M_\sigma^+ = \frac{e^{-\alpha\,\sigma}}{\alpha\,\mu_0}\,, \quad M_\sigma^- = -\infty\,.$$

4.2.2. The proof of Theorem 1

The idea is to define the mapping $T : X = W^{2,\beta}(\Omega)^d \times W^{1,\beta}(\Omega) \to X$ by taking $T(\mathbf{v}, q) = (\mathbf{w}, \pi)$ where $(\mathbf{w}, \pi)$ is the solution to the problem

$$-\Delta \mathbf{w} + \nabla \pi = b(q, \sigma)\, \mathbf{D}\, \mathbf{v}\ \nabla q \quad \text{in } \Omega\,, \tag{21}$$

$$\operatorname{div} \mathbf{w} = 0 \ \text{ in } \Omega\,, \quad \mathbf{w} = \mathbf{g} \ \text{ on } \partial\Omega\,, \quad \int_\Omega \pi = 0\,. \tag{22}$$

The classical result by Cattabriga [3](see Appendix A) implies, for any $\beta > d$ and any $(\mathbf{v}, q) \in W^{2,\beta}(\Omega)^d \times W^{1,\beta}(\Omega)$ the existence of such $(\mathbf{w}, \pi) \in W^{2,\beta}(\Omega) \times W^{1,\beta}(\Omega)$. Since we have imposed

$$\int_\Omega \pi = 0\,, \tag{23}$$

the solution is unique. Furthermore, assuming that

$$|(\mathbf{v}, q)|_X \equiv |\mathbf{v}|_{W^{2,\beta}(\Omega)} + |q|_{W^{1,\beta}(\Omega)} \le M\,, \quad \int_\Omega q = 0\,. \tag{24}$$

and using (45), we obtain

$$|\mathbf{w}|_{W^{2,\beta}(\Omega)} + |\pi|_{W^{1,\beta}(\Omega)} \le C_\beta \Big(|b(q,\sigma)|_{L^\infty(\Omega)} |\mathbf{D}\, \mathbf{v}|_{L^\infty(\Omega)} |\nabla q|_{L^\beta(\Omega)} + \tag{25}$$
$$+ |\mathbf{g}|_{W^{2-1/\beta,\beta}(\partial\Omega)} \Big) \le C_\beta \left(M^2 C(\beta,\infty)^2 |b(q,\sigma)|_{L^\infty(\Omega)} + |\mathbf{g}|_{W^{2-1/\beta,\beta}(\partial\Omega)} \right) .$$

For any $x \in \overline{\Omega}$

$$|q(x)| \le C(\beta, \infty)\ M \equiv \overline{M}\,. \tag{26}$$

As μ is convex in vicinity of $-\infty$, there exists some $s_0 < 0$ such that $s \mapsto \mu(s)$ is convex for all $s < s_0$.

For any $\eta < 0$ there exists $\sigma_0 < 0$ such that for any $\sigma < \sigma_0$

$$H(q(x), \sigma) < H(\overline{M}, \sigma) < \eta\,.$$

We choose $\eta < s_0$. Due to the convexity of μ we know that μ' is increasing for $s < -s_0$ so that for all $\sigma < \sigma_0$

$$b(q(x), \sigma) = \mu'(H(q(x), \sigma)) < \mu'(H(\overline{M}, \sigma)) = b(\overline{M}, \sigma)\,. \tag{27}$$

Now, due to the Lemma 1

$$\lim_{\sigma \to -\infty} |b(\overline{M}, \sigma)| = 0\,. \tag{28}$$

Thus, for any $x \in \overline{\Omega}$ and for any $\varepsilon > 0$, there exists σ_0 such that for any $\sigma < \sigma_0$

$$|b(q(x), \sigma)|_{L^\infty(\Omega)} \leq b(\overline{M}, \sigma) \leq \frac{\varepsilon}{C_\beta\, M^2 C(\beta, \infty)^2}\,.$$

For

$$G = |\mathbf{g}|_{W^{2-1/\beta,\beta}(\partial\Omega)} \tag{29}$$

we choose

$$M = 2\, C_\beta\ G \tag{30}$$

and for $\varepsilon < G$ (24) and (25) imply that

$$|(\mathbf{w}, \pi)|_X \equiv |\mathbf{w}|_{W^{2,\beta}(\Omega)} + |\pi|_{W^{1,\beta}(\Omega)} < M\,. \tag{31}$$

It proves that T maps the ball $B_M \subset X$ of radius M in itself. To apply the Tychonoff fixed point theorem it remains to prove that T is weakly continuous. As the ball B_M is metrizable in weak topology, weak sequential continuity will be enough. To do so, we assume that the sequence $(\mathbf{v}_n, q_n)$ converges weakly in X to some $(\mathbf{v}, q)$. Due to the compact embedding $W^{1,\beta}(\Omega) \subset C(\overline{\Omega})$ we know that $\mathbf{D}\mathbf{v}_n$ is bounded and strongly convergent in $C(\overline{\Omega})$, while ∇q_n is bounded in $L^\beta(\Omega)$.

Let $(\mathbf{w}_n, \pi_n) = T(\mathbf{v}_n, q_n)$. Since T maps B_M in itself, the sequence (w_n, π_n) is bounded in X. Furthermore, we can extract a subsequence, denoted by the same symbol, such that

$$\mathbf{w}_n \rightharpoonup \mathbf{w} \quad \text{weakly in } \ W^{2,\beta}(\Omega)^d \tag{32}$$

$$q_n \rightharpoonup q \quad \text{weakly in } \ W^{1,\beta}(\Omega)\,. \tag{33}$$

Compact embedding $W^{1,\beta}(\Omega) \subset C(\overline{\Omega})$ and (31) implies the strong convergence

$$\mathbf{w}_n \to \mathbf{w} \quad \text{in } \ C^1(\overline{\Omega})^d \tag{34}$$

$$q_n \to q \quad \text{in } \ L^\sigma(\Omega) \ \text{ and in } \ C(\overline{\Omega})\,. \tag{35}$$

Due to the continuity of b, we consequently get

$$b(q_n, \sigma) \to b(q, \sigma) \quad \text{in } \ C(\overline{\Omega})\,. \tag{36}$$

That is enough to pass to the limit in (21), (22). Proving that $T(v_n, q_n) \rightharpoonup T(v, q)$ weakly in X means that $T : B_M \to B_M$ is weakly continuous. Now the Tychonoff fixed point theorem implies the existence of solution $(\mathbf{u}, q) \in B_M$ of the transformed system (18), (19). To prove that $(\mathbf{u}, p)$, $p = H(q, \sigma)$, with H defined by (12), is the solution to the original system (3) we need to verify that $q(x)$ is in the range of function $H(\,\cdot\,, \sigma)$, $Im\, H(\,\cdot\,, \sigma) = [M_\sigma^-, M_\sigma^+]$, for some $\sigma < 0$. We recall that $M_\sigma^\pm$ were defined by (11) . Under the assumption (7) Lemma 1 states that $M_\sigma^- = -\infty$ so that we only need to verify that $q(x) \leq M_\sigma^+$, $\forall x \in \Omega$, which is fulfilled (again Lemma 1) for sufficiently large negative σ. Indeed, we know that $q(x) \leq \overline{M}$. It is therefore sufficient to chose σ such that $M_\sigma^+ \geq \overline{M}$.

4.3. Uniqueness theorem

The uniqueness of the solution can, in general, be proved only if the given boundary data is not too large and the pressure is not too high. The construction that we have used to prove the existence of the solution leads to the unique solution of that form. We recall that for such solution the velocity $\mathbf{u}$ and the transformed pressure $q = B(p,\sigma)$ remain inside the ball B_M in X for an appropriate choice of M. However, for general boundary data, we cannot rule out the possibility that there are some other solutions for which $(\mathbf{u}, q) \notin B_M$. For our solution (27) and (28) hold, so that the right hand side in the transformed equation (19) can be made as small as we want. Thus supposing that the problem (19) has two solutions $(\mathbf{u}, q), (\mathbf{w}, \eta)$ in $\mathbf{X}$. We denote by

$$\mathbf{E} = \mathbf{u} - \mathbf{w}\,, \qquad e = q - \eta\,.$$

Then, obviously

$$\mathbf{E} = 0 \;\text{ on }\; \partial\Omega\,. \tag{37}$$

On the other hand, due to (13), (15), for given $\varepsilon > 0$ we can choose σ in definition of b such that

$$|b(q,\sigma)\,\nabla q\, D\,\mathbf{u}|_{L^\beta(\Omega)} \le \frac{\varepsilon}{2\,C_\beta}\,, \quad |b(\eta,\sigma)\,\nabla \eta\, D\,\mathbf{w}|_{L^\beta(\Omega)} \le \frac{\varepsilon}{2\,C_\beta}\,, \tag{38}$$

where $C_\beta > 0$ is the constant from (45) Since the difference $(\mathbf{E}, e)$ satisfies the Stokes system

$$-\Delta \mathbf{E} + \nabla e = b(q,\sigma)\,\nabla q\, D\,\mathbf{u} - b(\eta,\sigma)\,\nabla \eta\, D\,\mathbf{w}$$

and the boundary condition (37), the standard a priori estimate (45) implies

$$|\mathbf{E}|_{W^{2,\beta}(\Omega)} + |e|_{W^{1,\beta}(\Omega)/\mathbf{R}} \le \varepsilon\,.$$

As ε was arbitrary, we conclude that $\mathbf{E} = 0$ and $e = const$. That proves the uniqueness of the velocity since it is independent on choice of σ. On the other hand q and η do depend on σ. However their difference does not since

$$e(x) = q(x) - \eta(x) = \int_{\pi(x)}^{p(x)} \frac{ds}{\mu(s)}\,,$$

with $p = H(q,\sigma)$, $\pi = H(\eta,\sigma)$. So, if we prescribe the mean value of the transformed pressure , i.e. if we put

$$\int_\Omega q = \int_\Omega \eta = 0\,, \tag{39}$$

we have $q = \eta$ and then

$$\int_{\pi(x)}^{p(x)} \frac{ds}{\mu(s)} = 0 \quad \text{for any} \quad x \in \Omega \,.$$

Functions $1/\mu, p, \pi$ are continuous and $1/\mu$ is positive, so that $p(x) = \pi(x)$. Thus, there can be only one solution constructed by transforming the pressure $q = B(p, \sigma)$ and solving (19).

As for the general case, we were only able to prove weaker result.
First of all, we need to assume that μ is increasing, which is physically reasonable. We also require that μ is convex, not only in vicinity of $-\infty$ but also in vicinity of $+\infty$. More precisely, we assume that there exists some $\xi_0 \geq 0$ such that $\xi \mapsto \mu(\xi)$ is convex for $\xi > \xi_0$. Next, suppose that μ is smooth and define

$$\overline{\mu'} = \sup_{\xi \leq \xi_0} |\mu'(\xi)| \,.$$

It is well known (see e.g. Galdi [5]) that, for any $\psi \in L^2(\Omega)$ such that

$$\int_\Omega \psi = 0$$

there exists $\mathbf{z} \in H^1(\Omega)$ such that

$$\begin{cases} \operatorname{div} \mathbf{z} = \psi \\ \mathbf{z} = 0 \;\text{ on }\; \partial\Omega \\ |\mathbf{z}|_{H^1(\Omega)} \leq C_{div} |\psi|_{L^2(\Omega)} \,, \end{cases} \tag{40}$$

with C_{div} depending only on the domain Ω.

We can now formulate the uniqueness result:

THEOREM 2. *Let $\overline{M}$ be defined by (26), (30) and (29). Assume that*

$$\overline{\mu'}\;\overline{M}\, C_{div} \left(1 + \sqrt{\mu(\xi_0)}\right) < 1 \,. \tag{41}$$

Then the problem (3) has only one solution $(\mathbf{u}, p) \in \mathbf{X} = W^{2,\beta}(\Omega)^d \times W^{1,\beta}(\Omega)$ such that

$$p(x) \leq \xi_0 \,, \quad x \in \Omega \,.$$

Proof. First of all, it is sufficient to prove the uniqueness of the pressure, since it implies the uniqueness of the velocity. Indeed, for given pressure the viscosity is given and the system is linear with respect to velocity.

If $p(x) \leq \xi_0$ then

$$\mu(p(x)) \leq \mu(\xi_0) \,, \quad x \in \Omega \,.$$

Let $(w, \pi) \in X$ be the solution constructed in Theorem 1 and let (u, p) be any other solution. Then, subtracting the equations (3) for $(\mathbf{u}, p)$, $(\mathbf{w}, \pi)$ and testing with $\mathbf{u} - \mathbf{w}$ gives

$$\int_\Omega \mu(p)\, |\, \mathbf{D}\,(\mathbf{u}-\mathbf{w})|^2 = \int_\Omega [\mu(p)-\mu(\pi)]\ \mathbf{D}\,(\mathbf{u}-\mathbf{w})\,\mathbf{D}\,\mathbf{w} \le$$
$$\le \overline{\mu'}\, |\mathbf{D}\,\mathbf{w}|_{L^\infty(\Omega)}\, |\mathbf{D}(\mathbf{u}-\mathbf{w})|_{L^2(\Omega)}\,.$$

Since

$$|\mathbf{D}\,\mathbf{w}|_{L^\infty(\Omega)} \le C(\infty,\beta)|\mathbf{w}|_{W^{2,\beta}(\Omega)} \le \overline{M}$$

we get

$$|\sqrt{\mu(p)}\ \mathbf{D}\,(\mathbf{u}-\mathbf{w})|_{L^2(\Omega)} \le \overline{\mu'}\ \ \overline{M}\,. \tag{42}$$

Now, taking $\mathbf{z}$ such that

$$\operatorname{div}\mathbf{z} = p-\pi \ \text{ in } \ \Omega\,, \quad \mathbf{z}=0 \text{ on } \partial\Omega$$
$$|\mathbf{z}|_{H^1(\Omega)} \le C_{div}\,|p-\pi|_{L^2(\Omega)}\,,$$

and testing (3) with it, we obtain, using (42)

$$\int_\Omega (p-\pi)^2 = \int_\Omega (p-\pi)\operatorname{div}\mathbf{z} = \int_\Omega [\mu(p)-\mu(\pi)]\ \mathbf{D}\,\mathbf{w}\ \mathbf{D}\,\mathbf{z} +$$
$$+\int_\Omega \mu(p)\,\mathbf{D}\,(\mathbf{u}-\mathbf{w})\,\mathbf{D}\,\mathbf{z} \le \overline{\mu'}\ \ \overline{M}\, C_{div}\,\left(1+\sqrt{\mu(\xi_0)}\right)\,|p-\pi|^2_{L^2(\Omega)}\,.$$

The result now follows from the assumption (41). □

A. Technical results

It is well known (see e.g. Cattabriga [3]) that the Stokes system

$$-\Delta\mathbf{v} + \nabla\pi = \mathbf{f}\ \ ,\ \ \operatorname{div}\mathbf{v}=0 \ \ \text{ in } \ \Omega \tag{43}$$
$$\mathbf{v} = \mathbf{h} \ \text{ on } \ \partial\Omega\,. \tag{44}$$

with $\mathbf{f} \in L^\beta(\Omega)^d$ and $\mathbf{h} \in W^{2-1/\beta,\beta}(\partial\Omega)^d$, satisfying

$$\int_{\partial\Omega} \mathbf{h}\,\cdot\,\mathbf{n} = 0\,,$$

admits a solution $(\mathbf{v}, \pi) \in \mathbf{X} = W^{2,\beta}(\Omega)^d \times W^{1,\beta}(\Omega)$ which is unique (π up to an additive constant) . Furthermore it satisfies the following estimate

$$|\mathbf{v}|_{W^{2,\beta}(\Omega)} + |\pi|_{W^{1,\beta}(\partial\Omega)/\mathbf{R}} \le C_\beta\left(|\mathbf{f}|_{L^\beta(\Omega)} + |\mathbf{h}|_{W^{2-1/\beta,\beta}(\partial\Omega)}\right)\,. \tag{45}$$

We recall that for $\beta > n$ the embedding $W^{1,\beta}(\Omega) \subset L^\infty(\Omega)$ holds true. We denote by $C(\beta,\infty)$ the constant such that

$$|\phi|_{L^\infty(\Omega)} \le C(\beta,\infty)\,|\phi|_{W^{1,\beta}(\Omega)} \ \ ,\ \ \forall\ \phi\in W^{1,\beta}(\Omega)\,. \tag{46}$$

Acknowledgement

This work was supported by the grant 3955 *Mathematical modeling and numerical simulations of processes in thin or porous domains* from Croatian science foundation.

References

[1] C. Barus, *Isotherms, isopiestics and isometrics relative to viscosity*, Amer. J. Sc. **45** (1893), 87–96.

[2] P.W. Bridgman, *The viscosity of liquids under pressure*, Proc. Nat. Acad. Sci. **11** (1925), 603–606

[3] L. Cattabriga, *Su un problema al contorno relativo al sistema di equazioni di Stokes*, Rend. Semin. Mat. Univ. Padova **31** (1961), 308–340.

[4] G. Delmaz and M. Godet, *Traction, load and film thickness in lightly-loaded lubricated point contacts* , J. Mech. Eng. Sci. **15** (1973) 400–409.

[5] G.P. Galdi, An introduction to the mathematical theory of the Navier-Stokes equations, Steady-state problems, Springer, 1994.

[6] F. Gazzola, *On stationary Navier-Stokes equations with a pressure dependent viscosity*, Rend. Inst. Lomb. Sci. Lett. Sez. A **128** (1994), 107–119.

[7] F. Gazzola, *A note on the evolution Navier-Stokes equations with a pressure dependent viscosity*, Z. Angew. Math. Phys **48** (1997), 760–773.

[8] F.Gazzola and P. Secchi, *Some results about stationary Navier-Stokes equations with a pressure-dependent viscosity*, Proceedings of the International conference on Navier-Stokes equations, Varenna, 1997, Pitman research notes, Math. Ser. 388, 1998, 174–183.

[9] D. Gilbarg and N.S. Trudinger, Elliptic partial differential equations of second order, Springer, 2001.

[10] E.M. Griest, W. Webb and R.W. Schiessler, *Effect of pressure on viscosity of higher hydrocarbons and their mixtures*, J. Chem. Phys. **29(4)** (1958), 711–720.

[11] J. Hron, J. Málek, J. Nečas and K.R. Rajagopal, *Numerical simulations and global existence of solutions for two-dimensional flows of fluids with shear and pressure dependent viscosities*, Math. Comput. Simulation **61** (2003), 297–315.

[12] J. Hron, J. Málek and K.R. Rajagopal, *Simple flows of fluids with pressure dependent viscosities*, Proc. R. Soc. Lond. A **457** (2001), 1603–1622.

[13] S.C. Jain, R. Sinhasan amd D.V. Singh, *Consideration of viscosity variation in determining the performance characteristics of circular bearings in the laminar and turbulent regimes*, Wear **86** (1983), 233–245.

[14] W.R. Jones, *Pressure viscosity measurement for several lubricants*, ASLE Trans. **18** (1975), 249–262.

[15] A. Kalogirou, S. Polyiadji and G.C. Georgiou, *Incompressible Poiseuille flow of Newtonian liquids with pressure-dependent viscosity*, J. Non-Newtonian Fluid Mech. **166** (2011), 413–419.

[16] J. Málek , J. Nečas and K.R. Rajagopal, *Global existence of solutions for flows fluids with pressure and shear dependent viscosities*, Appl. Math. Lett. **15** (2002), 961–967.

[17] E. Marušić-Paloka and S. Marušić, *Analysis of the Reynolds Equation for Lubrication in Case of Pressure-Dependent Viscosity*, Math. Probl. Eng. bf 2013 (2013) , 1–10.

[18] E. Marušić-Paloka and I. Pažanin, *A note on the pipe flow with a pressure-dependent viscosity*, J. Non-Newtonian Fluid Mech. **197** (2013) , 5–10.

[19] K.R. Rajagopal and A.Z. Szeri, *On an inconsistency in the derivation of the equations of hydrodynamic lubrication*, Proc. R. Soc. Lond. A **459** (2003), 2771–2786.

[20] M. Renardy, *Some remarks on the Navier-Stokes equations with pressure-dependent viscosity*, Comm. Part. Diff. Eq. **11(7)** (1986), 779–793.

[21] C.J.A. Roelands, *Correlation aspects of the viscosity-pressure relationship of lubricating oils*, PhD thesis, Delft University of technology, Netherlands, 1966.

[22] G.G. Stokes, *Notes on hydrodynamics. On the dynamical equations.* Cambridge and Dublin Mathematical Journal, III, (1848) 121–127.

Author's address:

Eduard Marušić-Paloka
Faculty of Science, Department of mathematics,
University of Zagreb,
Bijenička 30, 10000 Zagreb, Croatia
E-mail: `emarusic@math.hr`

Received December 1, 2013
Revised July 20, 2014

Rend. Istit. Mat. Univ. Trieste
Volume 46 (2014), 137–155

Similarity solutions for thawing processes with a convective boundary condition

Andrea N. Ceretani and Domingo A. Tarzia

Abstract. *An explicit solution of similarity type for thawing in a saturated semi-infinite porous media when change of phase induces a density jump and a convective boundary condition is imposed at the fixed face, is obtained if and only if an inequality for data is verified. Relationship between this problem and the problem with temperature condition studied in [8] is analized and conditions for physical parameters under which the two problems become equivalents are obtained. Furthermore, an inequality to be satisfied for the coefficient which characterizes the free boundary of each problem is also obtained.*

Keywords: Stefan problem, free boundary problem, phase-change process, similarity solution, density jump, thawing process, convective boundary condition, Neumann solution.
MS Classification 2010: 35R35 - 35C06 - 80A22

1. Introduction

In this paper, we consider the problem of thawing of a semi-infinite partially frozen porous media saturated with an incompressible liquid when change of phase induces a density jump and a convective boundary condition is imposed on the fixed face, with the aim of constructing similarity solutions (for a detailed exposition of the physical background we refer to [5, 7, 12, 14, 18]). In [8] and [10] (which generalized [19]) similarity solutions are obtained when a temperature and a heat flux condition are imposed at the fixed boundary, respectively. In this paper, we deal with the same physical situations as in [8, 10] and we study a one-dimensional model of the problem where the unknowns are the temperature $u(x,t)$ of the unfrozen zone $Q_1 = \{(x,t) : 0 < x < s(t), t > 0\}$, the temperature $v(x,t)$ of the frozen zone $Q_2 = \{(x,t) : x > s(t), t > 0\}$ and the free boundary $x = s(t)$, defined for $t > 0$, separating Q_1 and Q_2, which satisfies the following equations and boundary and initial conditions (we refer to [7] for a detailed explanation of the model):

$$u_t = d_U u_{xx} - b\rho\dot{s}(t)u_x \qquad 0 < x < s(t),\ t > 0 \quad (1)$$
$$v_t = d_F v_{xx} \qquad x > s(t),\ t > 0 \quad (2)$$
$$u(s(t),t) = v(s(t),t) = d\rho s(t)\dot{s}(t) \qquad t > 0 \quad (3)$$
$$k_F v_x(s(t),t) - k_U u_x(s(t),t) = \alpha\dot{s}(t) + \beta\rho s(t)(\dot{s}(t))^2 \qquad t > 0 \quad (4)$$
$$v(x,0) = v(+\infty,t) = -A \qquad x > 0,\ t > 0 \quad (5)$$
$$s(0) = 0 \quad (6)$$
$$k_U u_x(0,t) = \frac{h_0}{\sqrt{t}}(u(0,t) - B) \qquad t > 0 \quad (7)$$

with:

$$d_U = \alpha_U^2 = \frac{k_U}{\rho_U c_U}, \qquad d_F = \alpha_F^2 = \frac{k_F}{\rho_F c_F}, \qquad b = \frac{\epsilon\rho_W c_W}{\rho_U c_U}, \qquad d = \frac{\epsilon\gamma\mu}{K},$$

$$\rho = \frac{\rho_W - \rho_I}{\rho_W}, \qquad \alpha = \epsilon\rho_I l, \qquad \beta = \frac{\epsilon^2\rho_I(c_W - c_I)\gamma\mu}{K} = \epsilon d\rho_I(c_W - c_I) \neq 0.$$

where:
ϵ: porosity,
ρ_W and ρ_I: density of water and ice (g/cm^3),
c: specific heat at constant density $(cal/g\,^\circ C)$,
k_F and k_U: conductivity of the frozen and unfrozen zones $(cal/s\,cm\,^\circ C)$,
u: temperature of the unfrozen zone $(^\circ C)$,
v: temperature of the frozen zone $(^\circ C)$,
$u = v = 0$: melting point at atmospheric pressure,
l: latent heat at $u = 0$ (cal/g),
γ: coefficient in the Clausius-Clapeyron law $(s^2 cm^\circ C/g)$,
$\mu > 0$: viscosity of the liquid $(g/s\,cm)$,
$K > 0$: hydraulic permeability (cm^2),
$B > 0$: external boundary temperature at the fixed face $x = 0$ $(^\circ C)$,
$B_0 > 0$: temperature at the fixed face $x = 0$ $(^\circ C)$,
$-A < 0$: initial temperature $(^\circ C)$,
$h_0 > 0$: coefficient which characterizes the heat transfer at the fixed face $x = 0$ $(Cal/s^{\frac{1}{2}}\,cm^2\,^\circ C)$.

Remark 1.1. *The free boundary problem (1)-(7) reduces to the usual Stefan problem when*

$$\rho = 0$$

since in that case we have the cassical Stefan conditions on $x = s(t)$, i.e.:

$$u(s(t),t) = v(s(t),t) = 0\,, \quad t > 0$$

$$k_F v_x(s(t),t) - k_U u_x(s(t),t) = \alpha \dot{s}(t)$$

and therefore from now on we assume that $\rho \neq 0$.

The goal of this paper is to find the necessary and/or sufficient conditions for data (with three dimensionless parameters) in order to obtain an instantaneous phase-change process (1)-(7) with the corresponding explicit solution of the similarity type when a convective boundary condition of type (7) is imposed on the fixed face $x = 0$ [1, 4, 16, 23, 25]. We remark that the solution given in [25] is not correct for any data (in particular, for small heat transfer coefficient) for the classical two-phase Stefan problem ($\rho = 0$) which was improved in [20] obtaining the necessary and sufficient condition to get the corresponding explicit solution. Furthermore, we study the relationship between the problem (1)-(7) and the problem studied in [8], which consists of equations (1)-(6) and the following temperature boundary condition at the fixed face $x = 0$:

$$u(0,t) = B_0 \qquad\qquad t > 0, \tag{8}$$

where B_0 is the boundary temperature at the fixed face $x = 0$, with the aim of finding conditions under which the two problems become equivalent.

Recently Stefan-like problems were studied in [2, 3, 6, 9, 11, 15, 17, 21, 22, 24]. The plan is the following: first (Sect. 2) we obtain the necessary and sufficient condition in order to have a similarity solution of the free boundary problem (1)-(7) as a function of a positive parameter which must be the solution of a transcendental equation with three dimensionless parameters defined by the thermal coefficients, and initial and boundary conditions. We also find a monotonicity property between the coefficients which characterize the free boundary and the heat transfer at the fixed face $x = 0$. Then (Sect. 3) we give the necessary and/or sufficient conditions for the three real parameters involved in the trascendental equation in order to obtain an instantaneous phase-change process (1)-(7) with the corresponding similarity solution. We generalize results obtained for particular cases given in [13, 20]. Finally (Sect. 4), we analize the relationship between the problems (1)-(7) and (1)-(6) and (8), and we obtain conditions for data under which the two problems become equivalent. Furthermore, we obtain an inequality which satisfies the coefficient involved in the definition of the free boundary of the problem (1)-(6) and (8), which become the inequality obtained in [19] for the classical Stefan problem.

2. Similarity solutions

We have:

THEOREM 2.1. *The free boundary value problem (1)-(7) has the similarity solution:*

$$u(x,t) = \frac{Bg(p,\xi) + \frac{AMk_U}{2h_0\alpha_U}\xi^2 + (AM\xi^2 - B)\int_0^{\frac{x}{2\alpha_U\sqrt{t}}} exp(-r^2 + pr\xi)\,dr}{g(p,\xi) + \frac{k_U}{2h_0\alpha_U}} \tag{9}$$

$$v(x,t) = \frac{AM\xi^2 + A\ erf(\gamma_0\xi) - A(1+M\xi^2)\ erf\left(\frac{x}{2\alpha_F\sqrt{t}}\right)}{erfc(\gamma_0\xi)} \tag{10}$$

$$s(t) = 2\xi\alpha_U\sqrt{t} \tag{11}$$

if and only if the dimensionless coefficient $\xi > 0$ satisfies the following equation:

$$G(M,p,y) = y + Ny^3, \qquad y > 0 \tag{12}$$

involving three dimensionless parameters, N, M and p, defined by:

$$N = \frac{2\beta\rho\alpha_U^2}{\alpha} \in \mathbb{R}, \qquad M = \frac{2d\rho\alpha_U^2}{A} \in \mathbb{R}, \qquad p = 2b\rho \in \mathbb{R}, \tag{13}$$

where:

$$G(M,p,y) = \delta_1\left(1 - \frac{AM}{B}y^2\right)G_1(p,y) - \delta_2(1+My^2)G_2(y) \qquad y>0, \tag{14}$$

$$G_1(p,y) = \frac{exp((p-1)y^2)}{K_0 + g(p,y)} \qquad p\in\mathbb{R},\ y>0, \tag{15}$$

$$G_2(y) = \frac{exp(-\gamma_0^2y^2)}{erfc(\gamma_0 y)} \qquad M\in\mathbb{R},\ y>0, \tag{16}$$

$$g(p,y) = \int_0^y exp(-r^2 + pry)\,dr \qquad p\in\mathbb{R},\ y>0, \tag{17}$$

$$\delta_1 = \frac{k_U B}{2\alpha\alpha_U^2} > 0,\ \delta_2 = \frac{k_F A}{\alpha\alpha_U\alpha_F\sqrt{\pi}} > 0,\ K_0 = \frac{k_U}{2\alpha_U h_0} > 0,\ \gamma_0 = \frac{\alpha_U}{\alpha_F} > 0. \tag{18}$$

Proof. We will follow the method introduced in [8, 10]. First of all, we note that the function u defined by:

$$u(x,t) = \Phi(\eta) \qquad \text{with } \eta = \frac{x}{2\alpha_U\sqrt{t}}$$

is a solution of equation (1) if and only if Φ satisfies the following equation:

$$\frac{1}{2}\Phi''(\eta) + \left(\eta - \frac{b\rho}{\alpha_U}\dot{s}(t)\sqrt{t}\right)\Phi'(\eta) = 0$$

Then, if we consider the function $x = s(t)$ defined as in (11), for some $\xi > 0$ to be determined, we obtain that:

$$\Phi(\eta) = C_1 + C_2 \int_0^\eta \exp(-r^2 + 2b\rho\xi r)\, dr \tag{19}$$

where ξ, C_1 and C_2 are constant values.

Therefore, a solution of equation (1) is given by:

$$u(x,t) = C_1 + C_2 \int_0^{x/2\alpha_U\sqrt{t}} \exp(-r^2 + 2b\rho\xi r)\, dr \tag{20}$$

where C_1, C_2 and ξ are constant values to be determined.

On the other hand, it is well known that:

$$v(x,t) = C_3 + C_4 \operatorname{erf}\left(\frac{x}{2\alpha_F\sqrt{t}}\right), \tag{21}$$

is a solution of equation (2), where C_3 and C_4 are constant values to be determined and erf is the error function defined by:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-r^2)\, dr$$

From conditions (3), (5) and (7) we have that:

$$C_1 = \frac{Bg(2b\rho,\xi) + \frac{d\rho\alpha_U k_U}{h_0}\xi^2}{g(2b\rho,\xi) + \frac{k_U}{2h_0\alpha_U}}, \qquad C_2 = \frac{2d\rho\alpha_U^2\xi^2 - B}{g(2b\rho,\xi) + \frac{k_U}{2h_0\alpha_U}},$$

$$C_3 = \frac{2d\rho\alpha_U^2\xi^2 + A\operatorname{erf}(\gamma_0\xi)}{\operatorname{erfc}(\gamma_0\xi)}, \qquad C_4 = -\frac{2d\rho\alpha_U^2\xi^2 + A}{\operatorname{erfc}(\gamma_0\xi)}$$

and by imposing condition (4), we obtain that the functions s, u and v defined above by (9)-(11), are a solution of the free boundary value problem (1)-(7) if and only if the dimensionless parameter ξ satisfies the equation (12), and therefore the thesis holds. □

REMARK 2.2. *The similarity solution (19) is a generalized Neumann solution for the classical Stefan problem (see [1]) when a convective boundary condition at the fixed face $x = 0$ is imposed (see [20]).*

The following result give us a relationship between the coefficient ξ which characterizes the free boundary of the problem (1)-(7) and the coefficient $h_0 > 0$ which characterizes the heat transfer at the fixed face $x = 0$.

Corollary 2.3. *If the free boundary value problem (1)-(7) has a unique similarity solution of the type (9)-(11), $M > 0$, and the dimensionless coefficient $\xi > 0$ satisfies:*

$$0 < \xi < \sqrt{\frac{B}{AM}}, \tag{22}$$

then ξ is a strictly increasing function of $h_0 > 0$.

Proof. Let us assume that the free boundary value problem (1)-(7) has the similarity solution (9)-(11). Due to previous theorem we know that the dimensionless coefficient ξ satisfies equation (12). Since the LHS of equation (12) is a strictly increasing function of $h_0 > 0$, for all $y \in \left(0, \sqrt{\frac{B}{AM}}\right)$, as well as its limit when y tends to 0^+:

$$\lim_{y \to 0^+} \left(\delta_1 \left(1 - \frac{AM}{B} y^2\right) G_1(p, y) - \delta_2 (1 + My^2) G_2(y) \right) = \frac{\delta_1}{K_0} - \delta_2,$$

and the RHS of equation (12) is a strictly increasing function of $y > 0$, we can conclude that ξ is a strictly increasing function of $h_0 > 0$. □

3. Existence and uniqueness of similarity solutions

Due to previous theorem, in order to analize the existence and uniqueness of similarity solutions of the free boundary value problem (1)-(7), we can focus on the solvability of equation (12). With this aim, we split our analysis into four cases which correspond to four possible combinations of the signs of the dimensionless parameters M and N.

To prove the following results we will use properties of functions g, G_1 and G_2 involved in equation (12) which we proved in the Appendix.

Theorem 3.1. *If $M > 0$ and $N > 0$ then:*

1. *If:*
$$h_0 > \frac{A}{B} \frac{k_F}{\sqrt{\pi d_F}} \tag{23}$$
then equation (12) has at least one solution ξ, which satisfies $0 < \xi < \sqrt{\frac{B}{AM}}$.

2. *If $p \leq 1$ then equation (12) has a solution ξ, which satisfies $0 < \xi < \sqrt{\frac{B}{AM}}$, if and only if (23) holds. Moreover, when (23) holds, equation (12) has a unique solution ξ, which satisfies $0 < \xi < \sqrt{\frac{B}{AM}}$.*

Proof. Let $M > 0$ and $N > 0$.

1. We have proved in the Appendix that:

$$\lim_{y\to+\infty}\left(1-\frac{AM}{B}y^2\right)G_1(p,y)=\begin{cases}0 & p<2\\ -\infty & p\geq 2\end{cases}$$

and

$$\lim_{y\to+\infty}(1+My^2)G_2(y)=+\infty.$$

Then:

$$\lim_{y\to+\infty}\left[\delta_1\left(1-\frac{AM}{B}y^2\right)G_1(p,y)-\delta_2(1+My^2)G_2(y)\right]=-\infty.$$

Furthermore:

$$\lim_{y\to0^+}\left[\delta_1\left(1-\frac{AM}{B}y^2\right)G_1(p,y)-\delta_2(1+My^2)G_2(y)\right]=\frac{\delta_1}{K_0}-\delta_2.$$

Let us assume that $h_0>\frac{A}{B}\frac{k_F}{\sqrt{\pi d_F}}$. Since this inequality is equivalent to $\frac{\delta_1}{K_0}-\delta_2>0$, we have that the last limit is positive. Now taking into account that the RHS of (12) is an increasing function from 0 to $+\infty$, it follows that equation (12) has at least one positive solution.

Moreover, since:

$$\lim_{y\to0^+}\delta_1\left(1-\frac{AM}{B}y^2\right)G_1(p,y)=\frac{\delta_1}{K_0}\,,\quad \lim_{y\to0^+}\delta_2(1+My^2)G_2(y)=\delta_2\,,$$
$$\lim_{y\to+\infty}\delta_1\left(1-\frac{AM}{B}y^2\right)G_1(p,y)=-\infty\,,\ \lim_{y\to+\infty}\delta_2(1+My^2)G_2(y)=+\infty$$

and

$$\frac{\delta_1}{K_0}>\delta_2,$$

we have that the LHS of (12) has positive zeros $q_1<q_2<\cdots$. Thus we can find a solution ξ of (12) such that $\xi<q_1$. To prove that $0<\xi<\sqrt{\frac{B}{AM}}$, only remains to note that each q_k satisfies $q_k<\sqrt{\frac{B}{AM}}$.

2. Assume that $p\leq 1$ and $h_0>\frac{A}{B}\frac{k_F}{\sqrt{\pi d_F}}$. We know that (see [8]):

$$\frac{\partial}{\partial y}G_1(p,y)<0,\qquad y>0.$$

Then:

$$\frac{\partial}{\partial y}\left(1-\frac{AM}{B}y^2\right)G_1(p,y)=$$
$$\frac{-2AMy}{B}G_1(p,y)+\left(1-\frac{AM}{B}y^2\right)\frac{\partial}{\partial y}G_1(p,y)<0, \qquad 0<y<\sqrt{\frac{B}{AM}},$$

which implies that $\left(1-\frac{AM}{B}y^2\right)G_1(p,y)$ is strictly decreasing in the interval $\left(0,\sqrt{\frac{B}{AM}}\right)$. Furthermore, $\left(1-\frac{AM}{B}y^2\right)G_1(p,y)$ is negative in the interval $\left(\sqrt{\frac{B}{AM}},+\infty\right)$. On the other hand, we know from the Appendix that $(1+My^2)G_2(y)$ is a positive strictly incresing function of $y>0$. Therefore, the LHS of (12) is strictly decreasing in $\left(0,\sqrt{\frac{B}{AM}}\right)$ and negative in $\left(\sqrt{\frac{B}{AM}},+\infty\right)$. We also have that the limit of the LHS of (12) when y tends to 0^+ is positive because we are assuming that $h_0>\frac{A}{B}\frac{k_F}{\sqrt{\pi\alpha_F}}$. Then, since RHS of (12) is strictly increasing from 0 to $+\infty$, we have that equation (12) has a unique solution $\xi\in\left(0,\sqrt{\frac{B}{AM}}\right)$.

Now, let us assume that (12) has a solution $\xi\in\left(0,\sqrt{\frac{B}{AM}}\right)$. Suppose h_0 does not verify the inequality (23), that is $\frac{\delta_1}{K_0}-\delta_2\leq 0$. As LHS of (12) is strictly decreasing in $\left(0,\sqrt{\frac{B}{AM}}\right)$, follows that LHS of (12) is negative in $\left(0,\sqrt{\frac{B}{AM}}\right)$. Since RHS of (12) is positive in $\left(0,\sqrt{\frac{B}{AM}}\right)$, we have a contradiction. Therefore $h_0>\frac{A}{B}\frac{k_F}{\sqrt{\pi d_F}}$ and the thesis holds.

□

REMARK 3.2. *The inequality (23) was obtained in [20] for the particular case* $\rho=0$.

The following corollary summarizes previous results.

COROLLARY 3.3. *If* $M>0$, $N>0$, $p\leq 1$ *and* $h_0>\frac{A}{B}\frac{k_F}{\sqrt{\pi d_F}}$ *then the free boundary value problem (1)-(7) has a unique similarity solution of the type (9)-(11) and the dimensionless coefficient* ξ *is a strictly increasing function of the parameter* h_0 *on the interval* $\left(\frac{A}{B}\frac{k_F}{\sqrt{\pi d_F}},+\infty\right)$.

THEOREM 3.4. *If* $M>0$ *and* $N<0$ *then, if:*

$$h_0>\frac{A}{B}\frac{k_F}{\sqrt{\pi d_F}} \qquad \text{and} \qquad B<\frac{AM}{|N|},$$

then equation (12) has at least one solution ξ, which satisfies $0 < \xi < \sqrt{\frac{B}{AM}}$.

Proof. Let $M > 0$ and $N < 0$. We have that LHS of (12) is positive in $(0, q_1)$, where $q_1 < \sqrt{\frac{B}{AM}}$. On the other hand, RHS of (12) is positive in $\left(0, \sqrt{\frac{1}{|N|}}\right)$. Since $\sqrt{\frac{B}{AM}} < \sqrt{\frac{1}{|N|}}$, we can find a solution ξ of (12) which satisfies $0 < \xi < \sqrt{\frac{B}{AM}}$. □

THEOREM 3.5. *If $M < 0$ and $N > 0$ then:*

1. *If:*
$$h_0 > \frac{A}{B}\frac{k_F}{\sqrt{\pi d_F}}$$
and the LHS of (12) has positive zeros being q_1 the smaller one, then equation (12) has at least one solution ξ, which satisfies $0 < \xi < q_1$.

2. *If:*
$$h_0 \leq \frac{A}{B}\frac{k_F}{\sqrt{\pi d_F}} \qquad \text{and} \qquad N < \delta_2\sqrt{\pi}|M|\gamma_0,$$
then equation (12) has at least one positive solution.

Proof. Let $M < 0$ and $N > 0$.

1. It is analogous to the proof of the Theorem 3.1 (1).
2. Let us assume that $h_0 \leq \frac{A}{B}\frac{k_F}{\sqrt{\pi d_F}}$ and $N < \delta_2\sqrt{\pi}|M|\gamma_0$. From the asymptotic behavior of $\left(1 - \frac{AM}{B}y^2\right)G_1(p, y)$ and $(1 + My^2)G_2(y)$ when y tends to $+\infty$ (see Appendix), we have that:
$$\lim_{y\to+\infty} \frac{\delta_1\left(1 - \frac{AM}{B}y^2\right)G_1(p, y) - \delta_2(1 + My^2)G_2(y)}{y + Ny^3} =$$
$$= \begin{cases} -\frac{\delta_2\sqrt{\pi}\gamma_0 M}{N} & p \leq 2 \\ -\frac{AM(p-2)}{BN} - \frac{\delta_2\sqrt{\pi}\gamma_0 M}{N} & p > 2 \end{cases}.$$
Therefore:
$$\lim_{y\to+\infty} \frac{\delta_1\left(1 - \frac{AM}{B}y^2\right)G_1(p, y) - \delta_2(1 + My^2)G_2(y)}{y + Ny^3} > \frac{\delta_2\sqrt{\pi}\gamma_0|M|}{N}.$$
Then, it follows from the last inequality that the LHS of (12) tends to $+\infty$ faster than the RHS when y tends to $+\infty$. Now, taking into account that:
$$\lim_{y\to 0^+} \delta_1\left(1 - \frac{AM}{B}y^2\right)G_1(p, y) - \delta_2(1 + My^2)G_2(y) = \frac{\delta_1}{K_0} - \delta_2 < 0,$$
we have that equation (12) has at least one positive solution.

$\square$

REMARK 3.6. *When we have the condition $h_0 \leq \frac{A}{B}\frac{k_F}{\sqrt{\pi d_F}}$ in the classical Stefan problem with $\rho = 0$ we only obtain a heat transfer problem without a phase-change process [20].*

THEOREM 3.7. *If $M < 0$ and $N < 0$ then:*

1. *If:*
$$h_0 > \frac{A}{B}\frac{k_F}{\sqrt{\pi d_F}}$$
and the LHS of the equation (12) has positive zeros being q_1 the smaller one which satisfies:
   (a) *$q_1 < \sqrt{1/|N|}$, then equation (12) has at least two positive solutions, one of them satisfies $0 < \xi < \sqrt{1/|N|}$).*
   (b) *$q_1 = \sqrt{1/|N|}$, then equation (12) has $\xi = q_1$ as solution.*
2. *If:*
$$h_0 < \frac{A}{B}\frac{k_F}{\sqrt{\pi d_F}}$$
then equation (12) has at least one positive solution ξ.

Proof. Let $M < 0$ and $N < 0$.

1. (a) Assume that $h_0 > \frac{A}{B}\frac{k_F}{\sqrt{\pi d_F}}$.
We have proved in the Appendix that:
$$\lim_{y\to+\infty}\left(1 - \frac{AM}{B}y^2\right)G_1(p, y) = \begin{cases} 0 & p < 2 \\ +\infty & p \geq 2 \end{cases}$$
and
$$\lim_{y\to+\infty}(1 + My^2)G_2(y) = -\infty.$$
Then:
$$\lim_{y\to+\infty}\left[\delta_1\left(1 - \frac{AM}{B}y^2\right)G_1(p, y) - \delta_2(1 + My^2)G_2(y)\right] = +\infty.$$
Furthermore:
$$\lim_{y\to 0^+}\left[\delta_1\left(1 - \frac{AM}{B}y^2\right)G_1(p, y) - \delta_2(1 + My^2)G_2(y)\right] = \frac{\delta_1}{K_0} - \delta_2.$$
On the other hand, the RHS of (12) is positive in $\left(0, \sqrt{\frac{1}{|N|}}\right)$ and negative in $\left(\sqrt{\frac{1}{|N|}}, +\infty\right)$. Then, since $\frac{\delta_1}{K_0} - \delta_2 > 0$ and $q_1 < \sqrt{1/|N|}$

we have that equation (12) has at least two positive solutions, one of them satisfies $0 < \xi < \sqrt{\frac{1}{|N|}}$.

(b) It is inmediate.

2. Now, let us assume that $h_0 < \frac{A}{B}\frac{k_F}{\sqrt{\pi d_F}}$. We have:

$$\lim_{y\to+\infty}\left[\delta_1\left(1-\frac{AM}{B}y^2\right)G_1(p,y)-\delta_2(1+My^2)G_2(y)\right]=+\infty\,.$$

Furthermore:

$$\lim_{y\to 0^+}\left[\delta_1\left(1-\frac{AM}{B}y^2\right)G_1(p,y)-\delta_2(1+My^2)G_2(y)\right]=\frac{\delta_1}{K_0}-\delta_2.$$

Taking into account that $\lim_{y\to 0^+}\left(y+Ny^3\right)=0$ and $\lim_{y\to+\infty}\left(y+Ny^3\right)=-\infty$, since $\frac{\delta_1}{K_0}-\delta_2<0$ we have that equation (12) has at least one positive solution ξ.

□

REMARK 3.8. *Previous results imply that under the conditions specified in each case it is posible to find a similarity solution of the problem (1)-(7). Moreover, that solution guarantees a change of phase, that is,* $AM\xi^2 < u(0,t) < B$.

4. Relationship between the solutions of the Stefan problem with convective and temperature boundary conditions

In this section, we analyze the relationship between problem (1)-(7) and problem (1)-(6),(8) studied in [8], corresponding to a temperature condition at the fixed face $x = 0$.

In [8] it was proved that if $M > 0$, $N > 0$ and $p \leq 2$, then the problem (1)-(6),(8) has a unique similarity solution of the type:

$$U(x,t)=B_0+\frac{AM\omega^2-B_0}{g(p,y)}\int_0^{\frac{x}{2\alpha_U\sqrt{t}}}\exp(-r^2+pr\omega)\,dr \tag{24}$$

$$V(x,t)=\frac{AM\omega^2\operatorname{erf}\left(\frac{x}{2\alpha_F\sqrt{t}}\right)+A\left(\operatorname{erf}(\gamma_0\omega)-\operatorname{erf}\left(\frac{x}{2\alpha_F\sqrt{t}}\right)\right)}{\operatorname{erfc}(\gamma_0\omega)} \tag{25}$$

$$S(t)=2\omega\alpha_U\sqrt{t} \tag{26}$$

where ω is the unique positive solution of the trascendental equation:

$$\widetilde{G}(M,p,y)=y+Ny^3, \tag{27}$$

with:

$$\widetilde{G}(M,p,y)=\tilde{\delta}_1\left(1-\frac{AM}{B_0}y^2\right)\widetilde{G_1}(p,y)-\delta_2(1+My^2)G_2(M,y)\qquad y>0,\tag{28}$$

$$\widetilde{G}_1(p,y)=\frac{\exp\left((p-1)y^2\right)}{g(p,y)}\qquad p\in\mathbb{R},\ y>0,\tag{29}$$

$$\tilde{\delta}_1=\frac{k_U B_0}{2\alpha\,\alpha_U^2}.\tag{30}$$

Furthermore, $0<\omega<\sqrt{\frac{B_0}{AM}}$.

Henceforth, we will only deal with situations in which existence and uniqueness of similarity solutions of type (9)-(11) for problem (1)-(8) or of type (24)-(26) for problem (1)-(6),(8), are guarantee.

Theorem 4.1. *If $M>0$, $N>0$, $p\leq 1$ and $h_0>\frac{A}{B}\frac{k_F}{\sqrt{\pi d_F}}$ then the dimensionless coefficient ξ which characterizes the free boundary of the problem (1)-(7) satisfies:*

$$0<\xi(h_0)<\omega_\infty\qquad\qquad \forall\, h_0\in\left(\frac{A}{B}\frac{k_F}{\sqrt{\pi d_F}},+\infty\right)\tag{31}$$

where ω_∞ is the coefficient which characterizes the free boundary of the problem (1)-(6),(8) when the temperature condition is given by B.

Proof. Assume that $M>0$, $N>0$, $p\leq 1$ and $h_0>\frac{A}{B}\frac{k_F}{\sqrt{\pi d_F}}$. We know from Corollary 3.3 that the dimensionless coefficient ξ is a strictly increasing function of the coefficient h_0 on the interval $\left(\frac{A}{B}\frac{k_F}{\sqrt{\pi d_F}},+\infty\right)$. We also know that ξ satisfies equation (12) which became:

$$\delta_1\left(1-\frac{AM}{B_0}y^2\right)\widetilde{G_1}(p,y)-\delta_2(1+My^2)G_2(M,y)=y+Ny^3\tag{32}$$

when h_0 tends to $+\infty$. Only remains to note that this last equation (32) has a unique solution ω_∞ because (32) is the corresponding equation to problem (1)-(6),(8) when $B_0=B$, which has a unique similarity solution under the hypothesis considered here. □

Theorem 4.2. *If $M>0$, $N>0$, $p\leq 1$ and $h_0>\frac{A}{B}\frac{k_F}{\sqrt{\pi d_F}}$ then the similarity solution (9)-(11) of the the problem (1)-(7), coincide with the similarity solution (24)-(26) of the problem (1)-(6),(8), when the external boundary temperature at $x=0$ is defined as:*

$$B_0=\frac{2h_0\alpha_U Bg(p,\xi)+AMk_U\xi^2}{2h_0\alpha_U g(p,\xi)+k_U}\tag{33}$$

Moreover, the dimensionless parameters ξ and ω which characterize the free boundary in each problem, are equals.

REMARK 4.3. *Note that B_0 is positive since $B_0 = \frac{2h_0\alpha_U Bg(p,\xi)+AMk_U\xi^2}{2h_0\alpha_U g(p,\xi)+k_U} = u(0,t)$, $t>0$, and $u(0,t)>0$.*

Proof. Assume that $M>0$, $N>0$, $p\le 1$, $h_0 > \frac{A}{B}\frac{k_F}{\sqrt{\pi d_F}}$ and B_0 is defined as in (33). First of all, we note that function $\widetilde{G}(M,p,y)$, given in (28), can be written as:

$$\widetilde{G}(M,p,y) = \delta_1\left(\frac{B_0}{B} - \frac{AM}{B}y^2\right)\widetilde{G_1}(p,y) - \delta_2(1+My^2)G_2(y) \qquad y>0.$$

Because of the definition of B_0, given in (33), we have:

$$\frac{B_0}{B} - \frac{AM}{B}y^2 = \frac{g(p,\xi)\left(1-\frac{AM}{B}y^2\right)+\frac{AM}{B}K_0(\xi^2-y^2)}{g(p,\xi)+K_0}.$$

Then, equation (27) can be written as:

$$\begin{aligned}\delta_1\left(\frac{g(p,\xi)}{g(p,y)}\left(1-\frac{AM}{B}y^2\right)+\frac{AMK_0}{Bg(p,y)}(\xi^2-y^2)\right)G_1(p,y)\\ -\delta_2(1+My^2)G_2(y) = y+Ny^3.\end{aligned} \tag{34}$$

Now, since ξ satisfies equation (12), it follows that ξ satisfies equation (34). Therefore, ξ coincide with the coefficient ω which characterize the free boundary of the problem (1)-(8). Finally, from elementary calculations, we have that similarity solutions given in (9)-(11) and (24)-(26) are coincident. □

In the same way that the proof of the previous theorem, it can be shown the following result.

THEOREM 4.4. *If $M>0$, $N>0$ and $p\le 1$ then the similarity solution (24)-(26) of the problem (1)-(6),(8) coincide with the similarity solution (9)-(11) of the problem (1)-(7), when the coefficient which characterizes the heat transfer at the fixed face $x=0$ is defined by:*

$$h_0 = \frac{k_U(B_0 - AM\omega^2)}{2\alpha_U(B-B_0)g(p,\omega)} \tag{35}$$

and the boundary temperature B at $x=0$ is such that $B>B_0$.

Moreover, the dimensionless parameters ω and ξ which define the free boundary in each problem, are equals.

REMARK 4.5. *Note that* h_0 *is a positive number since* $B_0 - AM\omega^2 > 0$ *and* $B - B_0 > 0$.

We can conclude now that in the sense established in Theorems 4.2 and 4.4, Stefan problems with convective and temperature conditions at the fixed face $x = 0$ given by (1)-(7) and (1)-(6),(8), respectively, are equivalent when inequality (23) is verified by data.

THEOREM 4.6. *If* $M > 0$, $N > 0$ *and* $p \leq 1$ *then the coefficient* ω *which characterizes the free boundary of the problem (1)-(6),(8) satisfies the following inequality:*

$$\frac{B_0 - AM\omega^2}{g(p,\omega)} > \frac{2\alpha_U k_F A(B - B_0)}{\alpha_F k_U B\sqrt{\pi}}, \qquad \forall\, B > B_0 \tag{36}$$

Proof. Assume that $M > 0$, $N > 0$ and $p \leq 1$, and let $B > B_0$. We know from Theorem 4.4, that the problem (1)-(7) has a similarity solution of the type (9)-(11) when the external boundary temperature is given by B and the coefficient h_0 is defined as in (35). We also know that the coefficients which characterize the free boundary in problems (1)-(7) and (1)-(6), (8) are equals, that is $\xi = \omega$. Therefore, since $0 < \omega < \sqrt{\frac{B_0}{AM}}$ and $B > B_0$, we have that $0 < \xi < \sqrt{\frac{B}{AM}}$. Then, due to Theorem 3.1, part 2, inequality (23) holds. Only remains to note that inequality (23) becames inequality (36) when h_0 is defined as in (35). $\square$

By taking limit when B tends to $+\infty$ into both sides of inequality (36), we have the following corollary.

COROLLARY 4.7. *If* $M > 0$, $N > 0$ *and* $p \leq 1$ *then the coefficient* ω *which characterizes the free boundary of the problem (1)-(6),(8) satisfies the following inequality:*

$$\frac{B_0 - AM\omega^2}{g(p,\omega)} > \frac{2\alpha_U A k_F}{\alpha_F k_U \sqrt{\pi}} \tag{37}$$

REMARK 4.8. *For the classical Stefan problem with* $\rho = 0$ *the inequality (37) for the coefficient* ω*, which characterizes the free boundary* $s(t)$*, given by (26), become:*

$$erf(\omega) < \frac{B_0}{A}\frac{k_U}{k_F}\sqrt{\frac{d_F}{d_U}}$$

which was obtained in [19].

5. Appendix

PROPOSITION 5.1. *For any $M \in \mathbb{R}$ and $p \in \mathbb{R}$:*

1. *If $p < 2$ then:*
$$\lim_{y\to+\infty}\left(1-\frac{AM}{B}y^2\right)G_1(p,y)=0. \tag{38}$$

2. *If $p \geq 2$ then:*
$$\lim_{y\to+\infty}\left(1-\frac{AM}{B}y^2\right)G_1(p,y)=\begin{cases} -\infty & M>0 \\ +\infty & M<0 \end{cases} \tag{39}$$
and
$$\left(1-\frac{AM}{B}y^2\right)G_1(p,y)\simeq\begin{cases} -\dfrac{2AM}{B\sqrt{\pi}}y^2 & p=2 \\ \\ -\dfrac{AM(p-2)}{B}y^3 & p>2 \end{cases} \tag{40}$$
when $y \to +\infty$.

Proof. Let M and p real numbers.

1. Proof of (38) is an immediate consecuence of the following facts:
$$\lim_{y\to+\infty} G_1(p,y)=0 \qquad \text{and} \qquad \lim_{y\to+\infty} y^2G_1(p,y)=0 \tag{41}$$
Then, we will prove (41). We split the proof into three cases depending on the sign of the parameter p.
   (a) $0 < p \leq 1$
   Since $\lim\limits_{y\to+\infty} g(p,y)=+\infty$ (see [8]), we have:
$$\lim_{y\to+\infty} G_1(p,y)=0$$
   On the other hand,
$$\lim_{y\to+\infty}\left(y^2G_1(p,y)\right)^{-1} = \lim_{y\to+\infty}\left(\frac{K_0}{y^2\exp((p-1)y^2)}+\frac{g(p,y)}{y^2\exp((p-1)y^2)}\right)=+\infty$$
   since $\lim\limits_{y\to+\infty}\dfrac{g(p,y)}{y^2}=+\infty$ (see [10]). Then, $\lim\limits_{y\to+\infty} y^2G_1(p,y)=0$.

(b) $1<p<2$
First of all, we note that:

$$g(p,y)=\frac{\sqrt{\pi}}{2}\exp\left(\left(\frac{p}{2}y^2\right)^2\right)\left(\operatorname{erf}\left(\frac{p}{2}y\right)+\operatorname{erf}\left(\frac{2-p}{2}y\right)\right).$$

Then:

$$\lim_{y\to+\infty}G_1(p,y)=$$
$$\lim_{y\to+\infty}\frac{\exp\left(-\left(\frac{p}{2}-1\right)^2y^2\right)}{K_0\exp\left(-\left(\frac{p}{2}y\right)^2\right)+\frac{\sqrt{\pi}}{2}}\left(\operatorname{erf}\left(\frac{p}{2}y\right)+\operatorname{erf}\left(\frac{2-p}{2}y\right)\right)=0.$$

The proof of $\lim\limits_{y\to+\infty}y^2G_1(p,y)=0$ is similar to the previous one.

(c) $p\leq 0$
It is useful to write $G_1(p,y)$ in the following way:

$$G_1(p,y)=\Big[K_0\exp((1-p)y^2)+$$
$$\frac{\sqrt{\pi}}{2}\exp\left(\left(1-\frac{p}{2}\right)^2y^2\right)\left(\operatorname{erf}\left(\left(1-\frac{p}{2}\right)y\right)-\operatorname{erf}\left(-\frac{p}{2}y\right)\right)\Big]^{-1}.$$

Then, since:

$$\lim_{y\to+\infty}\Big[K_0\exp((1-p)y^2)+$$
$$\frac{\sqrt{\pi}}{2}\exp\left(\left(1-\frac{p}{2}\right)^2y^2\right)\left(\operatorname{erf}\left(\left(1-\frac{p}{2}\right)y\right)-\operatorname{erf}\left(-\frac{p}{2}y\right)\right)\Big]=+\infty,$$

we have that $\lim\limits_{y\to+\infty}G_1(p,y)=0$.

The proof of $\lim\limits_{y\to+\infty}y^2G_1(p,y)=0$ is similar to the previous one.

2. Now, we have:

$$\lim_{y\to+\infty}y^2G_1(p,y)=0. \tag{42}$$

In fact:

$$\lim_{y\to+\infty}G_1(p,y)=\lim_{y\to+\infty}\frac{2(p-1)y\exp((p-1)y^2)}{\frac{\partial}{\partial y}g(p,y)}\geq$$
$$\lim_{y\to+\infty}\frac{2(p-1)y\exp((p-1)y^2)}{\exp((p-1)y^2)+pyg(p,y)}$$

and

$$\lim_{y\to+\infty}\left(\frac{2(p-1)y\exp((p-1)y^2)}{\exp((p-1)y^2)+pyg(p,y)}\right)^{-1} = \lim_{y\to+\infty}\frac{1}{2(p-1)y}+\frac{1}{2y(p-1)}\frac{yg(p,y)}{\exp((p-1)y^2)}=0.$$

For the last limit, we are using the fact that $\lim_{y\to+\infty}\frac{yg(p,y)}{\exp((p-1)y^2)} = \frac{1}{p-2}$ (see [10]). Then $\lim_{y\to+\infty} y^2G_1(p,y)=0$.

Finally, (40) follows from similar arguments.

□

The following result is proved in [8].

PROPOSITION 5.2. *We have:*

$$\lim_{y\to+\infty}(1+My^2)G_2(y)=\begin{cases} +\infty & M>0 \\ -\infty & M<0 \end{cases} \tag{43}$$

and

$$(1+My^2)G_2(y)\simeq\sqrt{\pi}\gamma_0 My^3 \text{ as } y\to+\infty. \tag{44}$$

Furthermore, when $M>0$, $(1+My^2)G_2(y)$ is an increasing function of $y>0$.

Acknowledgment

This paper has been partially sponsored by the Project PIP No. 0534 from CONICET-UA and Grant from Universidad Austral (Rosario, Argentina).

REFERENCES

[1] V. ALEXIADES AND A. D. SOLOMON, *Mathematical modeling of melting and freezing processes*, Hemisphere Publishing Corp., Washington, 1993.

[2] A. C. BRIOZZO AND M. F. NATALE, *On a non-linear moving boundary problem for a difussion-convection equation*, J. Nonlinear Mech. **47** (2012), 712–718.

[3] A. C. BRIOZZO AND M. F. NATALE, *One-dimensional nonlinear Stefan problems in storm's materials*, Mathematics **2** (2014), 1–11.

[4] H. S. CARSLAW AND J. C. JAEGER, *Conduction of heat in solids*, Clarendon Press, Oxford, 1959.

[5] CH. CHARACH AND I. RUBINSTEIN, *Pressure-temperature effects in planar Stefan problems with density change*, J. Appl. Physics **71** (1992), 1128–1137.

[6] R. Cherniha and S. Kovalenko, *Lie symmetries of nonlinear boundary value problems*, Commun. Nonlinear Sci. Numer. Simul. **17** (2012), 71–84.

[7] A. Fasano, Z. Guan, M. Primicerio, and I. Rubinstein, *Thawing in saturated porous media*, Meccanica **28** (1993), 103–109.

[8] A. Fasano, M. Primicerio, and D. A. Tarzia, *Similarity solutions in a class of thawing processes*, Math. Models Methods Appl. Sci. **9** (1999), 1–10.

[9] S. S. Kovalenko, *Symmetry analysis and exact solutions of one class of (1+3)-dimensional boundary-value problems of the Stefan type*, Ukrainian Math. J. **63** (2012), 1534–1543.

[10] A. L. Lombardi and D. A. Tarzia, *Similarity solutions for thawing processes with a heat flux condition at the fixed boundary*, Meccanica **36** (2001), 251–264.

[11] S. L. Mitchel and M. Vynnycky, *On the numerical solution of two-phase Stefan problems with heat-flux boundary conditions*, J. Comput. Appl. Math. **264** (2014), 49–64.

[12] Y. Nakano, *Quasi-steady problems in freezing soils. I. Analysis on the steady growth of an ice layer*, Cold Regions Sci. Tech. **17** (1990), 207–226.

[13] M. F. Natale, E. A. Santillan Marcus, and D. A. Tarzia, *Explicit solutions for one-dimensional two-phase free boundary problems with either shrinkage or expansion*, Nonlinear Anal. Real World Appl. **11** (2010), 1946–1952.

[14] K. O'Neill and R. D. Miller, *Exploration of a rigid ice model of frost heave*, Water Resources Research **21** (1985), 281–296.

[15] S. Roscani and E. Santillan Marcus, *A new equivalence of Stefan's problems for the time-fractional-difussion equation*, Fract. Calc. Appl. Anal. **17** (2014), 371–381.

[16] L. Rubinstein, *The Stefan problem*, American Mathematical Society, Providence, 1971.

[17] N. N. Salva and D. A. Tarzia, *Explicit solution for a Stefan problem with variable latent heat and constant heat flux boundary conditions*, J. Math. Anal. Appl. **379** (2011), 240–244.

[18] F. Talamucci, *Analysis of the coupled heat-mass transport in freezing porous media*, Surv. Math. Industry **7** (1997), 93–139.

[19] D. A. Tarzia, *An inequality for the coefficient σ of the free boundary $s(t) = 2\sigma\sqrt{t}$ of the Neumann solution for the two-phase Stefan problem*, Quart. Appl. Math. **39** (1981-1982), 491–497.

[20] D. A. Tarzia, *An explicit solution for a two-phase unidimensional Stefan problem with a convective boundary condition at the fixed phase*, MAT-Serie A **8** (2004), 21–27.

[21] V. R. Voller, *Fractional Stefan problems*, Int. J. Heat Mass Transfer **74** (2014), 269–277.

[22] V. R. Voller and F. Falcini, *Two exact solutions of a Stefan problem with varying diffusivity*, Int. J. Heat Mass Transfer **58** (2013), 80–85.

[23] H. M. Weber, *Die partiellen differential-gleinchungen der mathematischen physik nach riemann's vorlesungen*, Braunschweig, 1901.

[24] Y. Zhou, Yi jiang Wang, and Wan kui Bu, *Exact solution for a Stefan problem with a latent heat a power functoin of position*, Int. J. Heat Mass Transfer **69** (2014), 451–454.

[25] S. M. Zubair and M. A. Chaudhry, *Exact solution of solid-liquid phase-change heat transfer when subjected to convective boundary conditions*, Wärme und Stoffübertragung **30** (1994), 77–81.

Authors' addresses:

Andrea N. Ceretani
CONICET-Depto. Matemática, Facultad de Ciencias Empresariales,
Universidad Austral, Paraguay 1950,
S2000FZF Rosario, Argentina.
E-mail: `aceretani@austral.edu.ar`

Domingo A. Tarzia
CONICET-Depto. Matemática, Facultad de Ciencias Empresariales,
Universidad Austral, Paraguay 1950,
S2000FZF Rosario, Argentina.
E-mail: `dtarzia@austral.edu.ar`

Received May 30, 2014
Revised July 29, 2014

Rend. Istit. Mat. Univ. Trieste
Volume 46 (2014), 157–180

Stratonovich-Weyl correspondence via Berezin quantization

Benjamin Cahen

Abstract. *Let G be a quasi-Hermitian Lie group and let K be a maximal compactly embedded subgroup of G. Let π be a unitary representation of G which is holomorphically induced from a unitary representation ρ of K. We introduce and study a notion of complex-valued Berezin symbol for an operator acting on the space of π and the corresponding notion of Stratonovich-Weyl correspondence. This generalizes some results already obtained in the case when ρ is a unitary character, see [19]. As an example, we treat in detail the case of the Heisenberg motion groups.*

Keywords: Stratonovich-Weyl correspondence, Berezin quantization, Berezin transform, quasi-Hermitian Lie group, coadjoint orbit, unitary representation, holomorphic representation, reproducing kernel Hilbert space, Heisenberg motion group.
MS Classification 2010: 22E46, 22E10, 22E20, 32M05, 32M10, 81S10, 46E22, 32M15.

1. Introduction

There are different ways to extend the usual Weyl correspondence between functions on $\mathbb{R}^{2n}$ and operators on $L^2(\mathbb{R}^n)$ to the general setting of a Lie group acting on a homogeneous space [1, 13, 29]. In this paper, we focuse on Stratonovich-Weyl correspondences. The notion of Stratonovich-Weyl correspondence was introduced in [42] and its systematic study began with the work of J.M. Gracia-Bondìa, J.C. Vàrilly and their co-workers (see [11, 23, 25, 27, 28]). The following definition is taken from [27], see also [28].

Definition 1.1. *Let G be a Lie group and π a unitary representation of G on a Hilbert space $\mathcal{H}$. Let M be a homogeneous G-space and let μ be a (suitably normalized) G-invariant measure on M. Then a Stratonovich-Weyl correspondence for the triple (G, π, M) is an isomorphism W from a vector space of operators on $\mathcal{H}$ to a space of (generalized) functions on M satisfying the following properties:*

1. *W maps the identity operator of $\mathcal{H}$ to the constant function* 1*;*

2. *the function $W(A^*)$ is the complex-conjugate of $W(A)$;*
3. *Covariance: we have $W(\pi(g)\,A\,\pi(g)^{-1})(x) = W(A)(g^{-1}\cdot x)$;*
4. *Traciality: we have*

$$\int_M W(A)(x)W(B)(x)\,d\mu(x) = \mathrm{Tr}(AB).$$

A basic example is the case when G is the $(2n+1)$-dimensional Heisenberg group H_n acting on $\mathbb{R}^{2n}$ by translations and π is a Schrödinger representation of H_n on $L^2(\mathbb{R}^n)$. In this case, the usual Weyl correspondence (see [26]) provides a Stratonovich-Weyl correspondence for the triple $(H_n, \pi, \mathbb{R}^{2n})$ [6, 40, 44].

Stratonovich-Weyl correspondences were constructed for various Lie group representations, in particular for the massive representations of the Poincaré group [23, 27].

In [19], we constructed and studied a Stratonovich-Weyl correspondence for a quasi-Hermitian Lie group G and a unitary representation π of G which is holomorphically induced from a unitary character of a compactly embedded subgroup K of G (see also [15] and [16]). In this case, M is taken to be a coadjoint orbit of G which is associated with π by the Kirillov-Kostant method of orbits [33, 34] and we can consider the Berezin calculus on M [9, 10]. Recall that the Berezin map S is an isomorphism from the Hilbert space of all Hilbert-Schmidt operators on $\mathcal{H}$ (endowed with the Hilbert-Schmidt norm) onto a space of square-integrable functions on a homogeneous complex domain [43]. In this situation, we can apply an idea of [25] (see also [3] and [4]) and construct a Stratonovich-Weyl correspondence for (G, π, M) by taking the isometric part W in the polar decomposition of S, that is, $W := (SS^*)^{-1/2}S$. Note that $B := SS^*$ is the so-called Berezin transform which have been intensively studied by many authors, see in particular [24, 38, 39, 43, 46].

In [19], we also showed that if the Lie algebra $\mathfrak{g}$ of G is reductive then W can be extended to a class of functions which contains $S(d\pi(X))$ for each $X \in \mathfrak{g}$ and that, for each simple ideal $\mathfrak{s}$ in $\mathfrak{g}$, there exists a constant $c \geq 0$ such that $W(d\pi(X)) = cS(d\pi(X))$ for each $X \in \mathfrak{s}$. Similar results have been obtained for different examples of non-reductive Lie groups, see in particular [21].

On the other hand, in [17] and [18] we also obtained a Stratonovich-Weyl correspondence for a non-scalar holomorphic discrete series representation of a semi-simple Lie group by introducing a generalized Berezin map.

In the present paper, we adapt the method and the arguments of [17] and [18] in order to generalize the results of [19] to the case when π is holomorphically induced from a unitary representation ρ of K (in a finite-dimensional vector space V) which is not necessarily a character. More precisely, we prove that the coadjoint orbit $\mathcal{O}$ of G associated with π is diffeomorphic to the product $\mathcal{D}\times o$ where $\mathcal{D}$ is a complex domain and o is the coadjoint orbit of K associated with ρ. Then, following [17], we introduce a Berezin calculus for $\mathrm{End}(V)$-valued functions on $\mathcal{D}$. By combining this calculus with the usual Berezin calculus s on o, we obtain a Berezin calculus S on $\mathcal{O}$ which is G-equivariant with respect to π. Thus, we get a Stratonovich-Weyl correspondence for the triple $(G, \pi, \mathcal{O})$ by taking the isometric part of S.

As an illustration, we consider the case when G is a Heisenberg motion group, that is, the semi-direct product of the $(2n+1)$-Heisenberg group H_n with a compact subgroup of the unitary group $U(n)$. Note that Heisenberg motion groups play an important role in the theory of Gelfand pairs, since the study of a Gelfand pair of the form (K_0, N) where K_0 is a compact Lie group acting by automorphisms on a nilpotent Lie group N can be reduced to that of the form (K_0, H_n) [7, 8].

In this case, the space $\mathcal{H}$ of π can be decomposed as $\mathcal{H}_0 \otimes V$ where $\mathcal{H}_0$ is the Fock space and we show that for each operator A on $\mathcal{H}$ of the form $A_1 \otimes A_2$ we have the decomposition formula $S(A)(Z, \varphi) = S_0(A_1)(Z)s(A_2)(\varphi)$ where S_0 denotes the Berezin calculus on $\mathcal{H}_0$. Moreover, we verify that the Berezin transform takes a simple form and then can be extended to the functions of the form $S(d\pi(X_1 X_2 \cdots X_p))$ for $X_1, X_2, \ldots, X_p \in \mathfrak{g}$ and we compute explicitely $W(d\pi(X))$ for $X \in \mathfrak{g}$.

2. Preliminaries

All the material of this section is taken from the excellent book of K.-H. Neeb, [37, Chapters VIII and XII], (see also [41, Chapter II] and, for the Hermitian case, [30, Chapter VIII] and [31, Chapter 6]).

Let $\mathfrak{g}$ be a real quasi-Hermitian Lie algebra, that is, a real Lie algebra for which the centralizer in $\mathfrak{g}$ of the center $\mathcal{Z}(\mathfrak{k})$ of a maximal compactly embedded subalgebra $\mathfrak{k}$ coincides with $\mathfrak{k}$ [37, p. 241]. We assume that $\mathfrak{g}$ is not compact. Let $\mathfrak{g}^c$ be the complexification of $\mathfrak{g}$ and $Z = X + iY \to Z^* = -X + iY$ the corresponding involution. We fix a compactly embedded Cartan subalgebra $\mathfrak{h} \subset \mathfrak{k}$, [37, p. 241], and we denote by $\mathfrak{h}^c$ the corresponding Cartan subalgebra of $\mathfrak{g}^c$. We write $\Delta := \Delta(\mathfrak{g}^c, \mathfrak{h}^c)$ for the set of roots of $\mathfrak{g}^c$ relative to $\mathfrak{h}^c$ and $\mathfrak{g}^c = \mathfrak{h}^c \oplus \sum_{\alpha\in\Delta} \mathfrak{g}_\alpha$ for the root space decomposition of $\mathfrak{g}^c$. Note that $\alpha(\mathfrak{h}) \in i\mathbb{R}$ for each $\alpha \in \Delta$ [37, p. 233]. Recall that a root $\alpha \in \Delta$ is called compact if $\alpha([Z, Z^*]) > 0$ holds for some element $Z \in \mathfrak{g}_\alpha$. All other roots are called non-compact [37, p. 235]. We write Δ_k, respectively Δ_p, for the set of compact, respectively non-compact, roots. Note that $\mathfrak{k}^c = \mathfrak{h}^c \oplus \sum_{\alpha\in\Delta_k} \mathfrak{g}_\alpha$ [37, p. 235]. Recall also that a subset $\Delta^+ \subset \Delta$ is called a positive system if there exists an element $X_0 \in i\mathfrak{h}$ such that $\Delta^+ = \{\alpha \in \Delta : \alpha(X_0) > 0\}$ and $\alpha(X_0) \neq 0$ for all $\alpha \in \Delta$. A positive system is then said to be adapted if for $\alpha \in \Delta_k$ and $\beta \in \Delta^+ \cap \Delta_p$ we have $\beta(X_0) > \alpha(X_0)$, [37, p. 236]. Here we fix a positive adapted system Δ^+ and we set $\Delta_p^+ := \Delta^+ \cap \Delta_p$ and $\Delta_k^+ := \Delta^+ \cap \Delta_k$, see [37,

p. 241].

Let G^c be a simply connected complex Lie group with Lie algebra $\mathfrak{g}^c$ and $G \subset G^c$, respectively, $K \subset G^c$, the analytic subgroup corresponding to $\mathfrak{g}$, respectively, $\mathfrak{k}$. We also set $K^c = \exp(\mathfrak{k}^c) \subset G^c$ as in [37, p. 506].

Let $\mathfrak{p}^+ = \sum_{\alpha \in \Delta_p^+} \mathfrak{g}_\alpha$ and $\mathfrak{p}^- = \sum_{\alpha \in \Delta_p^+} \mathfrak{g}_{-\alpha}$. We denote by P^+ and P^- the analytic subgroups of G^c with Lie algebras $\mathfrak{p}^+$ and $\mathfrak{p}^-$. Then G is a group of the Harish-Chandra type [37, p. 507], that is, the following properties are satisfied:

1. $\mathfrak{g}^c = \mathfrak{p}^+ \oplus \mathfrak{k}^c \oplus \mathfrak{p}^-$ is a direct sum of vector spaces, $(\mathfrak{p}^+)^* = \mathfrak{p}^-$ and $[\mathfrak{k}^+, \mathfrak{p}^\pm] \subset \mathfrak{p}^\pm$;
2. The multiplication map $P^+K^cP^- \to G^c$, $(z, k, y) \to zky$ is a biholomorphic diffeomorphism onto its open image;
3. $G \subset P^+K^cP^-$ and $G \cap K^cP^- = K$.

Moreover, there exists an open connected K-invariant subset $\mathcal{D} \subset \mathfrak{p}^+$ such that one has $GK^cP^- = \exp(\mathcal{D})K^cP^-$, [37, p. 497]. We denote by $\zeta : P^+K^cP^- \to P^+$, $\kappa : P^+K^cP^- \to K^c$ and $\eta : P^+K^cP^- \to P^-$ the projections onto P^+-, K^c- and P^--component. For $Z \in \mathfrak{p}^+$ and $g \in G^c$ with $g \exp Z \in P^+K^cP^-$, we define the element $g \cdot Z$ of $\mathfrak{p}^+$ by $g \cdot Z := \log \zeta(g \exp Z)$. Note that we have $\mathcal{D} = G \cdot 0$.

We also denote by $g \to g^*$ the involutive anti-automorphism of G^c which is obtained by exponentiating $X \to X^*$. We denote by $p_{\mathfrak{p}^+}$, $p_{\mathfrak{k}^c}$ and $p_{\mathfrak{p}^-}$ the projections of $\mathfrak{g}^c$ onto $\mathfrak{p}^+$, $\mathfrak{k}^c$ and $\mathfrak{p}^-$ associated with the direct decomposition $\mathfrak{g}^c = \mathfrak{p}^+ \oplus \mathfrak{k}^c \oplus \mathfrak{p}^-$.

The G-invariant measure on $\mathcal{D}$ is $d\mu(Z) := \chi_0(\kappa(\exp Z^* \exp Z))\, d\mu_L(Z)$ where χ_0 is the character on K^c defined by $\chi_0(k) = \mathrm{Det}_{\mathfrak{p}^+}(\mathrm{Ad}\, k)$ and $d\mu_L(Z)$ is a Lebesgue measure on $\mathcal{D}$ [37, p. 538].

Now, we construct a section of the action of G on $\mathcal{D}$, that is, a map $Z \to g_Z$ from $\mathcal{D}$ to G such that $g_Z \cdot 0 = Z$ for each $Z \in \mathcal{D}$. Such a section will be needed later. In [20], we proved the following proposition.

Proposition 2.1. *Let $Z \in \mathcal{D}$. There exists a unique element k_Z in K^c such that $k_Z^* = k_Z$ and $k_Z^2 = \kappa(\exp Z^* \exp Z)^{-1}$. Each $g \in G$ such that $g \cdot 0 = Z$ is then of the form $g = \exp(-Z^*)\, \zeta(\exp Z^* \exp Z) k_Z^{-1} h$ where $h \in K$. Consequently, the map $Z \to g_Z := \exp(-Z^*)\, \zeta(\exp Z^* \exp Z) k_Z^{-1}$ is a section for the action of G on $\mathcal{D}$.*

Note that we have

$$\begin{aligned} g_Z &= \exp(-Z^*)\, \zeta(\exp Z^* \exp Z) k_Z^{-1} \\ &= \exp Z \eta(\exp Z^* \exp Z)^{-1} \kappa(\exp Z^* \exp Z)^{-1} k_Z^{-1} \\ &= \exp Z \eta(\exp Z^* \exp Z)^{-1} k_Z \end{aligned}$$

and then $\kappa(g_Z)=k_Z$.

3. Representations

Let (ρ,V) be a (finite-dimensional) unitary irreducible representation of K with highest weight λ (relative to Δ_c^+). We also denote by ρ the extension of ρ to K^c and by $\tilde\rho$ the extension of ρ to K^cP^- which is trivial on P^-. First, we verify that the representation π of G which is associated with ρ as in [37, Proposition XII.2.1], can be obtained by holomorphic induction from ρ.

Let us introduce the Hilbert G-bundle $L:=G\times_\rho V$ over G/K. Recall that an element of L is an equivalence class

$$[g,v]=\{(gk,\rho(k)^{-1}v)\,:\,k\in K\}$$

where $g\in G$, $v\in V$ and that G acts on L by left translations: $g\,[g',v]:=[gg',v]$.

The projection map $[g,v]\to gK$ is then G-equivariant. The G-invariant Hermitian structure on L is given by

$$\langle [g,v],[g,v']\rangle=\langle v,v'\rangle_V$$

where $g\in G$ and v, $v'\in V$.

The space G/K being endowed with the complex structure defined in Section 2, let $\mathcal{H}^0$ be the space of all holomorphic sections s of L which are square-integrable with respect to the invariant measure μ_0 on G/K, that is,

$$\|s\|^2_{\mathcal{H}^0}=\int_{G/K}\langle s(p)\,,\,s(p)\rangle\,d\mu_0(p)<+\infty.$$

We can consider the action π_0 of G on $\mathcal{H}^0$ defined by

$$(\pi_0(g)\,s)(p)=g\,s(g^{-1}p).$$

Recall also that the map $gK\to\log\zeta(g)$ is a diffeomorphism from G/K onto $\mathcal{D}$ (see Section 2) whose inverse is the diffeomorphism σ from $\mathcal{D}$ onto G/K defined by $\sigma(Z)=g_ZK$. We can verify that σ intertwines the natural action of G on G/K and the action of G on $\mathcal{D}$ introduced in Section 2, that is, we have $\sigma(g\cdot Z)=g\sigma(Z)$ for each $Z\in\mathcal{D}$ and each $g\in G$. Then we have $\mu_0=(\sigma^{-1})^*(\mu)$.

Now, we will introduce a realization of π_0 on a space of functions on $\mathcal{D}$. To this aim, we associate with any $s\in\mathcal{H}^0$ the function $f_s:\mathcal{D}\to V$ defined by $s(\sigma(Z))=[g_Z,\tilde\rho(g_Z^{-1}\exp Z)f_s(Z)]$. Then, for each s and s' in $\mathcal{H}^0$, we have

$$\begin{aligned}\langle s(\sigma(Z)),s'(\sigma(Z))\rangle&=\langle\tilde\rho(g_Z^{-1}\exp Z)f_s(Z)\,,\,\tilde\rho(g_Z^{-1}\exp Z)f_{s'}(Z)\rangle_V\\&=\langle\tilde\rho(g_Z^{-1}\exp Z)^*\tilde\rho(g_Z^{-1}\exp Z)f_s(Z),f_{s'}(Z)\rangle_V\\&=\langle\tilde\rho(\kappa(\exp Z^*\exp Z))f_s(Z),f_{s'}(Z)\rangle_V\end{aligned}$$

since $g_Z^* g_Z = e$ (the unit element of G).

This implies that

$$\langle s\,,\, s'\rangle_{\mathcal{H}^0} = \int_{\mathcal{D}} \langle \rho(\kappa(\exp Z^* \exp Z)) f_s(Z)\,,\, f_{s'}(Z)\rangle_V \, d\mu(Z).$$

This leads us to introduce the Hilbert space $\mathcal{H}$ of all holomorphic functions $f : \mathcal{D} \to V$ such that

$$\|f\|^2_{\mathcal{H}} := \int_{\mathcal{D}} \langle \rho(\kappa(\exp Z^* \exp Z)) f(Z)\,,\, f(Z)\rangle_V \, d\mu(Z) < +\infty.$$

On the other hand, for each $s \in \mathcal{H}^0$, $g \in G$ and $Z \in \mathcal{D}$, we have

$$\begin{aligned}(\pi_0(g)\, s)(\sigma(Z)) &= gs(g^{-1}\sigma(Z))\\ &= g\,[g_{g^{-1}\cdot Z}, \tilde{\rho}(g^{-1}_{g^{-1}\cdot Z} \exp(g^{-1}\cdot Z)) f_s(g^{-1}\cdot Z)]\\ &= [g_Z, \tilde{\rho}(g_Z^{-1} g \exp(g^{-1}\cdot Z)) f_s(g^{-1}\cdot Z)].\end{aligned}$$

Then we get

$$\begin{aligned} f_{\pi_0(g)\, s}(Z) &= \tilde{\rho}(g_Z^{-1}\exp Z)^* \tilde{\rho}(g_Z^{-1} g \exp(g^{-1}\cdot Z)) f_s(g^{-1}\cdot Z)\\ &= \tilde{\rho}(\exp(-Z) g \exp(g^{-1}\cdot Z)) f_s(g^{-1}\cdot Z).\end{aligned}$$

Now, noting that

$$g^{-1}\exp Z = \exp(g^{-1}\cdot Z)\kappa(g^{-1}\exp Z)\eta(g^{-1}\exp Z),$$

we obtain

$$f_{\pi_0(g)\, s}(Z) = \rho(\kappa(g^{-1}\exp Z))^{-1} f_s(g^{-1}\cdot Z).$$

Let $J(g, Z) := \rho(\kappa(g \exp Z))$ for $g \in G$ and $Z \in \mathcal{D}$. Hence we can conclude that the equality

$$(\pi(g) f)(Z) = J(g^{-1}, Z)^{-1} f(g^{-1}\cdot Z)$$

defines a unitary representation π of G on $\mathcal{H}$ which is unitarily equivalent to π_0. This is precisely the representation of G introduced in [37, Proposition XII.2.1]. Note also that π is irreducible since ρ is irreducible, [37, p. 515].

We denote $K(Z, W) := \rho(\kappa(\exp W^* \exp Z))^{-1}$ for Z, $W \in \mathcal{D}$. The evaluation maps $K_Z : \mathcal{H} \to V, f \to f(Z)$ are continuous [37, p. 539]. The generalized coherent states of $\mathcal{H}$ are the maps $E_Z = K_Z^* : V \to \mathcal{H}$ defined by $\langle f(Z), v\rangle_V = \langle f, E_Z v\rangle$ for $f \in \mathcal{H}$ and $v \in V$.

We have the following result, see [37, p. 540].

PROPOSITION 3.1. *(1) There exists a constant $c_\rho > 0$ such that $E_Z^* E_W = c_\rho K(Z,W)$ for each Z, $W \in \mathcal{D}$.*
(2) For $g \in G$ and $Z \in \mathcal{D}$, we have $E_{g\cdot Z} = \pi(g) E_Z J(g,Z)^*$.

In the rest of this section, we give an explicit expression for the derived representation $d\pi$. We use the following notation. If L is a Lie group and X is an element of the Lie algebra of L then we denote by X^+ the right invariant vector field on L generated by X, that is, $X^+(h) = \frac{d}{dt}(\exp tX)h|_{t=0}$ for $h \in L$. Then, by differentiating the multiplication map from $P^+ \times K^c \times P^-$ onto $P^+K^cP^-$, we can easily prove the following result.

LEMMA 3.2. *Let $X \in \mathfrak{g}^c$ and $g = z\,k\,y$ where $z \in P^+$, $k \in K^c$ and $y \in P^-$. We have*

1. $d\zeta_g(X^+(g)) = (\mathrm{Ad}(z)\, p_{\mathfrak{p}^+}(\mathrm{Ad}(z^{-1})\, X))^+(z)$.
2. $d\kappa_g(X^+(g)) = (p_{\mathfrak{k}^c}(\mathrm{Ad}(z^{-1})\, X))^+(k)$.
3. $d\eta_g(X^+(g)) = (\mathrm{Ad}(k^{-1})\, p_{\mathfrak{p}^-}(\mathrm{Ad}(z^{-1})\, X))^+(y)$.

From this lemma, we deduce the following proposition (see also [37, p. 515]).

PROPOSITION 3.3. *For $X \in \mathfrak{g}^c$ and $f \in \mathcal{H}$, we have*

$$(d\pi(X)f)(Z) = d\rho(p_{\mathfrak{k}^c}(e^{-\operatorname{ad} Z}\, X))\, f(Z) - (df)_Z \left(\frac{\operatorname{ad} Z}{1 - e^{-\operatorname{ad} Z}} p_{\mathfrak{p}^+}(e^{-\operatorname{ad} Z}\, X) \right).$$

4. Berezin calculus

Here, we first introduce the Berezin calculus associated with ρ, see [5, 14, 45]. Let $\lambda \in (\mathfrak{h}^c)^*$ denote the highest weight of ρ relative to Δ_c^+. Let $\varphi_0 := -i\lambda \in (\mathfrak{h}^c)^*$. We also denote by φ_0 the restriction to $\mathfrak{k}$ of the trivial extension of φ_0 to $\mathfrak{k}^c$. Then the orbit $o(\varphi_0)$ of φ_0 under the coadjoint action of K is said to be associated with ρ [13, 45].

Note that a complex structure on $o(\varphi_0)$ is then defined by the diffeomorphism $o(\varphi_0) \simeq K/H \simeq K^c/H^cN^-$ where N^- is the analytic subgroup of K^c with Lie algebra $\sum_{\alpha\in\Delta_c^+} \mathfrak{g}_{-\alpha}$.

Without loss of generality, we can assume that V is a space of holomorphic functions on $o(\varphi_0)$ as in [14]. For each $\varphi \in o(\varphi_0)$ there exists a unique function $e_\varphi \in V$ (called a coherent state) such that $a(\varphi) = \langle a, e_\varphi\rangle_V$ for each $a \in V$. The Berezin calculus on $o(\varphi_0)$ associates with each operator B on V the complex-valued function $s(B)$ on $o(\varphi_0)$ defined by

$$s(B)(\varphi) = \frac{\langle Be_\varphi, e_\varphi\rangle_V}{\langle e_\varphi, e_\varphi\rangle_V}$$

which is called the symbol of B. In the following proposition, we recall some basic properties of the Berezin calculus, see for instance [5, 14, 22].

PROPOSITION 4.1. *1. The map $B \to s(B)$ is injective.*

2. For each operator B on V, we have $s(B^*) = \overline{s(B)}$.

3. For each $\varphi \in o(\varphi_0)$, $k \in K$ and $B \in \mathrm{End}(V)$, we have

$$s(B)(\mathrm{Ad}(k)\varphi) = s(\rho(k)^{-1} B \rho(k))(\varphi).$$

4. For each $U \in \mathfrak{k}$ and $\varphi \in o(\varphi_0)$, we have $s(d\rho(U))(\varphi) = i\beta(\varphi, U)$.

In order to define the Berezin symbol $S(A)$ of an operator A on $\mathcal{H}$, we first define the pre-symbol $S_0(A)$ of A as a $\mathrm{End}(V)$-valued function on $\mathcal{D}$, following [2, 17, 32].

Let $\mathcal{H}^s$ be the subspace of $\mathcal{H}$ generated by the functions $E_Z v$ for $Z \in \mathcal{D}$ and $v \in V$. Clearly, $\mathcal{H}^s$ is a dense subspace of $\mathcal{H}$. Let $\mathcal{C}$ be the space consisting of all operators A on $\mathcal{H}$ such that the domain of A contains $\mathcal{H}^s$ and the domain of A^* also contains $\mathcal{H}^s$. We define the pre-symbol $S_0(A)$ of $A \in \mathcal{C}$ by

$$S_0(A)(Z) = c_\rho^{-1}\, \rho(k_Z^{-1}) E_Z^* A E_Z \rho(k_Z^{-1})^*$$

and then the Berezin symbol $S(A)$ of A is defined as the complex-valued function on $\mathcal{D} \times o(\varphi_0)$ given by

$$S(A)(Z, \varphi) = s(S_0(A)(Z))(\varphi).$$

In order to establish that S_0 hence S are G-equivariant with respect to π, we need the following lemma.

LEMMA 4.2. *For $g \in G$ and $Z \in \mathcal{D}$, let $k(g, Z) := k_Z^{-1} \kappa(g \exp Z)^{-1} k_{g \cdot Z}$. Then we have $k(g, Z) = g_Z^{-1} g^{-1} g_{g \cdot Z}$. In particular, $k(g, Z)$ is an element of K.*

Proof. Let $g \in G$ and $Z \in \mathcal{D}$. We can write $g_Z = \exp Z k_Z y$ where $y \in P^-$. Then, on the one hand, we have

$$g g_Z = g \exp Z k_Z y = \exp(g \cdot Z) \kappa(g \exp Z) \eta(g \exp Z) k_Z y.$$

On the other hand, we can also write $g_{g \cdot Z} = \exp(g \cdot Z) k_{g \cdot Z} y'$ where $y' \in P^-$. Since $(g g_Z) \cdot 0 = g \cdot Z = g_{g \cdot Z} \cdot 0$, we see that $k := (g g_Z)^{-1} g_{g \cdot Z}$ is an element of K. Then, by replacing $g g_Z$ and $g_{g \cdot Z}$ by the above expressions we get

$$k = y^{-1} k_Z^{-1} \eta(g \exp Z)^{-1} \kappa(g \exp Z)^{-1} k_{g \cdot Z} y'.$$

Thus, applying κ, we obtain $k = k(g, Z)$ hence the result. $\square$

PROPOSITION 4.3. *1. Each $A \in \mathcal{C}$ is determined by $S_0(A)$.*

2. For each $A \in \mathcal{C}$ and each $Z \in \mathcal{D}$, we have $S_0(A^*)(Z) = S_0(A)(Z)^*$.

3. *For each $Z\in\mathcal{D}$, we have $S_0(I)(Z)=I_V$. Here I denotes the identity operator of $\mathcal{H}$ and I_V the identity operator of V.*

4. *For each $A\in\mathcal{C}$, $g\in G$ and $Z\in\mathcal{D}$, we have*

$$S_0(A)(g\cdot Z)=\rho(k(g,Z))^{-1}S_0(\pi(g)^{-1}A\pi(g))(Z)\rho(k(g,Z)).$$

Proof. The proof is similar to that of [17, Proposition 4.1]. Following [37, p. 15], we associate with any operator $A\in\mathcal{C}$ the function $K_A(Z,W):=E_Z^*AE_W$.

1. Let $A\in\mathcal{C}$. Since we have

$$\begin{aligned}\langle (Af)(Z),v\rangle_V&=\langle Af,E_Zv\rangle=\langle f,A^*E_Zv\rangle\\&=\int_{\mathcal{D}}\langle K(W,W)^{-1}f(W),(A^*E_Zv)(W)\rangle_V\,d\mu(W)\\&=\int_{\mathcal{D}}\langle K(W,W)^{-1}f(W),K_A(Z,W)^*v\rangle_V\,d\mu(W)\end{aligned}$$

we see that A is determined by K_A. Moreover, since $K_A(Z,W)$ is clearly holomorphic in the variable Z and anti-holomorphic in the variable W, we also see that K_A hence A is determined by $K_A(Z,Z)$ or, equivalently, by $S_0(A)(Z)$.

2. Clearly, for each $A\in\mathcal{C}$, Z, $W\in\mathcal{D}$, we have $K_{A^*}(Z,W)=K_A(W,Z)^*$. The result follows.

3. Let $Z\in\mathcal{D}$. We have

$$E_Z^*E_Z=c_\rho K(Z,Z)=c_\rho\rho(\kappa(\exp Z^*\exp Z))^{-1}=c_\rho\rho(k_Zk_Z^*).$$

The result therefore follows.

4. Let $A\in\mathcal{C}$, $g\in G$ and $Z\in\mathcal{D}$. We have

$$\begin{aligned}S_0(A)(g\cdot Z)&=\frac{1}{c_\rho}\,\rho(k_{g\cdot Z}^{-1})\,E_{g\cdot Z}^*AE_{g\cdot Z}\,\rho(k_{g\cdot Z}^{-1})^*\\&=\frac{1}{c_\rho}\,\rho(k_{g\cdot Z}^{-1})\,\rho(\kappa(g\exp Z))E_Z^*\pi(g)^{-1}A\pi(g)E_Z\rho(\kappa(g\exp Z))^*\rho(k_{g\cdot Z}^{-1})^*\\&=\frac{1}{c_\rho}\,\rho(k(g,Z))^{-1}\rho(k_Z^{-1})E_Z^*\pi(g)^{-1}A\pi(g)E_Z\rho(k_Z^{-1})^*\rho(k(g,Z))\\&=\rho(k(g,Z))^{-1}S_0(\pi(g)^{-1}A\pi(g))(Z)\rho(k(g,Z)).\end{aligned}$$

□

From this proposition and Proposition 4.1 we immediately deduce the following proposition.

PROPOSITION 4.4. 1. *Each $A\in\mathcal{C}$ is determined by $S(A)$.*

2. *For each $A\in\mathcal{C}$, we have $S(A^*)=\overline{S(A)}$.*

3. We have $S(I)=1$.

4. For each $A\in\mathcal{C}$, $g\in G$, $Z\in\mathcal{D}$ *and* $\varphi\in o(\varphi_0)$, *we have*

$$S(A)(g\cdot Z,\varphi)=S(\pi(g)^{-1}A\pi(g))(Z,\mathrm{Ad}(k(g,Z))\varphi).$$

5. Berezin symbols of representation operators

In this section, we give some simple formulas for the Berezin pre-symbol of $\pi(g)$ for $g\in G$ and for the Berezin symbol of $d\pi(X)$ for $X\in\mathfrak{g}^c$.

Proposition 5.1. *For* $g\in G$ *and* $Z\in\mathcal{D}$, *we have*

$$S_0(\pi(g))(Z)=\rho\left(k_Z^{-1}\kappa(\exp Z^*g^{-1}\exp Z)^{-1}(k_Z^{-1})^*\right).$$

Proof. For each $g\in G$, we have

$$\begin{aligned}S_0(\pi(g))(0)&=c_\rho^{-1}E_0^*\pi(g)E_0=c_\rho^{-1}E_0^*E_{g\cdot 0}J(g,0)^{*-1}\\&=K(0,g\cdot 0)J(g,0)^{*-1}=\rho(\kappa(g))^{*-1}=\rho(\kappa(g^{-1}))^{-1}\end{aligned}$$

by Proposition 3.1.

Now, by using G-equivariance of S_0 (see Proposition 4.3), we get

$$S_0(\pi(g))(Z)=S_0(\pi(g_Z^{-1}gg_Z))(0)=\rho(\kappa(g_Z^{-1}g^{-1}g_Z))^{-1}.$$

But writing $g_Z=\exp Zk_Zy$ with $y\in P^-$ we see that

$$g_Z^{-1}g^{-1}g_Z=g_Z^*g^{-1}g_Z=y^*k_Z^*\exp Z^*g^{-1}\exp Zk_Zy$$

hence $\kappa(g_Z^{-1}g^{-1}g_Z)=k_Z^*\kappa(\exp Z^*g^{-1}\exp Z)k_Z$. This gives the result. $\square$

Now, we aim to compute $S_0(d\pi(X))$ and $S(d\pi(X))$ for $X\in\mathfrak{g}^c$. For $\varphi\in\mathfrak{k}^*$, we also denote by φ the restriction to $\mathfrak{g}$ of the extension of φ to $\mathfrak{g}^c$ which vanishes on $\mathfrak{p}^\pm$. Then we have the following result.

Proposition 5.2. *1. For each* $g\in G$ *and* $Z\in\mathcal{D}$, *we have*

$$S_0(d\pi(X))(Z)=(d\rho\circ p_{\mathfrak{k}^c})(\mathrm{Ad}(g_Z^{-1})X).$$

2. For each $g\in G$, $Z\in\mathcal{D}$ *and* $\varphi\in o(\varphi_0)$, *we have*

$$S(d\pi(X))(Z,\varphi)=i\langle\mathrm{Ad}^*(g_Z)\varphi,X\rangle.$$

Proof. We can deduce the first statement from the preceding proposition. Indeed, by using Lemma 3.2 we get

$$\begin{aligned}&\frac{d}{dt}\,\rho(\kappa(\exp Z^*\exp(-tX)\exp Z)^{-1})|_{t=0}\\&\qquad=\rho(\kappa(\exp Z^*\exp Z)^{-1})(d\rho\circ p_{\mathfrak{k}^c})(\mathrm{Ad}(\zeta(\exp Z^*\exp Z)^{-1}\exp Z^*)X).\end{aligned}$$

Recall that we have $g_Z=\exp(-Z^*)\,\zeta(\exp Z^*\exp Z)k_Z^{-1}$. Then we obtain

$$\begin{aligned}S_0(d\pi(X))(Z)=&(d\rho\circ p_{\mathfrak{k}^c})(\mathrm{Ad}(k_Z\zeta(\exp Z^*\exp Z)^{-1}\exp Z^*)X)\\=&(d\rho\circ p_{\mathfrak{k}^c})(\mathrm{Ad}(g_Z^{-1})X).\end{aligned}$$

The second statement follows from the first and 4 of Proposition 4.1. $\square$

We are then lead to consider the map $\Psi:\mathcal{D}\times o(\varphi_0)\to\mathfrak{g}^*$ defined by $\Psi(Z,\varphi)=\mathrm{Ad}^*(g_Z)\varphi$. Note that by 4 of Proposition 4.4 and 2 of Proposition 5.2 we have

$$\Psi(g\cdot Z,\varphi)=\mathrm{Ad}^*(g)\,\Psi(Z,\mathrm{Ad}^*(k(g,Z))\varphi)\tag{1}$$

for each $g\in G$, $Z\in\mathcal{D}$ and $\varphi\in o(\varphi_0)$.

We say that $\xi_0\in\mathfrak{g}^*$ is *regular* if the stabilizer $G(\xi_0)$ of ξ_0 for the coadjoint action is connected and if the Hermitian form $(Z,W)\to\langle\xi_0,[Z,W^*]\rangle$ is not isotropic. Recall that we have denoted by $\varphi_0\in\mathfrak{g}^*$ the restriction to $\mathfrak{g}$ of the trivial extension to $\mathfrak{g}^c$ of $-i\lambda\in\mathfrak{h}^*$ where λ is the highest weight of ρ. Let $\mathcal{O}(\varphi_0)$ be the orbit of $\varphi_0\in\mathfrak{g}^*$ for the coadjoint action of G and let $K(\varphi_0)$ be the stabilizer of φ_0 for the coadjoint action of K. We assume that φ_0 is regular. Then we have the following result.

Lemma 5.3. *We have $G(\varphi_0)=K(\varphi_0)$.*

Proof. Let us denote by $\mathfrak{g}(\varphi_0)$ and $\mathfrak{k}(\varphi_0)$ the Lie algebras of $G(\varphi_0)$ and $K(\varphi_0)$. We first show that $\mathfrak{g}(\varphi_0)=\mathfrak{k}(\varphi_0)$.

Let $X\in\mathfrak{g}(\varphi_0)$. Then we have $\langle\varphi_0,[X,X']\rangle=0$ for each $X'\in\mathfrak{g}^c$. Now, we can write $X=Z+H+Y$ where $Z\in\mathfrak{p}^+$, $H\in\mathfrak{k}^c$ and $Y\in\mathfrak{p}^-$. Take $X'=Z$ in the preceding equation and recall that we have $\varphi_0|_{\mathfrak{p}^\pm}=0$ and $[\mathfrak{k}^c,\mathfrak{p}^\pm]\subset\mathfrak{p}^\pm$. Thus we get $\langle\varphi_0,[Z,Z^*]\rangle=0$ hence $Z=0$. Similarly, we obtain $Y=0$. This gives $X=H\in\mathfrak{k}(\varphi_0)$. This shows that $\mathfrak{g}(\varphi_0)=\mathfrak{k}(\varphi_0)$.

Now, $G(\varphi_0)$ is connected by hypothesis and $K(\varphi_0)$ is also connected by [35, Lemma 5]. Since $K(\varphi_0)\subset G(\varphi_0)$, we can conclude that $G(\varphi_0)=K(\varphi_0)$. $\square$

We are now in position to establish the following proposition.

Proposition 5.4. *The map Ψ is a diffeomorphism form $\mathcal{D}\times o(\varphi_0)$ onto $\mathcal{O}(\varphi_0)$.*

Proof. For each $g \in G$, one has

$$\mathrm{Ad}^*(g)\varphi_0 = \mathrm{Ad}^*(g)\Psi(0,\varphi_0) = \Psi(g\cdot 0, \mathrm{Ad}^*(k(g,0))\varphi_0).$$

This implies that Ψ takes values in $\mathcal{O}(\varphi_0)$ and that Ψ is surjective.

Now, let (Z,φ) and (Z',φ') in $\mathcal{D}\times o(\varphi_0)$ such that $\Psi(Z,\varphi) = \Psi(Z',\varphi')$. Then we have $\mathrm{Ad}^*(g_Z)\varphi = \mathrm{Ad}^*(g_{Z'})\varphi'$. Write $\varphi = \mathrm{Ad}^*(k)\varphi_0$ and $\varphi' = \mathrm{Ad}^*(k')\varphi_0$ where k, $k' \in K$. Thus we get $\mathrm{Ad}^*(g_Z k)\varphi_0 = \mathrm{Ad}^*(g_{Z'}k')\varphi_0$ and, by Lemma 5.3, there exists $k_0 \in K(\varphi_0)$ such that $g_{Z'}k' = g_Z k k_0$. Consequently, we have $Z' = (g_{Z'}k')\cdot 0 = (g_Z k k_0)\cdot 0 = Z$ hence $k' = kk_0$ and, finally, we obtain $\varphi' = \mathrm{Ad}^*(k')\varphi_0 = \mathrm{Ad}^*(kk_0)\varphi_0 = \mathrm{Ad}^*(k)\varphi_0 = \varphi$. This shows that Ψ is injective.

Now we have to show that Ψ is regular. By using Equation 1, it is sufficient to prove that Ψ is regular at $(0,\varphi)$ for $\varphi \in o(\varphi_0)$. Recall that we have

$$\Psi(Z,\varphi) = \mathrm{Ad}^*\left(\exp(-Z^*)\,\zeta(\exp Z^* \exp Z)k_Z^{-1}\right)\varphi.$$

Then, differentiating Ψ by using Lemma 3.2, we easily get

$$(d\Psi)(0,\varphi)(W, U^+(\varphi)) = \mathrm{ad}^*(W - W^* + U)\varphi$$

for each $W \in \mathfrak{p}^+$ and $U \in \mathfrak{k}^c$. Thus, for each $X \in \mathfrak{g}^c$, we have

$$\langle \varphi, [W - W^* + U, X]\rangle = 0.$$

Taking in particular $X = W^*$, we get $\langle \varphi, [W, W^*]\rangle = 0$. Since φ_0 hence φ is regular, we obtain $W = 0$ and, consequently, $\mathrm{ad}^*(U)\varphi = U^+(\varphi) = 0$. This finishes the proof. $\square$

Note that we have also the following result.

Proposition 5.5. *Assume that we have $[\mathfrak{p}^+, \mathfrak{p}^-] \subset \mathfrak{k}^c$ (this is the case, in particular, when $\mathfrak{g}$ is reductive). Let $\varphi^0 \in \mathfrak{h}^*$. As usual, we denote also by φ^0 the restriction to $\mathfrak{g}$ of the trivial extension of φ^0 to $\mathfrak{g}^*$. Then φ^0 is regular if and only if the Hermitian form $(Z,W) \to \langle \varphi^0, [Z, W^*]\rangle$ is not isotropic. In that case, we also have $G(\varphi_0) = K(\varphi_0)$.*

Proof. Assume that the Hermitian form $(Z,W) \to \langle \varphi^0, [Z, W^*]\rangle$ is not isotropic. Let $g \in G(\varphi^0)$. Write $g = (\exp Z)ky$ where $Z \in \mathfrak{p}^+$, $k \in K^c$ and $Y \in \mathfrak{p}^-$. Then we have $\mathrm{Ad}^*(k\exp Y)\varphi^0 = \mathrm{Ad}^*(\exp Z)\varphi^0$ and, for each $X \in \mathfrak{g}^c$,

$$\langle \varphi^0, \mathrm{Ad}(\exp Z)^{-1}X\rangle = \langle \varphi^0, \mathrm{Ad}(k\exp Y)^{-1}X\rangle.$$

Taking $X = Z^*$, we find $\langle \varphi^0, [Z, Z^*]\rangle = 0$ hence $Z = 0$. Similarly, we verify that $Y = 0$. This gives $g = k \in K^c \cap G(\varphi^0) = K(\varphi^0)$. Consequently, $G(\varphi^0)$ is connected and φ^0 is regular. $\square$

Moreover, by adapting the arguments of the proof of [19, Lemma 3.1], we also obtain the following proposition.

PROPOSITION 5.6. *Assume that $\mathcal{H} \neq (0)$. Then the Hermitian form $(Z, W) \to \langle \varphi^0, [Z, W^*] \rangle$ is not isotropic.*

6. Berezin transform and Stratonovich-Weyl correspondence

In this section, we introduce the Berezin transform and we review some of its properties. As an application, we construct a Stratonovich-Weyl correspondence for $(G, \pi, \mathcal{O}(\varphi_0))$.

Let us fix a K-invariant measure ν on $o(\varphi_0)$ normalized as in [14, Section 2]. Then the measure $\tilde{\mu} := \mu \otimes \nu$ on $\mathcal{D} \times o(\varphi_0)$ is invariant under the action of G on $\mathcal{D} \times o(\varphi_0)$ given by $g \cdot (Z, \varphi) := (g \cdot Z, \mathrm{Ad}(k(g, Z))^{-1}\varphi)$ and the measure $\mu_{\mathcal{O}(\varphi_0)} := (\Psi^{-1})^*(\tilde{\mu})$ is a G-invariant measure on $\mathcal{O}(\varphi_0)$.

We denote by $\mathcal{L}_2(\mathcal{H})$ (respectively $\mathcal{L}_2(V)$) the space of Hilbert-Schmidt operators on $\mathcal{H}$ (respectively V) endowed with the Hilbert-Schmidt norm. Since V is finite-dimensional, we have $\mathcal{L}_2(V) = \mathrm{End}(V)$. We denote by $L^2(\mathcal{D} \times o(\varphi_0))$ (respectively $L^2(\mathcal{D}), L^2(o(\varphi_0))$) the space of functions on $\mathcal{D} \times o(\varphi_0)$ (resp. $\mathcal{D}$, $o(\varphi_0)$) which are square-integrable with respect to the measure $\tilde{\mu}$ (respectively μ, ν). The following result is well-known, see for instance [15].

PROPOSITION 6.1. *For each $\varphi \in o(\varphi_0)$, let p_φ denote the orthogonal projection of V on the line generated by e_φ. Then the adjoint s^* of the operator $s : \mathcal{L}_2(V) \to L^2(o(\varphi_0))$ is given by*

$$s^*(a) = \int_{o(\varphi_0)} a(\varphi) p_\varphi \, d\nu(\varphi)$$

and the Berezin transform $b := ss^*$ is given by

$$b(a)(\psi) = \int_{o(\varphi_0)} a(\varphi) \frac{|\langle e_\psi, e_\varphi \rangle_V|^2}{\langle e_\varphi, e_\varphi \rangle_V \langle e_\psi, e_\psi \rangle_V} \, d\nu(\varphi)$$

for each $a \in L^2(o(\varphi_0))$

Following [18], we can easily obtain the following analogous results for S, see also [43].

PROPOSITION 6.2. *The map S is a bounded operator from $\mathcal{L}_2(\mathcal{H})$ to $L^2(\mathcal{D} \times o(\varphi_0))$. Moreover, S^* is given by*

$$S^*(f) = \int_{\mathcal{D} \times o(\varphi_0)} P_{Z,\varphi} f(Z, \varphi) d\mu(Z) d\nu(\varphi)$$

where $P_{Z,\varphi} := c_\rho^{-1} E_Z \rho(h_Z^{-1})^* p_\varphi \rho(h_Z^{-1}) E_Z^*$ *is the orthogonal projection of* $\mathcal{H}$ *on the line generated by* $E_Z\rho(h_Z^{-1})^* e_\varphi$.

From this result we easily deduce that the following proposition.

Proposition 6.3. *The Berezin transform* $B := SS^*$ *is a bounded operator of* $L^2(\mathcal{D}\times o(\varphi_0))$ *and, for each* $f \in L^2(\mathcal{D}\times o(\varphi_0))$*, we have the following integral formula*

$$B(f)(Z,\psi) = \int_{\mathcal{D}\times o(\varphi_0)} k(Z,W,\psi,\varphi)\, f(W,\varphi)\, d\mu(W) d\nu(\varphi)$$

where

$$k(Z,W,\psi,\varphi) := \frac{|\langle \rho(\kappa(g_Z^{-1} g_W))^{-1} e_\psi, e_\varphi\rangle_V|^2}{\langle e_\varphi, e_\varphi\rangle_V \langle e_\psi, e_\psi\rangle_V}.$$

Let us introduce the left-regular representation τ of G on $L^2(\mathcal{D}\times o(\varphi_0))$ defined by $(\tau(g)(f))(Z,\varphi) = f(g^{-1}\cdot(Z,\varphi))$. Clearly, τ is unitary. Moreover, since S is G-equivariant, we immediately verify that for each $f \in L^2(\mathcal{D}\times o(\varphi_0))$ and each $g\in G$, we have $B(\tau(g)f) = \tau(g)(B(f))$.

Now, we consider the polar decomposition of $S : \mathcal{L}_2(\mathcal{H}) \to L^2(\mathcal{D}\times o(\varphi_0))$. We can write $S = (SS^*)^{1/2} W = B^{1/2} W$ where $W := B^{-1/2} S$ is a unitary operator from $\mathcal{L}_2(\mathcal{H})$ to $L^2(\mathcal{D}\times o(\varphi_0))$. Then we have the following proposition.

Proposition 6.4.

1. *The map* $W : \mathcal{L}_2(\mathcal{H}) \to L^2(\mathcal{D}\times o(\varphi_0))$ *is a Stratonovich-Weyl correspondence for the triple* $(G,\pi,\mathcal{D}\times o(\varphi_0))$.
2. *The map* $\mathcal{W}$ *from* $\mathcal{L}_2(\mathcal{H})$ *to* $L^2(\mathcal{O}(\varphi_0), \mu_{\mathcal{O}(\varphi_0)})$ *defined by* $\mathcal{W}(f) = W(f\circ \Psi)$ *is a Stratonovich-Weyl correspondence for the triple* $(G,\pi,\mathcal{O}(\varphi_0))$.

7. Generalies on Heisenberg motion groups

We first introduce the Heisenberg group. For $z, w \in \mathbb{C}^n$, we denote $zw := \sum_{k=1}^n z_k w_k$. Consider the symplectic form ω on $\mathbb{C}^{2n}$ defined by

$$\omega((z,w),(z',w')) = \frac{i}{2}(zw' - z'w).$$

for $z, w, z', w' \in \mathbb{C}^n$. The $(2n+1)$-dimensional real Heisenberg group is $H_n := \{((z,\bar z), c) : z \in \mathbb{C}^n, c\in\mathbb{R}\}$ endowed with the multiplication

$$((z,\bar z),c)\cdot((z',\bar z'),c') = ((z+z',\bar z+\bar z'), c+c'+\tfrac{1}{2}\omega((z,\bar z),(z',\bar z'))). \tag{2}$$

Then the complexification H_n^c of H_n is $H_n^c := \{((z,w),c) : z,w\in\mathbb{C}^n, c\in\mathbb{C}\}$ and the multiplication of H_n^c is obtained by replacing $(z,\bar z)$ by (z,w) and $(z',\bar z')$

by (z',w') in Equation 2. We denote by $\mathfrak{h}_n$ and $\mathfrak{h}_n^c$ the Lie algebras of H_n and H_n^c.

Let K_0 be a closed subgroup of $U(n)$. Then K_0 acts on H_n by $k\cdot((z,\bar{z}),c)=((kz,\bar{kz}),c)$ and we can form the semi-direct product $G:=H_n\rtimes K_0$ which is called a Heisenberg motion group. The elements of G can be written as $((z,\bar{z}),c,h)$ where $z\in\mathbb{C}^n$, $c\in\mathbb{R}$ and $h\in K_0$. The multiplication of G is then given by

$$\begin{aligned}&((z,\bar{z}),c,h)\cdot((z',\bar{z}'),c',h')\\&\qquad=((z,\bar{z})+(hz',\bar{hz'}),c+c'+\tfrac{1}{2}\omega((z,\bar{z}),(hz',\bar{hz'})),hh').\end{aligned}$$

We denote by K_0^c the complexification of K_0. In order to describe the complexification G^c of G, it is convenient to introduce the action of K_0^c on $\mathbb{C}^n\times\mathbb{C}^n$ given by $k\cdot(z,w)=(kz,(k^t)^{-1}w)$ (here, the subscript t denotes transposition). The group G^c is then the semi-direct product $G^c=H_n^c\rtimes K_0^c$. The elements of G^c can be written as $((z,w),c,h)$ where z, $w\in\mathbb{C}^n$, $c\in\mathbb{C}$ and $h\in K_0^c$ and the multiplication law of G^c is given by

$$\begin{aligned}&((z,w),c,h)\cdot((z',w'),c',h')\\&\qquad=((z,w)+h\cdot(z',w'),c+c'+\tfrac{1}{2}\omega((z,w),h\cdot(z',w')),hh').\end{aligned}$$

We denote by $\mathfrak{k}_0$, $\mathfrak{k}_0^c$, $\mathfrak{g}$ and $\mathfrak{g}^c$ the Lie algebras of K_0, K_0^c, G and G^c. The derived action $\mathfrak{k}_0^c$ on $\mathbb{C}^n\times\mathbb{C}^n$ is $A\cdot(z,w):=(Az,-A^tw)$ and the Lie brackets of $\mathfrak{g}^c$ are given by

$$\begin{aligned}&[((z,w),c,A),((z',w'),c',A')]\\&\qquad=(A\cdot(z',w')-A'\cdot(z,w),\omega((z,w),(z',w')),[A,A']).\end{aligned}$$

Recall that, for each $X\in\mathfrak{g}^c$, we have $X^*=-\theta(X)$ where θ denotes conjugation over $\mathfrak{g}$. We can easily verify that if $X=((z,w),c,A)\in\mathfrak{g}^c$ then $X^*=((-\bar{w},-\bar{z}),c,\bar{A}^t)$.

Here we take $K=\{((0,0),c,h)\,:\,c\in\mathbb{R},\,h\in K_0\}$ for the maximal compactly embedded subgroup of G. Also, let $\mathfrak{h}_0$ be a Cartan subalgebra of $\mathfrak{k}_0$. Then we take $\mathfrak{h}:=\{((0,0),c,A)\,:\,c\in\mathbb{R},\,A\in\mathfrak{h}_0\}$ for the compactly embedded Cartan subalgebra of $\mathfrak{g}$, see Section 2. Moreover, we can choose the positive non-compact roots in such a way that $P^+=\{((z,0),0,I_n)\,:\,z\in\mathbb{C}^n\}$ and $P^-=\{((0,w),0,I_n)\,:\,w\in\mathbb{C}^n\}$. The $P^+K^cP^-$-decomposition of $g=((z_0,w_0),c_0,h)\in G^c$ is given by

$$g=((z_0,0),0,I_n)\cdot((0,0),c,h)\cdot((0,w_0),0,I_n)$$

where $c=c_0-\frac{i}{4}z_0w_0$. From this, we deduce that the action of the element $g=((z_0,w_0),c_0,h)$ of G on $Z=((z,0),0,0)\in\mathfrak{p}^+$ is given by $g\cdot Z=\log\zeta(g\exp Z)=((z_0+hz,0),0,0)$. Then we have $\mathcal{D}=\mathfrak{p}^+\simeq\mathbb{C}^n$.

We can also easily compute the section $Z \to g_Z$. We find that if $Z = ((z,0),0,0) \in \mathcal{D}$ then $g_Z = ((z,\bar{z}),0,I_n)$ and $k_Z = \kappa(g_Z) = ((0,0),-\frac{i}{4}|z|^2,I_n)$.

Now we compute the adjoint action of G^c. Let $g = (v_0,c_0,h_0) \in G^c$ where $v_0 \in \mathbb{C}^{2n}$, $c_0 \in \mathbb{C}$, $h_0 \in K_0^c$ and $X = (w,c,A) \in \mathfrak{g}^c$ where $w \in \mathbb{C}^{2n}$, $c \in \mathbb{C}$ and $A \in \mathfrak{k}_0^c$. We set $\exp(tX) = (w(t),c(t),\exp(tA))$. Then, since the derivatives of $w(t)$ and $c(t)$ at $t=0$ are w and c, we find that

$$\begin{aligned}\operatorname{Ad}(g)X &= \frac{d}{dt}(g\exp(tX)g^{-1})|_{t=0}\\ &= \big(h_0w - (\operatorname{Ad}(h_0)A)\cdot v_0, c + \omega(v_0,h_0w) - \tfrac{1}{2}\omega(v_0,(\operatorname{Ad}(h_0)A)\cdot v_0), \operatorname{Ad}(h_0)A\big).\end{aligned}$$

From this, we deduce the coadjoint action of G^c. Let us denote by $\xi = (u,d,\phi)$, where $u \in \mathbb{C}^{2n}$, $d \in \mathbb{C}$ and $\phi \in (\mathfrak{k}_0^c)^*$, the element of $(\mathfrak{g}^c)^*$ defined by

$$\langle \xi, (w,c,A)\rangle = \omega(u,w) + dc + \langle \phi, A\rangle.$$

Also, for $u, v \in \mathbb{C}^{2n}$, we denote by $v \times u$ the element of $(\mathfrak{k}_0^c)^*$ defined by $\langle v\times u, A\rangle := \omega(u, A\cdot v)$ for $A \in \mathfrak{k}_0^c$.

Now, let $\xi = (u,d,\phi) \in (\mathfrak{g}^c)^*$ and $g = (v_0,c_0,h_0) \in G^c$. Recall that we have $\langle \operatorname{Ad}^*(g)\xi, X\rangle = \langle \xi, \operatorname{Ad}(g^{-1})X\rangle$ for each $X \in \mathfrak{g}^c$. Then we obtain

$$\operatorname{Ad}^*(g)\xi = \big(h_0u - dv_0, d, \operatorname{Ad}^*(h_0)\phi + v_0 \times (h_0u - \tfrac{d}{2}v_0)\big)$$

By restriction, we also get the analogous formula for the coadjoint action of G. From this, we deduce that if a coadjoint orbit of G contains a point (u,d,ϕ) with $d \neq 0$ then it also contains a point of the form $(0,d,\phi_0)$. Such an orbit is called *generic*.

8. Representations of Heisenberg motion groups

We retain the notation of the previous section and introduce some additional notation. Let ρ_0 be a unitary irreducible representation of K_0 on a (finite-dimensional) Hilbert space V and let $\gamma \in \mathbb{R}$. Then we take ρ to be the representation of K on V defined by $\rho((0,0),c,h) = e^{i\gamma c}\rho_0(h)$ for each $c \in \mathbb{R}$ and $h \in K_0$. Thus, for each $Z = ((z,0),0,0)$, $W = ((w,0),0,0) \in \mathcal{D}$, we have $K(Z,W) = \rho(\kappa(\exp W^* \exp Z))^{-1} = e^{\gamma z\bar{w}/2} I_V$. Hence the Hilbert product on $\mathcal{H}$ is given by

$$\langle f,g\rangle = \int_{\mathcal{D}} \langle f(Z), g(Z)\rangle_V \, e^{-\gamma|z|^2/2}\, d\mu(Z)$$

where μ is the G-invariant measure on $\mathcal{D} \simeq \mathbb{C}^n$ defined by $d\mu(Z) := (\frac{\gamma}{2\pi})^n\, dx\, dy$. Here $Z = ((z,0),0,0)$ and $z = x+iy$ with x and y in $\mathbb{R}^n$. Note that we have $c_\rho = 1$. Moreover, for each $v \in V$, $Z = ((z,0),0,0)$, $W = ((w,0),0,0) \in \mathcal{D}$, we have $(E_W v)(Z) = K(Z,W)v = e^{\frac{\gamma}{2}z\bar{w}}v$.

On the other hand, we easily verify that, for each $g=((z_0,\bar z_0),c_0,h)\in G$ and $Z=((z,0),0,0),\in\mathcal{D}$, we have

$$J(g,Z)=\rho\left(\kappa(g\exp Z)\right)=\exp\left(i\gamma c_0+\tfrac{\gamma}{2}\bar z_0(hz)+\tfrac{\gamma}{4}|z_0|^2\right)\,\rho_0(h)$$

and consequently, we get the following formula for π:

$$(\pi(g)f)(Z)=\exp\left(i\gamma c_0+\tfrac{\gamma}{2}\bar z_0 z-\tfrac{\gamma}{4}|z_0|^2\right)\,\rho_0(h)\,f(h^{-1}(z-z_0),0,0)$$

where $g=((z_0,\bar z_0),c_0,h)\in G$ and $Z=((z,0),0,0),\in\mathcal{D}$.

Let $\phi_0\in\mathfrak{k}_0^*$. Assume that the orbit $o(\phi_0)$ of ϕ_0 for the coadjoint action of K_0 is associated with ρ_0 as in Section 4. Then, in the notation of Section 4, the coadjoint orbit of $\varphi_0:=((0,0),\gamma,\phi_0)$ for the coadjoint action of G is then associated with π. Note that the orbit $o(\varphi_0)$ of $\varphi_0:=((0,0),\gamma,\phi_0)$ for the coadjoint action of K can be identify to $o(\phi_0)$ via $\phi\to((0,0),\gamma,\phi)$.

In the present situation, Proposition 3.3 can be reformulated as follows.

Proposition 8.1. *Let $X=((a,b),c,A)\in\mathfrak{g}^c$. Then, for each $f\in\mathcal{H}$ and each $Z=((z,0),0,0),\in\mathcal{D}$, we have*

$$(d\pi(X)f)(Z)=d\rho_0(A)f(Z)+\gamma(ic-\tfrac{1}{2}bz)f(Z)-df_Z((a+Az,0),0,0).$$

Now consider the Hilbert space $\mathcal{H}_0$ of all holomorphic functions $f_0:\mathcal{D}\to\mathbb{C}$ such that

$$\|f\|_0^2=\int_{\mathcal{D}}|f(Z)|^2\,e^{-\gamma|z|^2/2}\,d\mu(Z)<+\infty.$$

Then for each $Z\in\mathcal{D}$ there exists a coherent state $e_Z^0\in\mathcal{H}_0$ such that $f(Z)=\langle f,e_Z^0\rangle_0$ for each $f\in\mathcal{H}_0$. More precisely, for each $Z=((z,0),0,0)$, $W=((w,0),0,0)\in\mathcal{D}$, we have $e_Z^0(W)=e^{\gamma\bar z w/2}$.

Clearly, one has $\mathcal{H}=\mathcal{H}_0\otimes V$. For $f_0\in\mathcal{H}_0$ and $v\in V$, we denote by $f_0\otimes v$ the function $Z\to f_0(Z)v$. Moreover, if A_0 is an operator of $\mathcal{H}_0$ and A_1 is an operator of V then we denote by $A_0\otimes A_1$ the operator of $\mathcal{H}$ defined by $(A_0\otimes A_1)(f_0\otimes v)=A_0f_0\otimes A_1v$.

Let π_0 be the unitary irreducible representation of H_n on $\mathcal{H}_0$ defined by

$$(\pi_0((z_0,\bar z_0),c_0)f_0)(Z)=\exp\left(i\gamma c_0+\tfrac{\gamma}{2}\bar z_0 z-\tfrac{\gamma}{2}|z_0|^2\right)\,f_0((z-z_0,0),0,0)$$

for each $Z=((z,0),0,0)\in\mathcal{D}$ and let σ_0 be the left-regular representation of K_0 on $\mathcal{H}_0$, that is, $(\sigma_0(h)f_0)(Z)=f_0(h^{-1}\cdot Z)$. Then we have

$$\pi((z_0,\bar z_0),c_0,h)=\pi_0((z_0,\bar z_0),c_0)\circ\sigma_0(h)\otimes\rho_0(h)$$

for each $z_0\in\mathbb{C}^n$, $c_0\in\mathbb{R}$ and $h\in K_0$. This is precisely Formula (3.18) in [7].

9. Berezin and Stratonovich-Weyl symbols for Heisenberg motion groups

In this section, we first establish a decomposition formula for the Berezin symbol of an operator on $\mathcal{H}$ of the form $A_0 \otimes A_1$ where A_0 is an operator of $\mathcal{H}_0$ and A_1 is an operator of V. As an application, we compute explicitely the Berezin and the Stratonovich-Weyl symbols of the representation operators.

We also need here the Berezin calculus for operators on $\mathcal{H}_0$. Recall that the Berezin symbol $S^0(A_0)$ of an operator A_0 on $\mathcal{H}_0$ is the function on $\mathcal{D}$ defined by

$$S^0(A_0)(Z) := \frac{\langle A_0\, e^0_Z\, ,\, e^0_Z\rangle}{\langle e^0_Z\, ,\, e^0_Z\rangle} = e^{-\gamma|z|^2/2}\,(A_0 e^0_Z)(Z),$$

see, for instance, [12]. In particular, S^0 is H_n-equivariant with respect to π_0. Let $B^0 := S^0(S^0)^*$ be the corresponding Berezin transform.

On the other hand, recall that $\varphi_0 = ((0,0),\gamma,\phi_0)$ and that we have identified the coadjoint orbit $o(\varphi_0)$ of K with the coadjoint orbit $o(\phi_0)$ of K_0. Then, for $\varphi = ((0,0),\gamma,\phi)$, we can identify the coherent state e_φ on $o(\varphi_0)$ with the coherent state e_ϕ on $o(\phi_0)$. Hence, the corresponding Berezin calculus can be also identified.

Let f_0 be a complex-valued function on $\mathcal{D}$ and f_1 be a complex-valued function on $o(\phi_0)$. Then we denote by $f_0 \otimes f_1$ the function on $\mathcal{D} \times o(\phi_0)$ defined by $f_0 \otimes f_1(Z,\phi) = f_0(Z) f_1(\phi)$.

Proposition 9.1. *Let A_0 be an operator on $\mathcal{H}_0$ and let A_1 be an operator on V. Let $A := A_0 \otimes A_1$. Then*

1. *For each $Z \in \mathcal{D}$, we have $S_0(A)(Z) = S^0(A_0)(Z)A_1$.*
2. *For each $Z \in \mathcal{D}$ and each $\phi \in o(\phi_0)$, we have $S(A)(Z,\phi) = S^0(A_0)(Z)s(A_1)(\phi)$, that is, $S(A) = S^0(A_0) \otimes s(A_1)$.*

Proof. Let $Z = ((z,0),0,0) \in \mathcal{D}$ and $v \in V$. We have

$$S_0(A)(Z)v = e^{-\gamma|z|^2/2} E_Z^* A E_Z v = e^{-\gamma|z|^2/2} A(E_Z v)(Z).$$

Now, recall that $E_Z v = e^0_Z \otimes v$. Then we get $A(E_Z v) = A_0 e^0_Z \otimes A_1 v$ and, consequently,

$$S_0(A)(Z)v = e^{-\gamma|z|^2/2}(A_0 e^0_Z)(Z) A_1 v = S^0(A_0)(Z)A_1.$$

This proves 1. Assertion 2 immediately follows from 1. □

The preceding proposition is useful to compute the Berezin symbol of an operator on $\mathcal{H}$ which is a sum of operators of the form $A_0 \otimes A_1$. This is precisely the case of the representation operators $\pi(g)$, $g \in G$ and $d\pi(X)$, $X \in \mathfrak{g}^c$ and then we have the following propositions.

PROPOSITION 9.2. *Let $g = ((z_0, \bar{z}_0), c_0, h) \in G$. For each $Z = ((z, 0), 0, 0) \in \mathcal{D}$ and each $\phi \in o(\phi_0)$, we have*

$$S(\pi(g))(Z, \phi) \\ = \exp \gamma \left(ic_0 + \tfrac{1}{2}\bar{z}_0 z - \tfrac{1}{4}|z_0|^2 - \tfrac{1}{2}|z|^2 + \tfrac{1}{2}\bar{z}h^{-1}(z - z_0) \right) s(\rho_0(h))(\phi).$$

Proof. Recall that, for each $g = ((z_0, \bar{z}_0), c_0, h) \in G$, we have

$$\pi(g) = \pi_0((z_0, \bar{z}_0), c_0) \circ \sigma(h) \otimes \rho_0(h).$$

Then the result follows from Proposition 9.1. □

PROPOSITION 9.3. *1. For each $X = ((a, b), c, A) \in \mathfrak{g}^c$, $Z = ((z, 0), 0, 0) \in \mathcal{D}$ and $\phi \in o(\phi_0)$, we have*

$$S(d\pi(X))(Z, \phi) = i\gamma c - \frac{\gamma}{2} \left(a\bar{z} + bz + \bar{z}(Az) \right) + s(d\rho_0(A))(\phi).$$

2. For each $X = ((a, b), c, A) \in \mathfrak{g}^c$ and $Z = ((z, 0), 0, 0) \in \mathcal{D}$ and $\phi \in o(\phi_0)$, we have $S(d\pi(X))(Z, \phi) = i\langle \Psi(Z, \phi), X \rangle$ where the diffeomorphism $\Psi : \mathcal{D} \times o(\phi_0) \to \mathcal{O}(\varphi_0)$ is defined by

$$\Psi(Z, \phi) = \left(-\gamma(z, \bar{z}), \gamma, \phi - \tfrac{\gamma}{2}(z, \bar{z}) \;\times (z, \bar{z}) \right).$$

Proof. Assertion 1 follows from Proposition 3.3 and Proposition 9.1 and Assertion 2 follows from the equality $\Psi(Z, \phi) = \mathrm{Ad}^*(g_Z)\varphi_0$. □

By adapting Proposition 6.3 to the present situation, we get the following decomposition of the Berezin transform $B = SS^*$.

PROPOSITION 9.4. *For each $f \in L^2(\mathcal{D} \times o(\phi_0))$, we have*

$$B(f)(Z, \psi) = \int_{\mathcal{D} \times o(\phi_0)} k(Z, W, \psi, \phi)\, f(W, \phi)\, d\mu(W) d\nu(\phi)$$

where

$$k(Z, W, \psi, \phi) = e^{-\gamma|z-w|^2/2} \, \frac{|\langle e_\psi, e_\phi \rangle_V|^2}{\langle e_\phi, e_\phi \rangle_V \langle e_\psi, e_\psi \rangle_V}.$$

In particular, for each $f_0 \in L^2(\mathcal{D})$ and $f_1 \in L^2(o(\phi_0))$, we have $B(f_0 \otimes f_1) = B_0(f_0) \otimes b(f_1)$.

Proof. We can compute $k(Z, W, \psi, \phi)$ (see Proposition 6.3) as follows. Let $Z = ((z, 0), 0, 0)$ and $W = ((w, 0), 0, 0) \in \mathcal{D}$. Then we have

$$g_Z^{-1} g_W = \left((-z + w, -\bar{z} + \bar{w}), -\frac{i}{4}(z\bar{w} - \bar{z}w), I_n \right).$$

Thus
$$\kappa(g_Z^{-1}g_W)=\left((0,0),-\frac{i}{4}(z\bar{z}+w\bar{w}-2\bar{z}w),I_n\right).$$
Consequently, we get
$$\rho(\kappa(g_Z^{-1}g_W))^{-1}=e^{-\gamma(|z|^2+|w|^2-2\bar{z}w)/4}\,I_V.$$
Since we have
$$|e^{-\gamma(|z|^2+|w|^2-2\bar{z}w)/4}|^2=e^{-\gamma|z-w|^2/2},$$
the first assertion follows from Proposition 6.3. The second assertion is an immediate consequence of the first one. □

In the following proposition, we study the form of the function
$$S(d\pi(X_1X_2\cdots X_q))$$
for $X_1,X_2,\ldots,X_q\in\mathfrak{g}^c$.

PROPOSITION 9.5. *Let $X_1,X_2,\ldots,X_q\in\mathfrak{g}^c$. Then*

1. *The function $S(d\pi(X_1X_2\cdots X_q))(Z,\phi)$ is a sum of terms of the form*
$$P(Z)Q(\bar{Z})s(d\rho_0(Y_1Y_2\cdots Y_r))(\phi)$$
where P, Q are polynomials of degree $\leq q$, $r\leq q$ and $Y_1,Y_2,\ldots,Y_r\in\mathfrak{k}_0^c$.
2. *We have $S(d\pi(X_1X_2\cdots X_q))\in L^2(\mathcal{D}\times o(\phi_0))$.*

Proof. 1. By using Proposition 3.3, we can verify by induction on q that, for each $X_1,X_2,\ldots,X_q\in\mathfrak{g}^c$, $d\pi(X_1X_2\cdots X_q)$ is a sum of terms of the form
$$P(Z)d\rho_0(Y_1Y_2\cdots Y_r)\partial_{i_1}\partial_{i_2}\cdots\partial_{i_s}$$
where P is a polynomial of degree $\leq q$, r, $s\leq q$ and $Y_1,Y_2,\ldots,Y_r\in\mathfrak{k}_0^c$. Here we write as usual $Z=((z,0),0,0)$ with $z\in\mathbb{C}^n$ and ∂_i stands for the derivative with respect to z_i.

Taking Proposition 9.1 into account, this implies that $S(d\pi(X_1X_2\cdots X_q))$ is a sum of terms of the form
$$P(Z)S^0(\partial_{i_1}\partial_{i_2}\cdots\partial_{i_s})(Z)\,s(d\rho_0(Y_1Y_2\cdots Y_r))(\phi).$$

But recall that $e_Z^0(W)=e^{\gamma\bar{z}w/2}$. Then we have
$$(\partial_{i_1}\partial_{i_2}\cdots\partial_{i_s}e_Z^0)(W)=\bar{w}_{i_1}\bar{w}_{i_1}\cdots\bar{w}_{i_s}e_Z^0(W).$$
Thus we see that
$$S^0(\partial_{i_1}\partial_{i_2}\cdots\partial_{i_s})(Z)=e_Z^0(Z)^{-1}(\partial_{i_1}\partial_{i_2}\cdots\partial_{i_s}e_Z^0)(Z)=\bar{w}_{i_1}\bar{w}_{i_1}\cdots\bar{w}_{i_s}.$$

The result follows.

2. This assertion is a consequence of 1. Indeed, the function $P(Z)Q(\bar{Z})$ with P, Q polynomials is clearly square-integrable with respect to μ_0. On the other hand, recall that V is finite-dimensional, that $o(\phi_0)$ is compact and that we have the property $|s(A_0)| \leq \|A_0\|_{\mathrm{op}}$ for each operator A_0 on V. Then we see that $s(d\rho_0(Y_1 Y_2 \cdots Y_s))$ is bounded hence square-integrable on $o(\phi_0)$. $\square$

In the general case, by contrast to the preceding proposition, the function $S(d\pi(X_1 X_2 \cdots X_q))$ is not usually square-integrable. However, when $\mathfrak{g}$ is reductive, we have proved that B can be extended to a class of fonctions which contains $S(d\pi(X_1 X_2 \cdots X_q))$ for $X_1, X_2, \ldots, X_q \in \mathfrak{g}^c$ and $q \leq q_\pi$ where q_π only depends on π, see [18, 19].

Finally, we compute $W(d\pi(X))$, $X \in \mathfrak{g}^c$ which is well-defined thanks to the preceding proposition. Consider the Stratonovich-Weyl correspondences $W := B^{-1/2}S$, $W_0 := B_0^{-1/2}S^0$ and $w := b^{-1/2}s$ on $\mathcal{D} \times o(\phi_0)$, $\mathcal{D}$ and $o(\phi_0)$, respectively. Clearly, for any A_0 operator on $\mathcal{H}_0$ and any A_1 operator on V, we have $W(A_0 \otimes A_1) = W_0(A_0) \otimes w(A_1)$ by Proposition 9.1 and Proposition 9.4.

Proposition 9.6. *For each $X = ((a,b),c,A) \in \mathfrak{g}^c$, $Z = ((z,0),0,0) \in \mathcal{D}$ and $\phi \in o(\phi_0)$, we have*

$$W(d\pi(X))(Z,\phi) = ic\gamma + w(d\rho_0(A))(\phi) + \frac{1}{2}\operatorname{Tr}(A) - \frac{\gamma}{2}\left(a\bar{z} + bz + \bar{z}(Az)\right).$$

Proof. Let $\Delta := 4\sum_{k=1}^n (\partial_{z_k}\partial_{\bar{z}_k})$ be the Laplace operator. Then it is well-known that we have $B_0 = \exp(\Delta/2\gamma)$, see [36]. Thus we get

$$W_0 = \exp(-\Delta/4\gamma)S^0$$

and, by applying Proposition 9.3 and Proposition 9.4, we find that

$$\begin{aligned} W(d\pi(X))(Z,\phi) &= ic\gamma + w(d\rho_0(A))(\phi) - \frac{\gamma}{2}\exp(-\Delta/4\gamma)\left(a\bar{z} + bz + \bar{z}(Az)\right) \\ &= ic\gamma + w(d\rho_0(A))(\phi) + \frac{1}{2}\operatorname{Tr}(A) - \frac{\gamma}{2}\left(a\bar{z} + bz + \bar{z}(Az)\right). \end{aligned}$$

$\square$

References

[1] S. T. Ali and M. Englis, *Quantization methods: a guide for physicists and analysts*, Rev. Math. Phys. **17** (2005), 391–490.

[2] S. T. Ali and M. Englis, *Berezin-Toeplitz quantization over matrix domains*, arXiv:math-ph/0602015v1.

[3] J. Arazy and H. Upmeier, *Weyl Calculus for Complex and Real Symmetric Domains*, Harmonic analysis on complex homogeneous domains and Lie groups (Rome, 2001). Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl. 13, no 3-4 (2002), pp. 165–181.

[4] J. Arazy and H. Upmeier, *Invariant symbolic calculi and eigenvalues of invariant operators on symmeric domains*, Function spaces, interpolation theory and related topics (Lund, 2000) pp. 151–211, de Gruyter, Berlin, 2002.

[5] D. Arnal , M. Cahen and S. Gutt, *Representations of compact Lie groups and quantization by deformation*, Acad. R. Belg. Bull. Cl. Sc. 3e série LXXIV, **45** (1988), 123–141.

[6] D. Arnal, J.-C. Cortet, *Nilpotent Fourier Transform and Applications*, Lett. Math. Phys. **9** (1985), 25–34.

[7] C. Benson, J. Jenkins, R. L. Lipsmann and G. Ratcliff, *A geometric criterion for Gelfand pairs associated with the Heisenberg group*, Pacific J. Math. **178** (1997), 1–36.

[8] C. Benson, J. Jenkins and G. Ratcliff, *The orbit method and Gelfand pairs associated with nilpotent Lie groups*, J. Geom. Anal. **9** (1999), 569–582.

[9] F. A. Berezin, *Quantization*, Math. USSR Izv. **8** (1974), 1109–1165.

[10] F. A. Berezin, *Quantization in complex symmetric domains*, Math. USSR Izv. **9** (1975), 341–379.

[11] C. Brif and A. Mann, *Phase-space formulation of quantum mechanics and quantum-state reconstruction for physical systems with Lie-group symmetries*, Phys. Rev. A **59** (1999), 971–987.

[12] B. Cahen, *Contractions of $SU(1,n)$ and $SU(n+1)$ via Berezin quantization*, J. Anal. Math. **97** (2005), 83–102.

[13] B. Cahen, *Weyl quantization for semidirect products*, Differential Geom. Appl. **25** (2007), 177–190.

[14] B. Cahen, *Berezin quantization on generalized flag manifolds*, Math. Scand. **105** (2009), 66–84.

[15] B. Cahen, *Stratonovich-Weyl correspondence for compact semisimple Lie groups*, Rend. Circ. Mat. Palermo **59** (2010), 331–354.

[16] B. Cahen, *Stratonovich-Weyl correspondence for discrete series representations*, Arch. Math. (Brno) **47** (2011), 41–58.

[17] B. Cahen, *Berezin quantization for holomorphic discrete series representations: the non-scalar case*, Beiträge Algebra Geom. **53** (2012), 461–471.

[18] B. Cahen, *Berezin transform for non-scalar holomorphic discrete series*, Comment. Math. Univ. Carolin. **53** (2012), 1–17.

[19] B. Cahen, *Berezin Quantization and Holomorphic Representations*, Rend. Sem. Mat. Univ. Padova **129** (2013), 277–297.

[20] B. Cahen, *Global Parametrization of Scalar Holomorphic Coadjoint Orbits of a Quasi-Hermitian Lie Group*, Acta Univ. Palacki. Olomuc., Fac. rer. nat., Mathematica **52** (2013), 35–48.

[21] B. Cahen, *Stratonovich-Weyl correspondence for the Jacobi group*, Comm. Math. **22** (2014), 31–48.

[22] M. Cahen, S. Gutt and J. Rawnsley, *Quantization on Kähler manifolds I, Geometric interpretation of Berezin quantization*, J. Geom. Phys. **7** (1990), 45–62.
[23] J. F. Cariñena, J. M. Gracia-Bondìa and J. C. Vàrilly, *Relativistic quantum kinematics in the Moyal representation*, J. Phys. A: Math. Gen. **23** (1990), 901–933.
[24] M. Davidson, G. Òlafsson and G. Zhang, *Laplace and Segal-Bargmann transforms on Hermitian symmetric spaces and orthogonal polynomials*, J. Funct. Anal. **204** (2003), 157–195.
[25] H. Figueroa, J. M. Gracia-Bondìa and J. C. Vàrilly, *Moyal quantization with compact symmetry groups and noncommutative analysis*, J. Math. Phys. **31** (1990), 2664–2671.
[26] B. Folland, *Harmonic Analysis in Phase Space*, Princeton Univ. Press, 1989.
[27] J. M. Gracia-Bondìa, *Generalized Moyal quantization on homogeneous symplectic spaces*, Deformation theory and quantum groups with applications to mathematical physics (Amherst, MA, 1990), pp. 93–114, Contemp. Math., 134, Amer. Math. Soc., Providence, RI, 1992.
[28] J. M. Gracia-Bondìa and J. C. Vàrilly, *The Moyal Representation for Spin*, Ann. Physics **190** (1989), 107–148.
[29] M. Gotay, *Obstructions to Quantization*, in: Mechanics: From Theory to Computation (Essays in Honor of Juan-Carlos Simo), J. Nonlinear Science Editors, Springer New-York, 2000, pp. 271–316.
[30] S. Helgason, *Differential Geometry, Lie Groups and Symmetric Spaces*, Graduate Studies in Mathematics, Vol. 34, American Mathematical Society, Providence, Rhode Island 2001.
[31] A. W. Knapp, *Representation theory of semi-simple groups. An overview based on examples*, Princeton Math. Series t. 36 , 1986.
[32] S. Kilic, *The Berezin symbol and multipliers on functional Hilbert spaces*, Proc. Amer. Math. Soc. **123** (1995), 3687–3691.
[33] A. A. Kirillov, *Lectures on the Orbit Method*, Graduate Studies in Mathematics Vol. 64, American Mathematical Society, Providence, Rhode Island, 2004.
[34] B. Kostant, *Quantization and unitary representations*, in: Modern Analysis and Applications, Lecture Notes in Mathematics 170, Springer-Verlag, Berlin, Heidelberg, New-York, 1970, pp. 87–207.
[35] B. Kostant, *Lie group representations on polynomial rings*, Amer. J. Math. **86** (1963), 327–404.
[36] S. Luo, *Polar decomposition and isometric integral transforms*, Int. Transf. Spec. Funct. **9** (2000), 313–324.
[37] K-H. Neeb, *Holomorphy and Convexity in Lie Theory*, de Gruyter Expositions in Mathematics, Vol. 28, Walter de Gruyter, Berlin, New-York 2000.
[38] T. Nomura, *Berezin Transforms and Group representations*, J. Lie Theory **8** (1998), 433–440.
[39] B. Ørsted and G. Zhang, *Weyl Quantization and Tensor Products of Fock and Bergman Spaces*, Indiana Univ. Math. J. **43** (1994), 551–583.
[40] N. V. Pedersen, *Matrix coefficients and a Weyl correspondence for nilpotent Lie groups*, Invent. Math. **118** (1994), 1–36.

[41] I. Satake, *Algebraic structures of symmetric domains*, Iwanami Sho-ten, Tokyo and Princeton Univ. Press, Princeton, NJ, 1971.
[42] R. L. Stratonovich, *On distributions in representation space*, Soviet Physics. JETP **4** (1957), 891–898.
[43] A. Unterberger and H. Upmeier, *Berezin transform and invariant differential operators*, Commun. Math. Phys. **164, 3** (1994), 563–597.
[44] N. J. Wildberger, *Convexity and unitary representations of a nilpotent Lie group*, Invent. Math. **89** (1989) 281–292.
[45] N. J. Wildberger, *On the Fourier transform of a compact semisimple Lie group*, J. Austral. Math. Soc. A **56** (1994), 64–116.
[46] G. Zhang, *Berezin transform on compact Hermitian symmetric spaces*, Manuscripta Math. **97** (1998), 371–388.

Author's address:

Benjamin Cahen
Université de Lorraine, Site de Metz, UFR-MIM
Département de mathématiques
Bâtiment A
Ile du Saulcy, CS 50128, F-57045, Metz cedex 01, France
E-mail: `benjamin.cahen@univ-lorraine.fr`

Received August 13, 2014

Rend. Istit. Mat. Univ. Trieste
Volume 46 (2014), 181–201

Two classes of real numbers and formal power series: quasi algebraic objects

Maurizio Citterio and Rodolfo Talamo

Abstract. *Motivated by the endeavour to extend to the algebraic irrationals the notion of best rational approximation to a given real number, we define the concept of quasi algebraic number and prove some results related to it. We apply these results to the study of the Schröder functional equation with quasi algebraic parameter. The main definitions can be transposed to the field of formal Laurent series over a finite field. In this respect we prove that every badly approximable series is quasi algebraic.*

Keywords: quasi algebraic objects, algebraic objects of best approximation.
MS Classification 2010: 11J17, 11J61.

1. Introduction

The present work originates from the purpose to extend the concept of best approximation r to a given real number ξ (which is defined for rational r) to the algebraic irrationals. In the literature there are two different notions of best approximations, namely the best approximations of the first and of the second kind [6], which are also referred to respectively as fair and good approximations [10]. A good approximation is also a fair one, while an analogous implication for algebraic approximations of higher degree does not hold: if $P(x)$ and $Q(x)$ are polynomials of degree $n > 1$, $P(\xi) \neq 0$, $Q(\xi) \neq 0$, $|P(\xi)| < |Q(\xi)|$, $H(Q) \leq H(P)$ (see the definition of height), it does not follow that $P(x)$ has a root which is closer to ξ than any root of $Q(x)$. Therefore we will be concerned with the extension of the notion of best approximation of the first kind which is more directly related with the distance between ξ and its approximation. For any irrational number ξ the sequence of its convergents is a uniquely determined sequence of rationals of best approximation. We have considered the problem of determining a sequence $\{\alpha_n\}$ of irrational algebraic numbers of bounded degree and of best approximation in a sense that we are going to define. The existence of such a sequence is not assured in general, as in many cases (like the Liouville numbers) the rationals outdo the algebraic numbers of higher degree in their approximation power.

There is a large literature concerning approximations by algebraic numbers (see [3], especially Chapter 2 and 3). The peculiar features of our approach are just these notions: after giving the appropriate definitions, the necessity of providing a suitable context for the existence of irrational algebraic numbers of best approximation has led us to considering a new class of irrational numbers, which we have called *quasi algebraic*. This class contains almost every real number in the sense of Lebesgue measure and has dense and uncountable complement in the set of real numbers. We prove some diophantine properties of quasi algebraic numbers which apply to the study of the Schröder functional equation. Our main definitions (quasi algebraic irrational numbers and irrational algebraic numbers of best approximation) can be transposed verbatim, mutatis mutandis, to the field $\mathbb{K}\left(\left(X^{-1}\right)\right)$ of formal Laurent series with coefficients in a field $\mathbb{K}$. We focus on the case $\mathbb{K}$ is a finite field.

We thank the referee for the valuable hints which have led to a substantial improvement of Theorem 3.1.

2. Notation and terminology

Let $a_0, a_1, \ldots, a_n$ be positive real numbers. Let's pose

$$[a_0, a_1, \ldots, a_n] = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots\cdots\cdots\cdots \cfrac{}{a_{n-1} + \cfrac{1}{a_n}}}}$$

In particular, if the a_i are integers, if we set

$$p_{-2} = 0\ , \quad p_{-1} = 1\ , \quad q_{-2} = 1\ , \quad q_{-1} = 0\ ,$$

and define inductively

$$p_i = a_i p_{i-1} + p_{i-2}\ , \quad q_i = a_i q_{i-1} + q_{i-2}\ , \quad (i \geq 0)\ ,$$

we have

$$\frac{p_n}{q_n} = [a_0, \ldots, a_n] \quad \text{and} \quad p_{n-1}q_n - p_n q_{n-1} = (-1)^n\ .$$

Every real number ξ can be developed in continued fraction

$$\xi = [a_0, a_1, \ldots, a_n, \ldots] = \lim_{n \to +\infty} \frac{p_n}{q_n}.$$

The fractions p_n/q_n are called the convergents of ξ. The development is finite if ξ is rational, infinite if ξ irrational, ultimately periodic if ξ is quadratic. The following identity holds:

$$\xi = [a_0, a_1, \dots, a_n, A_{n+1}],$$

where $A_{n+1} = [a_{n+1}, a_{n+2}, \dots]$. The a_i are called the partial quotients of ξ and the A_i the complete quotients. If $s_{n-1} = p_{n-1}/q_{n-1}$, $s_n = p_n/q_n$, $n > -1$ are two successive convergents and $a_{n+1} > 1$, a fraction of the form

$$s_{n,k} = \frac{kp_n + p_{n-1}}{kq_n + q_{n-1}} = \frac{p_{n,k}}{q_{n,k}} \ , \quad \text{with} \quad 1 \le k < a_{n+1}$$

is called *secondary convergent*. An irrational number ξ is said to be badly approximable or of constant type if there is a constant $C > 0$ such that

$$\left| \xi - \frac{p}{q} \right| \ge \frac{C}{q^2} \ ,$$

for every $p/q \in \mathbb{Q}$; ξ is of constant type if and only if the sequence of its partial quotients is bounded.

Given a polynomial with real coefficients

$$P(x) = a_n x^n + \cdots + a_0,$$

we define the *height*

$$H(P) := \max_{i \le n} |a_i| \ ,$$

and the *length*

$$L(P) := \sum_{i \le n} |a_i| \ .$$

Given an algebraic number α, we indicate with $H(\alpha)$ and $L(\alpha)$ the height and the length of its minimal polynomial over $\mathbb{Z}$. We denote by $d(\alpha)$ its degree.

Let n be a positive integer and ξ any real number. We say that an algebraic α of degree n is of best approximation for ξ if the following condition holds:

> for any algebraic β with $1 \le d(\beta) \le n$ and $H(\beta) \le H(\alpha)$, we have $|\alpha - \xi| \le |\beta - \xi|$.

We say that an irrational number ξ is quasi algebraic if there is an integer $m \ge 2$ such that the following condition $\mathcal{C}$ is fulfilled:

$\mathcal{C}$: for almost every rational r (i.e. except at most a finite number) there is an algebraic irrational $\alpha_{(r)}$ such that $d\left(\alpha_{(r)}\right) \le m$, $H\left(\alpha_{(r)}\right) \le H(r)$ and $\left|\xi - \alpha_{(r)}\right| \le |\xi - r|$.

If ξ is quasi algebraic and s is the least integer m for which the condition above holds, we say that s is the degree of ξ as a quasi algebraic number and write $s = \partial(\xi)$.

Two real numbers α, β are said to be equivalent if $\alpha = \dfrac{m\beta + p}{n\beta + q}$ with $m, n, p, q \in \mathbb{Z}$ and $|mq - np| = 1$.

If α is a real number, we denote by $\langle \alpha \rangle$ the distance between α and the set $\mathbb{Z}$ of integers.

3. Main results for real numbers

We take into account the set of all quasi algebraic numbers whose degree does not exceed 3 in order to investigate its metric properties as well as its relationship with the algebraic irrationals and with the numbers of constant type.

THEOREM 3.1. *1. Almost every real number in the sense of Lebesgue measure, including the algebraic irrationals, is quasi algebraic of degree* 2.

2. If $\bar{\xi}$ is of constant type, there exists at least one quasi algebraic ξ with $\partial(\xi) \leq$ 3 such that ξ and $\bar{\xi}$ are equivalent.

Proof. 1. Recall the definition of the function $w_n(\xi)$, for ξ real and n a positive integer: $w_n(\xi)$ is the supremum of the set of real numbers w with the property that there are infinitely many polynomials $P(x)$ of degree $\leq n$ with integer coefficients such that

$$0 < |P(\xi)| \leq H(P)^{-w} \ .$$

It is known that $w_1(\xi) = 1$ for almost all real numbers. Let us define the set $B_{1,2}$ of the real numbers ξ with the properties that $w_1(\xi) = 1$ and there is $H_0(\xi) > 0$ such that the following holds: for every $H \geq H_0$ there is a quadratic α with

$$H^{0.9} \leq H(\alpha) \leq H \quad \text{and} \quad |\xi - \alpha| \leq H(\alpha)^{-2.9} \ .$$

It follows from Corollaire 1 of [2], with $d = 2$ and $\varepsilon = 0, 1$, that $B_{1,2}$ contains almost every real number. Furthermore, as $w_2(\xi) = 2$, $w_1(\xi) = 1$ for all algebraic numbers of degree ≥ 3 (see [3, p.45]), from Théorème 1 of [2], taking $\lambda = 1$ and $\varepsilon = 0.1$ we obtain that every such algebraic belongs to $B_{1,2}$. Now we are going to show that every number in $B_{1,2}$ is quasi algebraic of degree 2. Let $\xi \in B_{1,2}$. Consider the interval $I_\xi = (\xi - \varepsilon_1, \xi + \varepsilon_1)$ where ε_1 is chosen in the following way:

a. if $|\xi| < 1$, then $-1 < \xi - \varepsilon_1 < \xi + \varepsilon_1 < 1$;

b. if $\xi > 1$, then $1 < \xi - \varepsilon_1$;

c. if $\xi < -1$, then $\xi + \varepsilon_1 < -1$.

Observe that, if $p/q \in I_\xi$, there is a constant C_ξ such that $q^{-1} > H\left(p/q\right)^{-1} C_\xi$; in fact, take C_ξ respectively equal 1, $\xi - \varepsilon_1$, $|\xi + \varepsilon_1|$ in the cases *a*, *b*, *c*. Fix two quadratics $\alpha_1, \alpha_2 \in I_\xi$, $\alpha_1 < \xi$, $\alpha_2 > \xi$ with $H(\alpha_1) = H_1$, $H(\alpha_2) = H_2$. Since $w_1(\xi) = 1$, there is a finite number of rationals $r_i = p_i/q_i$, $i \leq k$, such that

$$|q_i \xi - p_i| \leq H(r_i)^{-1.1} \tag{*}$$

Let $H_3 = \max_{i \leq k} H(r_i)$. From the defining condition of $B_{1,2}$, consider the H_0 associated with ξ and $H_4 = \max\left\{H_0, H_1, H_2, H_3, C_\xi^{-10}\right\}$. We are going to prove that our Condition $\mathcal{C}$ in the definition of quasi algebraic numbers is fulfilled, for $m = 2$. Take a rational r, $H(r) > H_4$. If $r \notin I_\xi$, then $|\xi - \alpha_1| < |\xi - r|$ if $r < \xi$, $|\xi - \alpha_2| < |\xi - r|$ if $r > \xi$, $H_1, H_2 < H(r)$. If $r = p/q \in I_\xi$, then it follows:

$$|\xi - r| > q^{-1} H(r)^{-1.1} > C_\xi H(r)^{-2.1} > H(r)^{-2.2} \, .$$

On the other hand, there is a quadratic $\alpha_{(r)}$, with $H(r)^{0.9} \leq H\left(\alpha_{(r)}\right) \leq H(r)$ and

$$\left|\xi - \alpha_{(r)}\right| \leq H\left(\alpha_{(r)}\right)^{-2.9} \leq H(r)^{-2.61} \, .$$

This completes the proof of part 1, as quadratic numbers are trivially quasi algebraic of degree 2, taking $\alpha_{(r)} = \xi$ for $H(r) \geq H(\xi)$.

2. Let $\bar{\xi}$ be a number of constant type. For what we proved in 1, we can suppose that $\bar{\xi}$ is a transcendental. Denote by $\xi = [a_0, a_1, \ldots, a_n, \ldots]$ any complete quotient of $\bar{\xi}$ associated with a maximal partial quotient. Therefore ξ and $\bar{\xi}$ are equivalent and $a_0 \geq a_i$, $i \geq 1$. Besides $a_0 \geq 2$ as $\bar{\xi}$ is not quadratic. We are going to prove that ξ is quasi algebraic with $\partial(\xi) \leq 3$. Divide the proof in three steps.

   i. Given a convergent $s_n = p_n/q_n$ of ξ, we prove that there is a cubic α_n satisfying the condition $\mathcal{C}$, with $r = s_n$. Observe that $\xi^2 > A_{n+1} = [a_{n+1}, \ldots]$, because $a_0 \geq 2$ by hypothesis. Consider

   $$\bar{x}(t) = \frac{tp_n + p_{n-1}}{tq_n + q_{n-1}} \, , \quad \bar{y}(t) = \sqrt{t} \, .$$

   Suppose at first that $s_n > s_{n-1}$, namely $\bar{x}(t)$ monotonically increasing. As $\bar{x}(A_{n+1}) = \xi$ and $\xi^2 > A_{n+1}$, we have $\bar{x}(\xi^2) > \bar{x}(A_{n+1}) = \xi = \bar{y}(\xi^2)$. As $\bar{y}(s_n^2) = s_n > \bar{x}(s_n^2)$, there is t_0, $\xi^2 < t_0 < s_n^2$, such that $\bar{x}(t_0) = \bar{y}(t_0)$. Then $t_0 = x_0^2$, $\xi < x_0 < s_n$. In the case $s_n < s_{n-1}$ and $\bar{x}(t)$ decreasing, we come, with symmetric arguments, to the inequality $s_n < x_0 < \xi$, while the equality $\bar{y}(t_0) = \bar{x}(t_0)$ holds with $s_n^2 < t_0 < \xi^2$. In both cases it follows:

$$\bar{x}(t_0) = \bar{x}(x_0^2) = \frac{x_0^2 p_n + p_{n-1}}{x_0^2 q_n + q_{n-1}} = \bar{y}(t_0) = x_0 \ .$$

Then x_0 is a solution of the cubic equation

$$q_n x^3 - p_n x^2 + q_{n-1}x - p_{n-1} = 0 \tag{1}$$

and $|\xi - x_0| < |\xi - s_n|$. It remains to show that the Equation (1) is irreducible in $\mathbb{Z}$, in order to conclude that $H(x_0) \leq H(s_n)$; this check is necessary, as the height of the product of two polynomials can be strictly less than the height of one of the factors. The first member of (1) is primitive, being the convergents reduced fractions. Moreover, it does not admit negative root, as p_n, q_n, p_{n-1} and q_{n-1} are positive. As the Equation (1) is equivalent to

$$x = \frac{x^2 p_n + p_{n-1}}{x^2 q_n + q_{n-1}} \ ,$$

α is a root of (1) if and only if $\bar{x}(\alpha^2) = \alpha = \bar{y}(\alpha^2)$, i.e. if and only if α^2 is a zero of the function $z(t) = \bar{x}(t) - \bar{y}(t)$. But $z'(t) \neq 0$ for $t > 0$; thus the only zero of $z(t)$ is x_0^2. Consequently, x_0 is the only real root of (1). If (1) were not irreducible over $\mathbb{Z}$, there would exist a factorization of the type:

$$q_n x^3 - p_n x^2 + q_{n-1}x - p_{n-1} = (bx - a)(lx^2 + mx + n)$$

with a, b, l, m, n in $\mathbb{Z}$. For what we have just proved above, $x_0 = a/b$ with $b > q_n$ since x_0 belongs to the interval of two successive convergents s_{n-1}, s_n. Equating terms yields $q_n = lb$, which is absurd. Therefore (1) is the minimal polynomial of x_0 and $H(x_0) \leq H(s_n)$.

ii. Given a secondary convergent of ξ, namely of the form

$$s_{n,k} = \frac{p_{n,k}}{q_{n,k}} = \frac{kp_n + p_{n-1}}{kq_n + q_{n-1}} \ , \quad 1 \leq k < a_{n+1} \ , \quad n \geq -1$$

(existing if $a_{n+1} > 1$), we prove that there is a quadratic $\alpha_{n,k}$ which satisfies the condition $\mathcal{C}$. Let's pose $\alpha_{n,k} = \left[\overline{a_0, \ldots, a_n, k}\right]$ if $n \geq 0$, $\alpha_{n,k} = \left[\overline{k}\right]$ if $n = -1$. We observe that $\alpha_{n,k}$ is the positive root of the equation:

$$q_{n,k}x^2 + (q_n - p_{n,k})x - p_n = 0 \ ,$$

being the fixed point of $\tilde{x}(t) = \dfrac{tp_{n,k} + p_n}{tq_{n,k} + q_n}$, as $|p_n q_{n,k} - p_{n,k} q_n| = 1$. Obviously $\alpha_{n,k}$ is quadratic and $H(\alpha_{n,k}) \leq H(s_{n,k})$. As in step *2i*, one can verify that $\alpha_{n,k}$ belongs to the interval of extremes ξ and $s_{n,k}$; thus $|\xi - \alpha_{n,k}| < |\xi - s_{n,k}|$.

iii. Consider now any rational r; set $\bar{h} = H(r)$. Let r^* be the rational (possibly coinciding with r) which attains the least distance with ξ among all the rationals of height less or equal to $\bar{h}$. It can be easily checked that r^* is either a convergent or a secondary convergent of ξ. In both cases, for what we have shown in *2i* and *2ii*, there is an algebraic $\alpha_{(r)}$ of degree 2 or 3 satisfying condition $\mathcal{C}$ with respect to r^* and therefore a fortiori with respect to r. This completes the proof. $\square$

The following proposition shows that there are numbers of constant type which are quasi algebraic of degree 2.

LEMMA 3.2. *Let m, n, p, q be positive integers, with $(m, n) = (p, q) = 1$ and $|mq - np| = 1$. Further, let k be a positive real number. Then:*

$$\left| \frac{km + p}{kn + q} - \frac{km + 2p}{kn + 2q} \right| < \left| \frac{km + p}{kn + q} - \frac{m}{n} \right|$$

Proof. Let $A = \dfrac{m}{n}$, $B = \dfrac{km + p}{kn + q}$, $C = \dfrac{km + 2p}{kn + 2q}$, we have $C < B < A$ or $A < B < C$ according to whether $mq - np$ equals 1 or -1. As $A + C - 2B = (A - B) - (B - C) = \dfrac{2q(mq - np)}{n(kn + q)(kn + 2q)}$, which is positive in the first case and negative in the second, our claim is proved. $\square$

PROPOSITION 3.3. *Let $\xi = [a_0, a_1, \ldots, a_n, \ldots]$ be transcendental of constant type, $m_0 = \min\limits_{a_i \neq 0}\{a_i\}$, $M_0 = \max\limits_{i \geq 0}\{a_i\}$. If $M_0/m_0 \leq 2$, then $\partial(\xi) = 2$, i.e. ξ is quasi quadratic.*

Proof. Suppose firstly $a_0 \geq 1$. Pose $\alpha_1 = \left[\overline{m_0, M_0}\right]$, $\alpha_2 = \left[\overline{M_0, m_0}\right]$. Observe that α_1 and α_2 are respectively the minimum and the maximum of all irrationals whose partial quotients are included between m_0 and M_0; moreover $\alpha_2/\alpha_1 = M_0/m_0$. Then $\alpha_1 < \xi < \alpha_2$. To prove that $\partial(\xi) = 2$, owing to the results of Theorem 3.1, we need only to consider the convergents of ξ as the quadratic associated with a secondary convergent can be constructed exactly in the same way as in step *2ii* of it. In the notations of the theorem, let $s_N = p_N/q_N$, $N \geq 0$ and $\bar{x}(t)$ the related map. Define $\alpha_{(N)} = [\overline{a_0, a_1, \ldots, a_N}]$. Of course $\alpha_1 \leq \alpha_{(N)} \leq \alpha_2$ and $\alpha_1 < A_{N+1} < \alpha_2$. We distinguish two cases, since $A_{N+1} \neq \alpha_{(N)}$, as A_{N+1} is transcendental.

1. $A_{N+1} < \alpha_{(N)}$. As $\bar{x}(\alpha_{(N)}) = \alpha_{(N)}$, $\bar{x}(A_{N+1}) = \xi$, $\lim_{t\to+\infty} \bar{x}(t) = s_N$ and $\bar{x}(t)$ is monotone, $\alpha_{(N)}$ belongs to the interval between ξ and s_N, and thus $|\xi - \alpha_{(N)}| < |\xi - s_N|$.

2. $A_{N+1} > \alpha_{(N)}$. As $\alpha_2/\alpha_1 = M_0/m_0 \leq 2$ and $A_{N+1} < \alpha_2$, it follows $A_{N+1}/2 < \alpha_{(N)}$ from $\alpha_1 \leq \alpha_{(N)}$. If in Lemma 3.2 we pose $m = p_N$, $n = q_N$, $p = p_{N-1}$, $q = q_{N-1}$, $k = A_{N+1}$, we have $A = s_N$, $B = \xi = \bar{x}(A_{N+1})$, $C = \bar{x}(k/2)$. From already known facts about $\bar{x}(t)$ and the inequality $A_{N+1}/2 < \alpha_{(N)}$, it follows that $\alpha_{(N)}$ lies between B and C; therefore we have $|\xi - \alpha_{(N)}| < |\xi - C|$. Since from Lemma 3.2 follows $|\xi - C| < |\xi - s_N|$ and $H(\alpha_{(N)}) \leq H(s_N)$, because the minimal polynomial of $\alpha_{(N)}$ is

$$q_N x^2 + (q_{N-1} - p_N)x - p_{N-1} \ ,$$

the proof is complete in the case $a_0 \geq 1$.

If $a_0 = 0$, put $\alpha_{(N)} = [0, \overline{a_1, \ldots, a_N}]$ and $\bar{\alpha}_{(N)} = [\overline{a_1, \ldots, a_N}]$; proceed like before considering the two cases $A_{N+1} > \bar{\alpha}_{(N)}$ and $A_{N+1} < \bar{\alpha}_{(N)}$ and taking into account the fact that now $\bar{x}\left(\bar{\alpha}_{(N)}\right) = \alpha_{(N)}$. □

In [3, p.39] it is observed that the Thue-Morse word on $\{1, 2\}$ (see [11]) "is too well approximable by quadratic numbers to be algebraic, but we do not know precisely how well it is approximable". As a matter of fact, it follows directly from Proposition 3.3 that this number is quasi quadratic. Therefore the above question about the Thue-Morse word con be considered in the more general framework of quasi quadratic numbers. The following proposition quantifies the goodness of the approximation of a quasi quadratic number by quadratics.

Proposition 3.4. *Let ξ be a quasi quadratic transcendental number. There is a uniquely determined sequence of quadratics of best approximation $\{\alpha_n\}$, $n \in \mathbb{N}$, converging to ξ and such that $|\xi - \alpha_n| = O\left(H(\alpha_n)\right)^{-2}$.*

Proof. We are going to define inductively an integer r_n and a quadratic α_n. Let r_1 be the least integer such that the defining condition $\mathcal{C}$ of quasi algebraic numbers holds (for $m = 2$) for every convergent p_i/q_i of ξ, with $i \geq r_1$. Let $h_1 = h\left(p_{r_1}/q_{r_1}\right)$. Among all quadratics of height equal or less h_1, let α_1 be the nearest to ξ; it is unique for the transcendence of ξ. Suppose we have defined r_n and α_n; let r_{n+1} be the least integer i such that $|\xi - p_i/q_i| < |\xi - \alpha_n|$ and let $h_{n+1} = h\left(p_{r_{n+1}}/q_{r_{n+1}}\right)$. Among all quadratics whose height is less or equal h_{n+1}, let α_{n+1} be the nearest to ξ. As $h_n < (|\xi| + 1)\, q_{r_n}$, from the hypotheses on ξ it follows:

$$|\xi - \alpha_n| < \left|\xi - \frac{p_{r_n}}{q_{r_n}}\right| < \frac{1}{q_{r_n}^2} < \frac{(|\xi| + 1)^2}{h_n^2} < \frac{(|\xi| + 1)^2}{[H(\alpha_n)]^2} \ .$$

It is easily checked that α_n are of best approximation. Besides, the construction determines them univocally. $\square$

REMARK 3.5. *Observe that the convergents s_n of a real number ξ are rationals of best approximation such that $|\xi - s_n| = O\left(H(s_n)\right)^{-2}$, namely the same properties which characterize the quadratics α_n with respect to the quasi quadratic ξ. Roughly speaking, a quasi quadratic number admits a sequence of quadratic convergents; they approximate ξ better than the rational convergents s_n, for less or equal height.*

Furthermore, if ξ satisfies the hypotheses of Proposition 3.3, the integer r_1 in Proposition 3.4 can be taken equal zero. Consequently, for every ration r there is a quadratic of best approximation α_r with $H(\alpha_r) \leq H(r)$ and $|\alpha_r - \xi| < |r - \xi|$.

Now we are going to prove a diophantine property of quasi algebraic numbers. It will enable us to demonstrate the existence of uncountably many real numbers which are not quasi algebraic. Moreover it ensures the existence of a solution for the Schröder functional equation with quasi algebraic parameter.

THEOREM 3.6. *Let ξ be a transcendental quasi algebraic number in the unit interval, $\partial(\xi) = s$. Then the following hold:*

i. there is a real positive constant K depending only on ξ such that, for any $p, q \in \mathbb{Z}$, $q \geq 1$,

$$|q\xi - p| > Kq^{-s} \;;$$

ii. for any $\varepsilon > 0$ the series

$$\sum_{h\geq 1} h^{-(s+\varepsilon)} \langle h\xi \rangle^{-1}$$

converges and its sum $S(\varepsilon)$ is $O\left(\zeta'(1+\varepsilon)\right)$ for $\varepsilon \to 0^+$, where ζ' is the derivative of the Riemann ζ function.

Proof. *i.* For $n \geq n_0$, to every convergent s_n of ξ can be associated an algebraic irrational $\alpha_{(s_n)} = \alpha_n$ with $d(\alpha_n) = r_n \leq s$, $H(\alpha_n) \leq H(s_n)$ (following our definition of quasi algebraic number which excludes possibly a finite number of rationals for Condition $\mathcal{C}$) and $|\alpha_n - \xi| < |s_n - \xi|$. We are going to apply Güting's Theorem to α_n and the polynomial $P_n(x) = q_n x - p_n$. From [5] follows, for $n \geq n_0$:

$$|q_n \alpha_n - p_n| \geq \frac{1}{L(\alpha_n)\ (q_n + |p_n|)^{r_n - 1}} \;.$$

As $L(\alpha_n) \leq H(\alpha_n)\,(r_n + 1)$ we get, dividing by q_n:

$$|\alpha_n - s_n| \geq \frac{1}{H(\alpha_n)\ (r_{n+1})\ (1 + |s_n|)^{r_n - 1}\, q_n^{r_n}} \;.$$

Since $H(\alpha_n) \le H(s_n) \le q_n\,(|\xi|+1)$ and $r_n \le s$ we obtain:

$$|\alpha_n - s_n| \ge \frac{1}{q_n^{s+1}(s+1)\,(|\xi|+1)\,(|\xi|+2)^{s-1}}\ .$$

If we put $C = \dfrac{1}{(s+1)\,(|\xi|+1)\,(|\xi|+2)^{s-1}}$, we get:

$$|\alpha_n - s_n| \ge \frac{C}{q_n^{s+1}}\ , \quad \forall n \ge n_0\ . \tag{2}$$

As $|\xi - s_n| < \dfrac{1}{a_{n+1}q_n^2}$ and $|\xi - \alpha_n| < |\xi - s_n|$, this yields:

$$|\alpha_n - s_n| \le |\alpha_n - \xi| + |\xi - s_n| < \frac{2}{a_{n+1}q_n^2}\ . \tag{3}$$

From (2) and (3) we have

$$a_{n+1} < \frac{2}{C}\,q_n^{s-1}\ , \quad \forall n \ge n_0\ .$$

Determine now constants $C_1, \dots, C_{n_0}$ such that:

$$a_1 < C_1 q_0^{s-1}\ ,\ \dots\ ,\ a_{n_0} < C_{n_0} q_{n_0-1}^{s-1}\ .$$

If $D = \sup\left(C_1, \dots, C_{n_0}, \frac{2}{C}\right)$ we get:

$$a_{n+1} < Dq_n^{s-1},\ n \ge 0. \tag{4}$$

From the equation:

$$|\alpha q_n - p_n| = \frac{1}{A_{n+1}q_n + q_{n-1}}, \tag{5}$$

as $A_{n+1} < a_{n+1} + 1$, we get eventually from (4):

$$|\alpha q_n - p_n| > (D+2)^{-1} q_n^{-s}.$$

As $\xi \in [0,1]$, for any $p, q \in \mathbb{Z}$, $q \ge 1$, there is a convergent p_n/q_n, $n \ge 0$, such that $\left|\xi - \dfrac{p}{q}\right| \ge \left|\xi - \dfrac{p_n}{q_n}\right|$, $q_n \le q$. We get eventually $|q\xi - p| = q\left|\xi - \dfrac{p}{q}\right| \ge q\left|\xi - \dfrac{p_n}{q_n}\right| \ge |q_n\xi - p_n| > Kq_n^{-s} > Kq^{-s}$, $K = (D+2)^{-1}$.

ii. From (4) and Theorem 2 of [7], it follows that

$$\sum_{h}^{N} h^{-s}\langle h\xi\rangle^{-1} = O(\log N)\ ,\ N \geq 2\ .$$

Define $v_h = h^{-s}\langle h\xi\rangle^{-1}$, $u_h = h^{-\varepsilon}$, $h > 1$. Then, as $\{v_h\}$ and $\{u_h\}$ are both positive and $\{u_h\}$ is decreasing, summation by parts can be applied, namely (see [7, p.231]):

$$\sum_{h=1}^{N} u_h v_h = \sum_{h=1}^{N-1}(u_h - u_{h+1})\sum_{j=1}^{h} v_j + u_N \sum_{j=1}^{N} v_j\ .$$

Consequently, there is $\tilde{M} > 0$ such that

$$\begin{aligned}
&\sum_{h=1}^{N} h^{-(s+\varepsilon)}\langle h\xi\rangle^{-1} \leq \\
&\quad \leq \langle\xi\rangle^{-1}\left(1-2^{-\varepsilon}\right) + \tilde{M}\sum_{2}^{N-1}\frac{(h+1)^{\varepsilon}-h^{\varepsilon}}{h^{2\varepsilon}}\log h + \tilde{M}N^{-\varepsilon}\log N \leq \\
&\qquad \leq \langle\xi\rangle^{-1}\left(1-2^{-\varepsilon}\right) + \tilde{M}\sum_{2}^{N-1} h^{-(1+\varepsilon)}\log h + \tilde{M}N^{-\varepsilon}\log N\ .
\end{aligned}$$

Thus $\sum_{h\geq 1} h^{-(s+\varepsilon)}\langle h\xi\rangle^{-1}$ converges for any $\varepsilon > 0$; moreover, for ε sufficiently small,

$$S(\varepsilon) < \left(\tilde{M}+1\right)|\zeta'(1+\varepsilon)|\ . \qquad \square$$

Corollary 3.7. *The set of quasi algebraic numbers has dense and uncountable complement in* $\mathbb{R}$.

Proof. Set $\alpha = [a_0, \ldots, a_n, \ldots]$ any real number such that $a_{n+1} > q_n^n$, $n \geq n_0$. Clearly the set of such α is dense because n_0 and the first $n_0 + 1$ partial quotients are quite arbitrary; moreover, for every $n \geq n_0$, there are countably many ways to choose a_{n+1}; then the α are uncountably many. Let s be a positive real number. For any convergent p_n/q_n of α we have, from (5):

$$\langle q_n\alpha\rangle = |q_n\alpha - p_n| < \frac{1}{a_{n+1}q_n} < \frac{1}{q_n^{n+1}}\ .$$

Thus $q_n^{-s}\langle q_n\alpha\rangle^{-1} > q_n^{n-s+1}$. For $n > s$ the terms tend to infinity; consequently the series diverges for any $s > 0$ and α cannot be quasi algebraic for *i.* $\square$

4. The Schröder functional equation

Let $f(z) = \lambda z + a_2 z^2 + \cdots + a_n z^n + \ldots$, $a_i \in \mathbb{C}$, $\lambda = \mathrm{e}^{2\pi i \alpha}$, $\alpha \in [0,1]$ be analytical in a neighborhood of zero. The problem of the existence of an analytic solution $\phi(z) = z + b_2 z^2 + \cdots + b_n z^n + \ldots$ (called Königs function of f) of the Schröder functional equation

$$\phi(\lambda z) = f(\phi(z)) \tag{6}$$

is closely related with the diophantine properties of α. Let

$$\Gamma = \left\{\alpha \in [0,1] : \exists \varepsilon, \mu > 0 \text{ such that } |q\alpha - p| > \varepsilon q^{-\mu} \text{ for any } p, q \in \mathbb{Z}, q \geq 1\right\}.$$

Siegel [14, 15] proved that the analytical solution of (6) exists if $\alpha \in \Gamma$. In the paper [16] the author tackled the problem of approximating a given Königs function with a sequence of "quadratic Königs functions" (namely deriving from an α quadratic). Now we want to give an improved version of the result in [16] and apply it to the case in which α is a quasi algebraic.

Property 1. *Let $f(z) = \lambda z + a_2 z^2 + \cdots + a_n z^n + \ldots$ analytical in a neighborhood of zero, $\lambda = \mathrm{e}^{2\pi i \alpha}$, $\alpha \in \Gamma$, $\alpha \in [0,1]$ and $\{\lambda_n\}_{n \in \mathbb{N}}$, $\lambda_n = \mathrm{e}^{2\pi i \alpha_n}$, $\alpha_n \in \Gamma$ any sequence such that $\lim_{n\to\infty} \alpha_n = \alpha$ and α_n, α have the same constants ε, μ as elements of Γ. Let $f_n(z) = f(z) + (\lambda_n - \lambda)z$ and denote by $\phi(z)$, $\phi_n(z)$ the Königs functions of $f(z)$, $f_n(z)$. Then $\phi(z) = \lim_{n\to\infty} \phi_n(z)$ uniformly in a neighborhood of zero.*

Proof. If $\phi_k(z) = z + \sum_{n\geq 2} b_{n,k} z^n$, $\phi(z) = z + \sum_{n \geq 2} b_n z^n$, write $S_{n_0,k}(z) = z + \sum_{n \leq n_0} b_{n,k} z^n$, $S_{n_0}(z) = z + \sum_{n \leq n_0} b_n z^n$, $R_{n_0,k}(z) = \phi_k(z) - S_{n_0,k}(z)$, $R_{n_0}(z) = \phi(z) - S_{n_0}(z)$. In [16, p.54] we observe that the functions ϕ, ϕ_k are analytic in an open disc whose radius r depends only on ε and μ. Let $r' < r$. Obviously the maps ϕ, ϕ_k converge absolutely in the closed disc of radius r'. From [14] follows the existence of a series of constants $\sum_{i \geq 1} \bar{b}_i$ ($\bar{b}_i > 0$) whose coefficients depend on r', ε, μ (majoring series) such that $|b_n z^n| < \bar{b}_n$, $|b_{n,k} z^n| < \bar{b}_n$, $n \geq 1, k \geq 1, |z| \leq r'$. (See also [15, p.124] for the series $\psi = \sum_{n \geq 2} \dfrac{a_k}{\lambda^k - \lambda} z^n$ which is the first step of the iteration process). Remark that the coefficients b_n of ϕ, which are deduced inductively from the Schröder functional equation (6), have the form

$$b_n = \frac{P_n(\lambda, a_2, \ldots, a_n)}{\lambda^{n-1}(\lambda - 1)^{r_1}(\lambda^2 - 1)^{r_2} \ldots (\lambda^{n-1} - 1)^{r_{n-1}}},$$

where the P_k are polynomial with integer coefficients in $\lambda, a_2, \ldots, a_n$ and the r_i are positive integers. Therefore the b_n are continuous functions of α, for $\alpha \in \Gamma$ and $\lambda = \mathrm{e}^{2\pi i \alpha}$, as the denominators never vanish. Given now $\eta > 0$, determine n_0 such that $\sum_{n > n_0} \bar{b}_n < \eta/2$; then $|R_{n_0,k}(z)| < \eta/2$, $|R_{n_0}(z)| < \eta/2$, for any

$k \in \mathbb{N}$, $|r| \leq r'$. For the continuity of b_n on the set Γ, there is an integer k_0 such that $|b_n - b_{n,k}| < \dfrac{\eta}{2(n_0+1)(r')^{n_0}}$, $k \geq k_0$ and $2 \leq n \leq n_0$. Thus $|\phi(z) - \phi_n(z)| < \eta$, $k > k_0$ and $|z| \leq r'$ and the sequence $\{\phi_n(z)\}$ converges to $\phi(z)$. □

Remark 4.1. *The quoted result in [16] relied on the existence of a converging subsequence of $\{\phi_n\}$ (i.e. on the existence of a subsequence of the sequence of quadratics previously constructed in the lemma). Our property shows that* any *sequence $\{\alpha_n\}$ converging to α yields the convergence of $\{\phi_n\}$, provided that the α_n and α have the same constants ε and μ as elements of Γ.*

Property 2. *Let $f(z) = \lambda z + a_2 z^2 + \cdots + a_n z^n + \ldots$ analytical in a neighborhood of zero, $\lambda = \mathrm{e}^{2\pi i\alpha}$, α quasi algebraic. Then the Schröder functional equation (6) has an analytic solution; moreover there is a sequence of quadratics $\{\alpha_n\}_{n\in\mathbb{N}}$ such that the Königs functions $\phi_n(z)$of the maps $f_n(z)$ defined as in Property 1 converge uniformly to $\phi(z)$ in a neighborhood of zero.*

Proof. From *i.* of Theorem 3.6, it follows that $\alpha \in \Gamma$ with constants $\varepsilon = K$, $\mu = s = \partial(\alpha)$. Therefore the analytical ϕ exists for the quoted Siegel theorem. If $\alpha = [a_0, a_1, \ldots, a_n, \ldots]$ and $a_0 = 0$, define $\alpha_k = [a_{0,k}, a_{1,k}, \ldots, a_{n,k}, \ldots]$ where $a_{n,k} = a_n$ if $n \leq k$, while $a_{n,k} = 1$ if $n > k$, $k > 1$. In [16, Lemma, p.52] it is shown that the a_k have the same constants ε, μ of α, if $\alpha \in \Gamma$. Therefore the statement follows from Property 1. □

5. Formal Laurent series on a finite field

Let $\mathbb{K}$ be a finite field. Denote by $\mathbb{K}[X]$ the ring of polynomials, by $\mathbb{K}(X)$ the field of rational functions and by $\mathbb{K}\left(\left(X^{-1}\right)\right)$ the field of formal Laurent series with coefficients in $\mathbb{K}$. If $f \in \mathbb{K}\left(\left(X^{-1}\right)\right)$, $f = \sum_{n \leq n_0} f_n X^n$, $n_0 \in \mathbb{Z}$, $f_{n_0} \neq 0$, $n_0 = \deg(f)$, an absolute value is defined by setting $|0| = 0$, $|f| = e^{n_0}$ for $f \neq 0$. The resulting structure of complete metric space on $\mathbb{K}\left(\left(X^{-1}\right)\right)$ is the completion of the non Archimedean valuation on $\mathbb{K}(X)$ defined by $|f/g| = e^{\deg(f)-\deg(g)}$ (see [4]). Every $f \in \mathbb{K}\left(\left(X^{-1}\right)\right)$ can be uniquely developed in continued fraction (see [12]): $f = [a_0(x), a_1(x), \ldots, a_n(x), \ldots]$ where the partial quotients $a_i(x) \in \mathbb{K}[X]$, $i \geq 0$ and $|a_i(x)| > 1$ for $i > 0$. We indicate by $s_n(x) = p_n(x)/q_n(x)$, $p_i(x)$, $q_i(x) \in \mathbb{K}[X]$, $i \geq 0$, the convergents of f and by $A_i(x)$ the complete quotients. If $P(y) = p_n(x)y^n + p_{n-1}(x)y^{n-1} + \cdots + p_0(x)$ is a polynomial with coefficients $p_i(x) \in \mathbb{K}[X]$, the height $H(P)$ is defined as the maximum of the absolute values of the $p_i(x)$, $0 \leq i \leq n$. An $\alpha \in \mathbb{K}\left(\left(X^{-1}\right)\right)$ is algebraic if it is a root of a polynomial $P(y) \in \mathbb{K}[X][Y]$; the degree $d(\alpha)$is the degree of its minimal polynomial, while the height $H(\alpha)$ is the above defined height of the same minimal polynomial. If we replace rational numbers with rational functions and the absolute value of a real number with the absolute value of a formal Laurent series in the definition of quasi algebraic irrationals, we obtain the corresponding definition for $f \in \mathbb{K}\left(\left(X^{-1}\right)\right)$, and the same has been done (see [8, p.208]) in the literature for the notion of badly approximable element of $\mathbb{K}\left(\left(X^{-1}\right)\right)$; also the concept of algebraic irrational of best approximation can be extended this way to $\mathbb{K}\left(\left(X^{-1}\right)\right)$.

We are going to prove the

Theorem 5.1. *Let $\xi \in \mathbb{K}\left(\left(X^{-1}\right)\right)$, $\xi \notin \mathbb{K}(X)$ and badly approximable. Then it is quasi algebraic.*

Proof. Let $\xi = \sum_{n \le k_\xi} \xi_n X^n$, $k_\xi \in \mathbb{Z}$, $\xi_n \in \mathbb{K}$. Consider its development in continued fraction, $\xi = [a_0(x), a_1(x), \dots, a_n(x), \dots]$, $a_i(x) \in \mathbb{K}[X]$. Suppose first $|\xi| > 1$; obviously we can assume without loss of generality that it is transcendental. As it is badly approximable, there is a constant $C > 0$ such that, for every $n \ge 0$:

$$\left|\xi - \frac{p_n(x)}{q_n(x)}\right| \ge \frac{C}{|q_n(x)|^2} .$$

As (see [4, p.71]):

$$\left|\xi - \frac{p_n(x)}{q_n(x)}\right| = \frac{1}{|a_{n+1}(x)|\,|q_n(x)|^2} , \tag{7}$$

then $|a_{n+1}(x)| \le 1/C$, $\forall n \ge 0$.
Since$|\xi| > 1$ we have $|a_0(x)| > 1$. Therefore, if $p = \mathrm{char}(\mathbb{K})$, there is $h = p^{m_0}$, m_0 nonnegative integer such that, for every $n \ge 0$

$$|a_0(x)|^h \ge |a_n(x)| \ . \tag{8}$$

Set $m = h + 1$. Given any non-constant rational function $r \in \mathbb{K}(X)$, $r \notin \mathbb{K}$, $r = \sum_{n \le k_r} r_n X^n$, $k_r \in \mathbb{Z}$, $r_n \in \mathbb{K}$, we are going to prove the existence of $\alpha_{(r)} \in \mathbb{K}\left(\left(X^{-1}\right)\right)$, $\alpha_{(r)}$ algebraic irrational, $d(\alpha_{(r)}) \le m$, $H(\alpha_{(r)}) \le H(r)$ and $\left|\alpha_{(r)} - \xi\right| \le |r - \xi|$. We distinguish various cases according to the development of r in continued fraction:

1) r is a convergent of ξ, namely there is a $N \ge 0$ such that

$$r = [a_0(x), a_1(x), \dots, a_N(x)] = s_N(x).$$

Let's pose

$$\alpha_{(r)} = \left[a_0(x), \dots, a_N(x), a_0^H(x), \dots, a_N^h(x), a_0^{h^2}(x), \dots, a_N^{h^2}(x), \dots\right] .$$

Obviously $\alpha_{(r)}$ is irrational; since h is a power of the characteristic p of the field $\mathbb{K}$, we get:

$$\alpha_{(r)}^h = \left[a_0^h(x), \dots, a_N^h(x), a_0^{h^2}(x), \dots, a_N^{h^2}(x), \dots\right] .$$

Besides, from the formula which connects the elements of $\mathbb{K}\left(\left(X^{-1}\right)\right)$ with their complete quotients (see [4, p.71]) we obtain, as $\alpha^h_{(r)}$ is a complete quotient of $\alpha_{(r)}$:

$$\alpha_{(r)} = \frac{p_N(x)\alpha^h_{(r)} + p_{N-1}(x)}{q_N(x)\alpha^h_{(r)} + q_{N-1}(x)} \ . \tag{9}$$

Therefore $\alpha_{(r)}$ is a root of the equation:

$$q_n(x)y^m - p_N(x)y^{m-1} + q_{N-1}(x)y - p_{N-1}(x) = 0 \ .$$

Unlike what happens for polynomial with coefficients in $\mathbb{Z}$ (cfr. Theorem 3.1), the height of the product of two polynomials with coefficients in $\mathbb{K}[X]$ equals the product of the two heights; the verification is routine. Then:

$$\begin{aligned} H(\alpha_{(r)}) &\leq \max\left\{|q_N(x)|, |p_N(x)|, |q_{N-1}(x)|, |p_{N-1}(x)|\right\} \\ &= |p_N(x)| = H(s_N(x)) \end{aligned}$$

and $2 \leq d(\alpha_{(r)}) \leq m$. It remains to be checked that $\left|\alpha_{(r)} - \xi\right| \leq |r - \xi|$. Consider the Laurent development of $\alpha_{(r)}$, namely $\alpha_{(r)} = \sum_{n \leq k_\alpha} \alpha_n X^n$, $k_\alpha \in \mathbb{Z}$, $\alpha_n \in \mathbb{K}$. From (7) we obtain

$$\begin{aligned} |\xi - r| = e^{-t_0} \ , &\quad t_0 = 2\sum_{i=1}^{N} \deg\left[a_i(x)\right] + \deg\left[a_{N+1}(x)\right] \ , \\ \left|\alpha_{(r)} - r\right| = e^{-t_1} \ , &\quad t_1 = 2\sum_{i=1}^{N} \deg\left[a_i(x)\right] + \deg\left[a^h_0(x)\right] \ . \end{aligned} \tag{10}$$

From (8) it follows $t_0 \leq t_1$. From the Laurent development of ξ, r, $\alpha_{(r)}$, and (10) we have:

$$\begin{aligned} \xi_n = r_n \ \text{ for } \ n > -t_0 \ , &\quad \xi_{-t_0} \neq r_{-t_0} \ , \\ \alpha_n = r_n \ \text{ for } \ n > -t_1 \ , &\quad \alpha_{-t_1} \neq r_{-t_1} \ . \end{aligned} \tag{11}$$

Then, for $n > -t_0 \geq -t_1$ we have $\xi_n = \alpha_n$ and consequently $\left|\xi - \alpha_{(r)}\right| \leq |\xi - r|$, namely what we wanted to prove. Observe that, if $t_0 \neq t_1$, from (11) it follows that:

$$\left|\xi - \alpha_{(r)}\right| = \max\left\{|\xi - r|, \left|\alpha_{(r)} - r\right|\right\} \ .$$

More generally, arguing as above, the following fact can be stated, applying (10) and (11):

REMARK 5.2. *If ξ, ξ' are elements of $\mathbb{K}\left(\left(X^{-1}\right)\right)$ having the same convergent s, then:*

$$|\xi - s| \neq |\xi' - s| \quad \textit{implies} \quad |\xi - \xi'| = \max\{|\xi - s|, |\xi' - s|\} \ ,$$
$$|\xi - s| = |\xi' - s| \quad \textit{implies} \quad |\xi - \xi'| \leq |\xi - s| = |\xi' - s| \ .$$

2) $r = [b_0(x), b_1(x), \ldots, b_k(x)]$, $k > N \geq 0$ with $a_i(x) = b_i(x)$ for $0 \leq i \leq N$ and $a_{N+1}(x) \neq b_{N+1}(x)$. In other words, r is not a convergent of ξ, but r and ξ have in common the convergent $s_N(x)$.

2A) $|a_{N+1}(x)| \neq |b_{N+1}(x)|$.
Define $\alpha(r)$ as in point 1). As $s_N(x)$ is at the same time a convergent of r, ξ and $\alpha(r)$, a repeated application of the above Remark 5.2 yields:

$$|\xi - r| = \max\{|\xi - s_N(x)|, |r - s_N(x)|\} \ ,$$

as $|a_{N+1}(x)| \neq |b_{N+1}(x)|$ by the hypothesis and $\left|\xi - \alpha_{(r)}\right| \leq |\xi - s_N(x)|$ from (8). Thus $\left|\xi - \alpha_{(r)}\right| \leq |\xi - r|$. Of course degree and height of $\alpha_{(r)}$ are like in 1).

2B) $|a_{N+1}(x)| = |b_{N+1}(x)|$.
Consider the following rational functions, together with their Laurent developments:

$$\bar{r} = [a_0(x), \ldots, a_N(x), a_{N+1}(x)] = \frac{p_{N+1}}{q_{N+1}} = \sum_{n \leq k_{\bar{r}}} \bar{r}_n x^n \ ,$$
$$\tilde{r} = [a_0(x), \ldots, a_N(x), b_{N+1}(x)] = \frac{\tilde{p}_{N+1}}{\tilde{q}_{N+1}} = \sum_{n \leq k_{\tilde{r}}} \tilde{r}_n x^n \ ,$$

$\bar{r}_n, \tilde{r}_n \in \mathbb{K}$, $k_{\bar{r}}, k_{\tilde{r}} \in \mathbb{Z}$.
Let's prove the following two inequalities (12) and (13):

$$|\bar{r} - \xi| < |\tilde{r} - \xi| \ . \tag{12}$$

As the hypothesis $|a_{N+1}(x)| = |b_{N+1}(x)|$, then $|q_{N+1}(x)| = |\tilde{q}_{N+1}(x)|$. Consequently

$$|\bar{r} - \xi| = \frac{1}{|a_{N+2}(x)| \ |q_{N+1}(x)|^2} < \frac{1}{|\tilde{q}_{N+1}(x)|^2} \ .$$

If it were, by contradiction, $|\bar{r} - \xi| \geq |\tilde{r} - \xi|$, this would imply $|\tilde{r} - \xi| < \dfrac{1}{|\tilde{q}_{N+1}(x)|^2}$. But (see [12, p.140]) from this inequality it follows that $\tilde{r}$ is a convergent of ξ. This is a contradiction, as $a_{N+1}(x) \neq b_{N+1}(x)$.
In exactly the same way one proves the inequality:

$$|\tilde{r} - r| < |\bar{r} - r| \ . \tag{13}$$

Now we prove:

$$|\bar{r} - \xi| < |r - \xi| \ . \tag{14}$$

Taking into account the coefficients of the Laurent development for ξ, r, $\tilde{r}$, $\bar{r}$ (see above), from (12) we have:

$$\begin{aligned} \bar{r}_n &= \xi_n \ \text{ for } \ n > -\bar{t}_1 \ , \\ \tilde{r}_n &= \xi_n \ \text{ for } \ n > -\bar{t}_0 > -\bar{t}_1 \ , \end{aligned} \tag{15}$$

with $\bar{t}_0$, $\bar{t}_1$ positive integers. This yields $\tilde{r}_{-\bar{t}_0} \neq \xi_{-\bar{t}_0}$, $\bar{r}_{-\bar{t}_0} = \xi_{-\bar{t}_0}$. Now if we had $r_{-\bar{t}_0} = \xi_{-\bar{t}_0}$, this would imply $r_{-\bar{t}_0} = \bar{r}_{-\bar{t}_0}$ and $r_{-\bar{t}_0} \neq \tilde{r}_{-\bar{t}_0}$ from (15) contradicting what we have shown in (13). Thus $r_{-\bar{t}_0} \neq \xi_{-\bar{t}_0}$ proving (14).

If we pose:

$$\alpha_{(r)} = \Big[a_0(x), \ldots, a_{N+1}(x), a_0^h(x), \ldots \\ \ldots, a_{N+1}^h(x), a_0^{h^2}(x), \ldots, a_{N+1}^{h^2}(x), \ldots\Big] \ ,$$

as r is both a convergent of $\alpha_{(r)}$ and of ξ, from (8) and Remark 5.2 at the end of point 1), it follows that $\left|\alpha_{(r)} - \xi\right| \leq |\bar{r} - \xi|$, and finally $\left|\alpha_{(r)} - \xi\right| < |r - \xi|$ from (14). The conditions $H(\alpha_{(r)}) \leq H(r)$ and $d(\alpha_{(r)}) \leq m$ are fulfilled while $\alpha_{(r)}$ is a root of the equation:

$$q_{N+1}(x) y^m - p_{N+1}(x) y^{m-1} + q_N(x) y - p_N(x) = 0 \ ,$$

following the analogous (with respect to $s_{N+1}(x)$ and $s_N(x)$) of formula (9).

3) $r = [b_0(x), \ldots, b_k(x)]$, $k \geq 0$ and $b_0(x) \neq a_0(x)$.
In this case the quadratics suffice to outdo the rationals.

3A) $|a_0(x)| \leq |b_0(x)|$.
We set $\alpha_{(r)} = \left[\overline{a_0(x)}\right]$ and then $|\xi - r| = e^{\deg[a_0(x) - b_0(x)]} \geq 1$, while $\left|\alpha_{(r)} - \xi\right| < 1$. Moreover, as the minimal polynomial of $\alpha_{(r)}$ is:

$$y^2 - a_0(x) y - 1 = 0 \ ,$$

we have:

$$H(\alpha_{(r)}) = e^{\deg[a_0(x)]} \leq e^{\deg[b_0(x) + \cdots + b_k(x)]} = H(r) \ ,$$

and $d(\alpha_{(r)}) = 2 \leq m$.

3B) $|a_0(x)| > |b_0(x)| > 1$.
We pose $\alpha_{(r)} = \left[\overline{b_0(x)}\right]$ and proceed as in 3A).

3C) $|b_0(x)| \leq 1$.
As we have supposed $r \notin \mathbb{K}$, we have $b_0(x) = b \in \mathbb{K}$ and $|b_1(x)| > 1$. Pose $\alpha_{(r)} = \left[b, \overline{b_1(x)}\right]$. In this case (as in 3B)) $\left|\alpha_{(r)} - \xi\right| = |r - \xi|$; it can be checked that $\alpha_{(r)}$ is a root of the equation:

$$y^2 - (2\,b - b_1(x))\,y - b\,b_1(x) + b^2 - 1 = 0\ .$$

Therefore $H(\alpha_{(r)}) \leq |b_1(x)| \leq H(r)$ and $d(\alpha_{(r)}) = 2 \leq m$.

As we have considered all the possible cases for $r \in \mathbb{K}(x)$, leaving out only the constant functions which are finitely many, our proof is complete, under the initial hypothesis $|\xi| > 1$. Before to tackle the conclusive part of the demonstration, we need to point out the following two facts:

(A) Let $f = f_n X^n + f_{n-1} X^{n-1} + \ldots$ and $g = g_n X^n + g_{n-1} X^{n-1} + \ldots$ be two non null elements of $\mathbb{K}\left(\left(X^{-1}\right)\right)$ with the same leading coefficient $f_n = g_n$; $\bar{f} = 1/f = \bar{f}_{-n} X^{-n} + \ldots$, $\bar{g} = 1/g = \bar{g}_{-n} X^{-n} + \ldots$ their inverses, and $k \geq 0$ an integer. Then:

$$f_n = g_n, f_{n-1} = g_{n-1}, \ldots, f_{n-k} = g_{n-k}, f_{n-(k+1)} \neq g_{n-(k+1)}$$

imply

$$\bar{f}_{-n} = \bar{g}_{-n}, \bar{f}_{-n-1} = \bar{g}_{-n-1}, \ldots, \bar{f}_{-n-k} = \bar{g}_{-n-k}, \bar{f}_{-n-(k+1)} \neq \bar{g}_{-n-(k+1)}$$

By induction on k; the property is trivially true for $k = 0$; suppose it holds for all integers h with $0 \leq h \leq k-1$; then, as $-n+i-k > -n-k$ for $i = 1, \ldots, k$, by the inductive hypothesis we get:

$$\bar{f}_{-n-k} = -\frac{\displaystyle\sum_{i=1}^{k} f_{n-i}\,\bar{f}_{-n+i-k}}{f_n} = -\frac{\displaystyle\sum_{i=1}^{k} g_{n-i}\,\bar{g}_{-n+i-k}}{g_n} = \bar{g}_{-n-k}\ .$$

Reversing the roles of the coefficients f_j and $\bar{f}_j$ in the preceding formula yields $\bar{f}_{-n-(k+1)} \neq \bar{g}_{-n-(k+1)}$.

(B) If $\alpha \in \mathbb{K}\left(\left(X^{-1}\right)\right)$ is algebraic and $a \in \mathbb{K}$, then $1/\alpha$ and $(\alpha + a)$ are algebraic of the same degree and height as α.

If α is a root of the equation:

$$p_n(x) y^n + p_{n-1}(x) y^{n-1} + \cdots + p_0(x) = 0\ ,$$

then $1/\alpha$ satisfies:

$$p_0(x) y^n + p_1(x) y^{n-1} + \cdots + p_n(x) = 0\ ,$$

and $(a+\alpha)$ satisfies:

$$p_n(x)(a-y)^n - p_{n-1}(x)(a-y)^{n-1} + \cdots + (-1)^n p_0(x) = 0\ .$$

This prove our assertion. (Note incidentally that, if $\mathbb{F}$ is a number field, $\alpha \in \mathbb{F}$ is algebraic over $\mathbb{Q}$ and a is an integer, then the height of $(a+\alpha)$ depends on a).

Suppose now $|\xi| \leq 1$. Then $a_0(x) = a \in \mathbb{K}$. Consider the complete quotient $\xi' = A_1(x)$. Clearly it is badly approximable and $|\xi'| > 1$; thus it is quasi algebraic for the first part of this theorem. Let $\partial(\xi') = m'$. Given $r = [b_0(x), \dots, b_k(x)]$, $r \in \mathbb{K}(x)$, $r \notin \mathbb{K}$, $k \geq 0$ we are going to define $\alpha_{(r)}$ as follows: if $b_0(x) \neq a$ put $\alpha_{(r)} = \left[\overline{b_0(x)}\right]$ for $|b_0(x)| > 1$, $\alpha_{(r)} = \left[a', \overline{b_1(x)}\right]$ for $b_0(x) = a' \neq a$. Trivially $\left|\alpha_{(r)} - \xi\right| = |r-\xi|$, $H(\alpha_{(r)}) \leq H(r)$ and $d(\alpha_{(r)}) = 2 \leq m'$. Suppose then $r = [a, b_1(x), \dots, b_k(x)]$, $k \geq 1$ as $r \neq \mathbb{K}$. Define $r' = B_1(x)$:

- If r' pertains to cases 1) or 2) (with respect to ξ') of the first part of the theorem, define $\alpha_{(r')}$ accordingly and set $\alpha_{(r)} = \left[a, \alpha_{(r')}\right]$. As ξ', r', $\alpha_{(r')}$ have the same leading term (the leading term of $b_1(x) = a_1(x)$), applying (A) we have that $\left|\alpha_{(r')} - \xi'\right| \leq |r' - \xi'|$ implies $\left|\dfrac{1}{\alpha_{(r')}} - \dfrac{1}{\xi'}\right| \leq \left|\dfrac{1}{r'} - \dfrac{1}{\xi'}\right|$. Thus $\left|\alpha_{(r)} - \xi\right| = \left|\dfrac{1}{\alpha_{(r')}} - \dfrac{1}{\xi'}\right| \leq \left|\dfrac{1}{r'} - \dfrac{1}{\xi'}\right| = |r-\xi|$. From (B) it follows $H(\alpha_{(r)}) = H(r)$.
- r' is in the cases 3A) or 3B) (the case 3C) cannot occur as $|b_1(x)| > 1$). Then define $\alpha_{(r)} = \left[a, \overline{a_1(x)}\right]$ if $|a_1(x)| \leq |b_1(x)|$ or $\alpha_{(r)} = \left[a, \overline{b_1(x)}\right]$ if $|a_1(x)| > |b_1(x)|$ and everything runs as before. This completes the proof. $\square$

The following corollary and the successive example confirm that the results for $\mathbb{K}\left(\left(X^{-1}\right)\right)$ are slightly stronger than the corresponding ones for number fields (cfr. Theorem 3.1).

Corollary 5.3. *In the hypotheses of Theorem 5.1, let* $\xi = [a_0(x), \dots, a_n(x), \dots]$ *such that* $|a_i(x)| \leq |a_0(x)|$ *for any* $i \geq 0$*. Then* ξ *is quasi quadratic.*

Proof. In formula (8) of the theorem, h can be taken equal 1, then $\partial(\xi) = 2$. $\square$

In [1] it is shown the example of a badly approximable cubic $\alpha \in \mathbb{Z}_2\left(\left(X^{-1}\right)\right)$. Let $a_i(x)$ be any partial quotient of maximal degree of α. Then $A_i(x)$ is obviously cubic; moreover it is quasi quadratic from the preceding corollary. Then the quasi algebraic degree of α is strictly smaller than its algebraic degree.

6. Concluding remarks

Apart from metrical properties, the only effective examples of transcendental quasi quadratic numbers that we know at the present time are those which satisfy the hypotheses of Proposition 3.3. Among other possible examples, it would be nice to know whether some of the classical numbers (like e, π, etc.) enjoy this property; however, our main question remains the following:

is every badly approximable real number quasi algebraic?

References

[1] L.E. Baum and M.M. Sweet, *Continued fractions of algebraic power series in characteristic* 2, Ann. of Math. (2) **103** (1976), 593–610.

[2] Y. Bugeaud, *Approximation par des nombres algébriques*, J. Number Theory **84** (2000), 15–33.

[3] Y. Bugeaud, *Approximation by algebraic numbers*, Cambridge Tracts in Mathematics, vol. 160, Cambridge University Press, Cambridge, 2004.

[4] B. de Mathan, *Approximations diophantiennes dans un corps local*, Bull. Soc. Math. France Suppl. Mém. **21** (1970), 93.

[5] R. Güting, *Approximation of algebraic numbers by algebraic numbers*, Michigan Math. J. **8** (1961), 149–159.

[6] A.Ya. Khintchine, *Continued fractions*, P. Noordhoff Ltd., Groningen, 1963, (trans. from 3rd Russian ed. by Peter Wynn).

[7] A.H. Kruse, *Estimates of* $\sum_{k=1}^{N} k^{-s} \langle kx \rangle^{-t}$, Acta Arith. **12** (1966/1967), 229–261.

[8] A. Lasjaunias, *Diophantine approximation and continued fraction expansions of algebraic power series in positive characteristic*, J. Number Theory **65** (1997), 206–225.

[9] K. Mahler, *Lectures on transcendental numbers*, Lecture Notes in Mathematics, Vol. 546, Springer-Verlag, Berlin, 1976.

[10] I. Niven, H.S. Zuckerman, and H.L. Montgomery, *An introduction to the theory of numbers*, fifth ed., John Wiley & Sons Inc., New York, 1991.

[11] M. Queffélec, *Transcendance des fractions continues de Thue-Morse*, J. Number Theory **73** (1998), 201–211.

[12] W.M. Schmidt, *On continued fractions and Diophantine approximation in power series fields*, Acta Arith. **95** (2000), 139–166.

[13] J. Shallit, *Real numbers with bounded partial quotients: a survey*, Enseign. Math. (2) **38** (1992), 151–187.

[14] C.L. Siegel, *Iteration of analytic functions*, Ann. of Math. (2) **43** (1942), 607–612.

[15] C.L. Siegel and J.K. Moser, *Lectures on celestial mechanics*, Springer-Verlag, New York, 1971, Translation by Charles I. Kalme, Die Grundlehren der mathematischen Wissenschaften, Band 187.

[16] R. Talamo, *Quadratic irrational numbers and Koenigs functions*, Rend. Istit. Mat. Univ. Trieste **21** (1989), 51–57.

Authors' addresses:

Maurizio Citterio
Dipartimento di Matematica
Politecnico di Milano
via Bonardi, 9
20133 Milano; Italia
E-mail: `maurizio.citterio@polimi.it`

Rodolfo Talamo
Dipartimento di Matematica
Politecnico di Milano
via Bonardi, 9
20133 Milano; Italia
E-mail: `rodolfo.talamo@polimi.it`

Received September 26, 2013
Revised September 12, 2014

Rend. Istit. Mat. Univ. Trieste
Volume 46 (2014), 203–229

Resolution of the ideal sheaf of a generic union of conics in $\mathbb{P}^3$: I

Olivier Rahavandrainy

Abstract. *We work over an algebraically closed field $\mathcal{K}$ of characteristic zero. Let Y be the generic union of $r \geq 2$ skew conics in $\mathbb{P}^3_{\mathcal{K}}$, $\mathcal{I}_Y$ its ideal sheaf and v the least integer such that $h^0(\mathcal{I}_Y(v)) > 0$. We first establish a conjecture (concerning a maximal rank problem) which allows to compute, by a standard method, the minimal free resolution of $\mathcal{I}_Y$ if $r \geq 5$ and $\dfrac{v(v+2)(v+3)}{12v+2} < r < \dfrac{(v+1)(v+2)(v+3)}{12v+6}$. At the second time, we give the first part of the proof of that conjecture.*

Keywords: Projective space, scheme, sheaf, minimal free resolution
MS Classification 2010: 14N05, 14F05

1. Introduction

We work over an algebraically closed field $\mathcal{K}$ of characteristic zero. We denote by $\mathbb{P}^3$ the projective space $\mathrm{Proj}(\mathcal{K}[x_0, x_1, x_2, x_3])$ of dimension 3 over $\mathcal{K}$, and by $\mathcal{O}$ its structural sheaf.

For $a \in \mathbb{N}, m \in \mathbb{Z}$, and for a coherent sheaf $\mathcal{F}$ on $\mathbb{P}^3$, we put:

$$a\mathcal{O}(m) = \underbrace{\mathcal{O}(m) \oplus \cdots \oplus \mathcal{O}(m)}_{\text{a times}}, \mathcal{F}(m) = \mathcal{F} \otimes \mathcal{O}(m), h^i(\mathcal{F}(m)) = \dim_{\mathcal{K}} H^i(\mathcal{F}(m)).$$

It is well known (Hilbert's syzygies theorem) that the graded $\mathcal{K}[x_0, x_1, x_2, x_3]$-module, $\Gamma_*(\mathcal{F}) = \bigoplus_{n \in \mathbb{Z}} H^0(\mathcal{F}(n))$, has a minimal graded free resolution of length at most 4. After sheafifing, we get a minimal free resolution of $\mathcal{F}$:

$$0 \to \mathcal{E}_4 \to \mathcal{E}_3 \to \mathcal{E}_2 \to \mathcal{E}_1 \to \mathcal{E}_0 \to \mathcal{F} \to 0, \tag{1}$$

where each $\mathcal{E}_j$ is of the form $\bigoplus_{i=1}^{N_j} a_{ij}\ \mathcal{O}(-n_{ij})$, with $N_j, n_{ij}, a_{ij} \in \mathbb{N}$.

However, if one wants to get more information about the N_j's, the n_{ij}'s and the a_{ij}'s, many problems arise, namely the postulation problem (see [2, 3, 9, 10] and references therein). So, one cannot always calculate completely that resolution.

Let v be the least integer such that $h^0(\mathcal{F}(v)) \neq 0$ and consider Conditions (C_1), (C_2) and (C_3) below:

(C_1) $\mathcal{F}$ is $v+1$-regular and $h^0(\mathcal{F}(k)) \cdot h^1(\mathcal{F}(k)) = 0$, for any $k \in \mathbb{Z}$,

(C_2) $h^0(\Omega \otimes \mathcal{F}(k+1)) \cdot h^1(\Omega \otimes \mathcal{F}(k+1)) = 0$, for any $k \in \mathbb{Z}$,

(C_3) $h^0(\Omega^* \otimes \mathcal{F}(k+1)) \cdot h^1(\Omega^* \otimes \mathcal{F}(k+1)) = 0$, for any $k \in \mathbb{Z}$,

where Ω (resp. Ω^*) is the cotangent bundle (resp. the tangent bundle) over $\mathbb{P}^3$.

The following facts (illustrated in Proposition 2.6 for a particular case) are well known:
- If Conditions (C_1) and (C_2) are both satisfied with $h^1(\Omega \otimes \mathcal{F}(v+1)) \neq 0$, then one knows exactly $\mathcal{E}_0$, $\mathcal{E}_1$ and $\mathcal{E}_2$ in (1).
- If (C_1) and (C_2) are satisfied with $h^1(\Omega \otimes \mathcal{F}(v+1)) = 0$, then we need Condition (C_3) to get our target.

In the case where $\mathcal{F}$ is the ideal sheaf of a generic union of r skew lines in $\mathbb{P}^3$, Condition (C_1) holds (see [7]). M. Idà proved ([9]) that Condition (C_2) holds also if $r \neq 4$. We do not know whether Condition (C_3) may be satisfied. So, the minimal free resolution of $\mathcal{F}$ is well known, for infinitely many (but not for all) values of r.

Now, if $\mathcal{F}$ is a general instanton bundle (with Chern classes $c_1 = 0$ and $c_2 > 0$), then (see [6, 13, 14]) Conditions (C_1), (C_2) and (C_3) are all satisfied and we know completely the resolution of $\mathcal{F}$, without exception.

The case of a general stable bundle $\mathcal{F}$ of rank two, on $\mathbb{P}^3$ (with $c_1 = -1$ and $c_2 = 2p \geq 6$), is not yet completely solved: Conditions (C_1) and (C_2) hold (see [6, 15]), but Condition (C_3) is not proved to be true.

In this paper, we are interested in the ideal sheaf $\mathcal{I}_Y$ of the generic union $Y := Y_r$ of r skew conics in $\mathbb{P}^3$, with $r \in \mathbb{N}^*$. E. Ballico showed ([2]) that Condition (C_1) holds if $r \geq 5$. We conjecture that Condition (C_2) would be also satisfied (see Conjecture 1.1) for any $r \in \mathbb{N}^*$, and we will give the first part of its proof.

Note that if $\mathcal{F} = \mathcal{I}_Y$, then (C_2) (resp. (C_3)) means that the natural (restriction) map $r_Y(n) : H^0(\Omega(n)) \to H^0(\Omega(n)_{|Y})$ (resp. $r_Y^*(n) : H^0(\Omega^*(n)) \to H^0(\Omega^*(n)_{|Y})$) has *maximal rank* (i.e., it is injective or surjective). So, we may establish our conjecture as:

CONJECTURE 1.1. *Let Y be the generic union of r skew conics in $\mathbb{P}^3$, $r \in \mathbb{N}^*$, and let Ω be the cotangent bundle on $\mathbb{P}^3$. Then for any integer n, the natural map from $H^0(\Omega(n))$ to $H^0(\Omega(n)_{|Y})$ has maximal rank.*

We remark that (see Theorem 5.2 in [5], p. 228) there exists a positive integer n_0 (depending on Ω and Y) such that $h^1(\Omega(n) \otimes \mathcal{I}_Y) = 0$, for any $n \geq n_0$. Therefore, the restriction map $r_Y(n)$ is always surjective for any such n. We also get: $h^0(\Omega(n) \otimes \mathcal{I}_Y) = h^0(\Omega(n)) = 0$, for any $n \leq 1$. Our Conjecture is then true for $n \notin \{2, \dots, n_0 - 1\}$.

We give in Section 3, the main idea to prove such a maximal rank problem. But before that, we recall (Section 2) the standard method to get the minimal free resolution of $\mathcal{I}_Y$. Section 4 is devoted to notations, definitions and several results which are necessary to our (first part of the) proof in Section 5. Finally, we give in Section 6 some Maple programs which help us for computations.

2. Standard method

We adapt here the standard method to our situation where $\mathcal{F}$ is the ideal sheaf $\mathcal{I}_Y$ of the generic union Y of r skew conics in $\mathbb{P}^3$. In this case, the form of the minimal free resolution of $\mathcal{I}_Y$ is:

$$0 \to \mathcal{E}_2 \to \mathcal{E}_1 \to \mathcal{E}_0 \to \mathcal{I}_Y \to 0 \tag{2}$$

where for $0 \leq j \leq 3$, $\mathcal{E}_j = \bigoplus_{i=1}^{N_j} a_{ij}\ \mathcal{O}(-n_{ij})$, with $N_j, n_{ij}, a_{ij} \in \mathbb{N}$.

We need the two following lemmata.

Lemma 2.1. *i) For any $k \in \mathbb{N}$, one has:*

$$h^0(\mathcal{I}_Y(k)) - h^1(\mathcal{I}_Y(k)) = \binom{k+3}{3} - (2k+1)r,\ h^2(\mathcal{I}_Y(k)) = h^3(\mathcal{I}_Y(k-3)) = 0.$$

ii) If $r \geq 5$ then:

a) $h^0(\mathcal{I}_Y(k)) \cdot h^1(\mathcal{I}_Y(k)) = 0$ *for any* $k \in \mathbb{Z}$,
b) $h^0(\mathcal{I}_Y(k)) = \max(0, \binom{k+3}{3} - (2k+1)r)$ *for any* $k \in \mathbb{Z}$,
c) $v = \min\{m \in \mathbb{N}\ /\ \binom{m+3}{3} - (2m+1)r \geq 1\} \geq 5$,
d) $h^1(\mathcal{I}_Y(v)) = 0,\ h^2(\mathcal{I}_Y(v-1)) = 0$ *and* $h^3(\mathcal{I}_Y(v-2)) = 0$.

Proof. i): consider cohomologies in the exact sequence:

$$0 \to \mathcal{I}_Y(l) \to \mathcal{O}(l) \to \mathcal{O}_Y(l) \to 0,$$

and remark that

$h^2(\mathcal{I}_Y(l)) = h^1(\mathcal{O}_Y(l)) = r \cdot h^1(\mathcal{O}_{\mathbb{P}^1}(2l)) = r \cdot h^0(\mathcal{O}_{\mathbb{P}^1}(-2l-2)) = 0$ if $l \geq 0$,
and $h^3(\mathcal{I}_Y(l)) = h^3(\mathcal{O}(l)) = h^0(\mathcal{O}(-l-4)) = 0$ if $l \geq -3$.

ii): a) is obtained from [2]. Parts b), c) and d) immediately follow. □

Now, put $\mathbb{I}_k = H^0(\mathcal{I}_Y(k))$ and $\mathbb{I} = \bigoplus_{k \geq 0} \mathbb{I}_k$, the homogeneous ideal of Y.

We get by Castelnuovo-Mumford Lemma ([11, p. 99]) and by Lemma 2.1:

Lemma 2.2. *If $r \geq 5$, the sheaf $\mathcal{I}_Y$ is $v+1$-regular, $\mathbb{I}_k = (0)$ if $k < v$ and $\mathbb{I}$ is generated by $\mathbb{I}_v \oplus \mathbb{I}_{v+1}$.*

As consequences, we know more about the minimal free resolution of $\mathcal{I}_Y$, for $r \geq 5$:

Corollary 2.3. (see [14] and [9, Proposition 7.2.1]) *If $r \geq 5$, then the $\mathcal{O}$-modules $\mathcal{E}_0, \mathcal{E}_1, \mathcal{E}_2$ involved in (2) are*

$$\begin{aligned}
\mathcal{E}_0 &= \alpha_1\mathcal{O}(-v) \oplus \beta_1\mathcal{O}(-v-1),\\
\mathcal{E}_1 &= \alpha_2\mathcal{O}(-v-1) \oplus \beta_2\mathcal{O}(-v-2),\\
\mathcal{E}_2 &= \alpha_3\mathcal{O}(-v-2) \oplus \beta_3\mathcal{O}(-v-3),
\end{aligned}$$

where

$$(\star) : \begin{cases}
\alpha_1 = h^0(\mathcal{I}_Y(v)),\\
\beta_1 = h^1(\Omega \otimes \mathcal{I}_Y(v+1)),\\
\alpha_2 = h^0(\Omega \otimes \mathcal{I}_Y(v+1)),\\
\beta_2 = h^1(\Omega^* \otimes \mathcal{I}_Y(v-2)),\\
\alpha_3 = h^0(\Omega^* \otimes \mathcal{I}_Y(v-2)),\\
\beta_3 = h^1(\mathcal{I}_Y(v-1)),\\
\alpha_2 - \beta_1 = 4h^0(\mathcal{I}_Y(v)) - h^0(\mathcal{I}_Y(v+1)),\\
\alpha_3 - \beta_2 = \alpha_2 - \beta_1 - \beta_3 - \alpha_1 + 1, \text{ by considering ranks.}
\end{cases}$$

Corollary 2.4. *We suppose that $r \geq 5$.*

i) If $r_Y(v+1)$ has maximal rank, then $\mathcal{E}_0$ is completely known.

ii) If $r_Y(v+1)$ is injective but not surjective, then $\mathcal{E}_0$, $\mathcal{E}_1$ and $\mathcal{E}_2$ are completely known.

iii) If $r_Y(v+1)$ is surjective and if $r^*_Y(v-2)$ has maximal rank, then $\mathcal{E}_0$, $\mathcal{E}_1$ and $\mathcal{E}_2$ are completely known.

Proof. i): The integers α_1 and β_3 are already known. We see that $r_Y(v+1)$ has maximal rank if and only if $\alpha_2\beta_1 = 0$. We can precise the exact value of β_1 since $\alpha_2 - \beta_1 = 4h^0(\mathcal{I}_Y(v)) - h^0(\mathcal{I}_Y(v+1))$.
ii): In this case, $\beta_1 \neq 0$ and $\alpha_2 = 0$. Thus, by minimality, $\alpha_3 = 0$. We obtain β_2 from $(\star)$ in Corollary 2.3.
iii): We get $\beta_1 = 0$ and $\alpha_3\beta_2 = 0$. Again, $(\star)$ gives the values of α_2, α_3 and β_2. □

Relations between r and v are given by

Lemma 2.5. *i) One has:*

$$\frac{v(v+1)(v+2)}{12v-6} \leq r < \frac{(v+1)(v+2)(v+3)}{12v+6}.$$

ii) If $\alpha_2\ \beta_1 = 0$, then

$$\beta_1 \neq 0 \iff \beta_1 > 0 \iff \frac{v(v+2)(v+3)}{12v+2} < r < \frac{(v+1)(v+2)(v+3)}{12v+6}.$$

Proof. i): One has, from Lemma 2.1:

$$\begin{array}{l} h^0(\mathcal{I}_Y(v)) > 0,\ h^1(\mathcal{I}_Y(v)) = 0,\ h^0(\mathcal{I}_Y(v-1)) = 0 \text{ and } h^1(\mathcal{I}_Y(v-1)) \geq 0, \\ \binom{v+3}{3} - (2v+1)r = h^0(\mathcal{I}_Y(v)) - h^1(\mathcal{I}_Y(v)) = h^0(\mathcal{I}_Y(v)) > 0, \\ \binom{v+2}{3} - (2v-1)r = h^0(\mathcal{I}_Y(v-1)) - h^1(\mathcal{I}_Y(v-1)) = -h^1(\mathcal{I}_Y(v-1)) \leq 0. \end{array}$$

ii): $\beta_1 > 0$ and $\alpha_2 = 0$. Hence we get from $(\star)$ in Corollary 2.3:

$$\begin{aligned} (6v+1)r - \frac{v(v+2)(v+3)}{2} &= \tbinom{v+4}{3} - (2v+3)r - 4\tbinom{v+3}{3} + 4(2v+1)r) \\ &= h^0(\mathcal{I}_Y(v+1)) - 4h^0(\mathcal{I}_Y(v)) \\ &= \beta_1 > 0. \end{aligned}$$

□

Proposition 2.6 follows from Corollary 2.4 and Lemma 2.5.

Proposition 2.6. *Let Y be the generic union of $r \geq 5$ skew conics in $\mathbb{P}^3$. If $r_Y(v+1)$ has maximal rank and if $\frac{v(v+2)(v+3)}{12v+2} < r < \frac{(v+1)(v+2)(v+3)}{12v+6}$, then $\mathcal{I}_Y$ has the following minimal free resolution:*

$$0 \to \beta_3\mathcal{O}(-v-3) \to \beta_2\mathcal{O}(-v-2) \to \alpha_1\mathcal{O}(-v) \oplus \beta_1\mathcal{O}(-v-1) \to \mathcal{I}_Y \to 0,$$

where:

$$\begin{cases} \alpha_1 = \frac{1}{6}(v+1)(v+2)(v+3) - (2v+1)r, \\ \beta_1 = (6v+1)r - \frac{1}{2}v(v+2)(v+3), \\ \beta_2 = (6v-1)r - \frac{1}{2}v(v+1)(v+3), \\ \beta_3 = (2v-1)r - \frac{1}{6}v(v+1)(v+2). \end{cases}$$

Remark 2.7. i) The first twenty values of r and the corresponding values of v, for which Proposition 2.6 holds, are:

r	5	6	9	11	13	15	18	20	23	26
v	5	6	8	9	10	11	12	13	14	15

r	29	32	35	39	42	43	46	47	50	51
v	16	17	18	19	20	20	21	21	22	22

Hence, $v_1 = 5$ is the minimal value of n in Conjecture 1.1, that we shall consider. However, it is natural to treat also the case $n \leq 4$ (see Section 5.4).

ii) If $r \in \{2, 3, 4\}$, then $\mathcal{I}_Y$ does not satisfy Condition (C_1) in Section 1 (see [2]). So we cannot apply Proposition 2.6. In that case, the minimal free resolution of $\mathcal{I}_Y$ would be obtained by direct (but delicate) computations. We will do it in the future.

3. How to prove Conjecture 1.1?

A maximal rank problem (depending on a natural number n) can be proved by using the so called *Horace method* (see Section 3.1 and [8]). It is an induction proof (on n) where each step requires more or less sophisticated conditions (equations and inequations satisfied by many integers), called *adjusting conditions* (see e.g. the hypotheses of Proposition 4.12). If n is sufficiently large, then those conditions are not difficult to realize, whereas for "small" values of n, one must verify them case by case: the *initial cases*. A priori, for each n (large or not), many complicated calculations arise (see e.g. [9] or [13]). So we often use Maple computations.

3.1. The Horace method (see [8])

We omit here to recall the notion of specialization of a subscheme (see e.g. [15, Section 3.1]).
Let E be a bundle on a quasi-projective scheme T and let Z be a subscheme of T. We consider the restriction map $\rho : H^0(E) \to H^0(E_{|Z})$. We say that:
- Z is numerically E-settled if $h^0(E) = h^0(E_{|Z})$,
- Z is E-settled if ρ has maximal rank.
If Δ is a Cartier divisor on T and Z_s is a specialization of Z, then we put:
$Z'' = Z_s \cap \Delta$ (*trace* of Z_s on Δ),
$Z' = res_\Delta Z_s$ (residual scheme: scheme such that its ideal sheaf is the kernel of the natural morphism : $\mathcal{O} \to \mathcal{H}om(\mathcal{I}_{Z_s}, \mathcal{O}_\Delta)$).

From the residual exact sequence (cf. [8, p. 353]):

$$0 \to I_{Z'}(-\Delta) \to I_Z \to I_{Z'',\Delta} \to 0,$$

we get the following lemmata.

Lemma 3.1. *If Z_s is numerically E-settled, then, Z' is numerically $E(-\Delta)$-settled if and only if Z'' is numerically $E_{|\Delta}$-settled.*
*In this case, we say that Z_s is a (E, Δ)-**adjusted specialization of** Z.*

Lemma 3.2. *Let i be a natural number. If $h^i(E(-\Delta) \otimes I_{Z'}) = 0$ (condition called dègue) and if $h^i(E \otimes I_{Z'',\Delta}) = 0$ (dîme), then $h^i(E \otimes I_Z) = 0$.*

Remark 3.3. i) We call *adjusting conditions*, the conditions for which, the specialization Z_s of Z is numerically E-settled.

ii) We say that one *exploits* a divisor if one applies the Horace method with it.

iii) Again, to prove the dègue and the dîme, we may apply the Horace method and so on... It leads, after a finite number of steps, to simpler statements, because for each "dègue", the bundle degree decreases, and for each "dîme", the subscheme dimension decreases.

3.2. A first step of the proof

Conjecture 1.1 says that, for any integer n, the natural map $r_Y(n)$ from $H^0(\Omega(n))$ to $H^0(\Omega(n)_{|Y})$ has maximal rank, Ω being the cotangent bundle over $\mathbb{P}^3$.

As mentioned at the end of Section 1, the map $r_Y(n)$ is injective if $n \leq 1$ and it is surjective if $n \geq n_0$, for some $n_0 \in \mathbb{N}^*$. It remains then the case: $2 \leq n \leq n_0 - 1$. For $2 \leq n \leq 4$, see Section 5.4.1. Now, we suppose that $n \geq 5$. We would like to apply exactly the idea described in [9]. We put:

$$\mathcal{X}^* = \mathbb{P}(\Omega),\ L_n = \mathcal{O}_{\mathcal{X}^*}(1) \otimes \pi^*\mathcal{O}(n),\ Y^* = \pi^{-1}(Y),$$
where $\pi : \mathcal{X}^* \to \mathbb{P}^3$ is the canonical projection.

We remark that L_n is a bundle of rank 1, so we may define (Section 5.1) a subscheme $T^*(n)$, not depending on r, contained in Y^* or containing Y^*, such that $h^0(L_n) = h^0(L_{n|T^*(n)})$ (see [9, Section 1.1]).

Let $\rho_n : H^0(L_n) \to H^0(L_{n|T^*(n)})$ be the restriction map. If ρ_n is bijective and if $Y^* \subset T^*(n)$ (resp. $Y^* \supset T^*(n)$), then $r_Y(n)$ is surjective (resp. injective). So, we get Conjecture 1.1. The bijectivity of ρ_n is equivalent to $H(n)$: $H^0(L_n \otimes \mathcal{I}_{T^*(n)}) = 0$, where $\mathcal{I}_{T^*(n)}$ is the ideal sheaf of $T^*(n)$.

The equality $H(n)$ is proved by using the *Horace method*. For that, we build another subscheme $T'^*(n)$ of $\mathcal{X}^*$ such that $h^0(L_n) = h^0(L_{n|T'^*(n)})$, in such a manner that if the natural map $\rho'_{n-2} : H^0(L_{n-2}) \to H^0(L_{n-2|T'^*(n-2)})$ is bijective, then we get $H(n)$. We remark also that the bijectivity of ρ'_n is equivalent to $H'(n)$: $H^0(L_n \otimes \mathcal{I}_{T'^*(n)}) = 0$, and $H'(n)$ may be proved by the Horace method, and so on...

In Section 5, we define the schemes $T^*(n)$ and $T'^*(n-2)$ and we prove the implication: $H'(n-2) \Rightarrow H(n)$ for any $n \geq 5$. Unfortunately, contrary to what happened in [9] and [15], the statement $H'(n)$ is more difficult to prove because the adjusting conditions are more complicated. We shall try to look more carefully at this situation, in a forthcoming paper, in order to complete the proof of this Conjecture.

4. Preliminary results

In the rest of the paper, Q denotes a smooth quadric surface in $\mathbb{P}^3$, Ω the cotangent bundle over $\mathbb{P}^3$, $\overline{\Omega}$ the restriction of Ω on Q, $\mathcal{X}^* = \mathbb{P}(\Omega)$.
$\pi : \mathcal{X}^* \to \mathbb{P}^3$, $p_1, p_2 : Q \cong \mathbb{P}^1 \times \mathbb{P}^1 \ \to \ \mathbb{P}^1$ are the canonical projections.
We put: $C^* = \pi^{-1}(C)$ for a subscheme C of $\mathbb{P}^3$, and for two integers a and b:

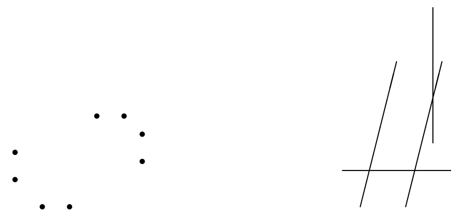
$$\begin{aligned} &\mathcal{O}_Q(a,b) = p_1^*\mathcal{O}_{\mathbb{P}^1}(a) \otimes p_2^*\mathcal{O}_{\mathbb{P}^1}(b),\ \overline{\Omega}(a,b) = \Omega \otimes \mathcal{O}_Q(a,b),\ \overline{\Omega}(a) = \overline{\Omega}(a,a),\\ &K_{a,b} = \mathcal{O}_{Q^*}(1) \otimes \pi^*\mathcal{O}_Q(a,b),\ K_a = K_{a,a}. \end{aligned}$$

We denote by $[\dfrac{a}{b}]$ the quotient (by Euclidean division) of a by b, and by $\{\dfrac{a}{b}\}$ the remainder.

4.1. Definitions (see [1] and [9])

- A s-point is a point of $\mathcal{X}^*$.
- A d-point represents two s-points lying in a same fiber $\pi^{-1}(x)$, $x \in \mathbb{P}^3$.
- A t-point (resp. t-curve) represents three non-collinear points lying in a same fiber $\pi^{-1}(x)$ (resp. inverse image of a curve in $\mathbb{P}^3$, under π).
- A grille of type (p,q) is a set of pq points of Q, which are the intersection of p lines of type $(1,0)$ and q lines of type $(0,1)$.
- A *four-point* is a set of 4 points, $[P] = \{P_1, \ldots, P_4\} \subset Q$, such that $P_1, P_2 \in \ell \setminus \ell'$ and $P_3, P_4 \in \ell' \setminus \ell$, for some lines $\ell, \ell' \subset Q$ of type $(1,0)$ and $(0,1)$. In other words, $P_1, \ldots, P_4$ are cocyclic but 3 by 3 non collinear.
For example, the intersection of Q with a degenerate conic transverse to Q, such that the singular point does not lie on Q, is a four-point.
- A *bamboo* (see [2]) is a union of 4 lines $L_1, \ldots, L_4$ such that: $L_i \cap L_j \neq \emptyset$ if and only if $|i-j| \leq 1$.
- the first infinitesimal neighborhood of a point x in $\mathbb{P}^3$, denoted by $\xi(x)$, is the subscheme of $\mathbb{P}^3$, having ${\mathcal{I}_x}^2$ as ideal sheaf.
- A triple-point (resp. double-point) is a subscheme of $\mathbb{P}^3$, supported by a point having ideal locally defined by $(x_1, x_2)^2$ (resp. by (x_1^2, x_2)) in $\mathcal{K}[x_1, x_2]$.
For example, $\xi(x) \cap Q$ is a triple-point of Q if $x \in Q$.
- A t-first infinitesimal neighborhood (resp. a t-grille) is the inverse image of a first infinitesimal neighborhood (resp. of a grille), under π.

- We say that t-points, d-points, and s-points are *collinear* (resp. *cocyclic*) in Q^* if their projections on Q lie on the same line (resp. same conic).



two four-points a bamboo

4.2. Examples of specialization

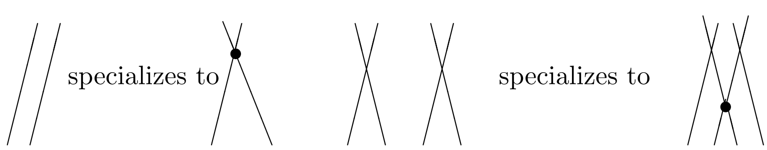
We give some specializations, traces and residual schemes which are useful in Sections 4.3, 5.2 and 5.4 (see also [15, Section 4.4]).

Lemma 4.1. *i) The trace* (*resp. residual scheme*) *of a finite union of subschemes equals the union of traces* (*resp. of residual schemes*).

ii) If ℓ and ℓ' are two lines in Q, intersecting at the point x, then $\ell \cup \ell' \cup \xi(x)$ is a specialization of two skew lines in $\mathbb{P}^3$. Moreover, the residual scheme $res_Q(\ell \cup \ell' \cup \xi(x))$ equals $\{x\}$.

iii) If $[L] = (L_1, \ldots, L_4)$ is a bamboo and if $\{x\} = L_2 \cap L_3$, then the union $[L] \cup \xi(x)$ is a specialization of two skew conics in $\mathbb{P}^3$.

Proof. i): see [8], 4.4. ii): see [7], 2.1.1. iii) follows from i) and ii). □



two skew lines 2 skew singular conics

Lemma 4.2. (see [15, Lemme 4.2])

i) If ℓ is a line and if $x \in \ell$, then $\ell \cap \xi(x)$ is a double-point and $res_\ell(\xi(x))$ is the (*simple*) *point x.*

ii) If C is a rational curve of type $(1,2)$ on Q and if $x \in C$, then $C \cap \xi(x)$ is a double-point and $res_C(\xi(x))$ is the (*simple*) *point x.*

4.3. Lemmata on the quadric Q

First, we recall some general results which we can apply in Lemma 4.10 and in Proposition 4.12. Let E be a bundle on a quasi-projective scheme T and let Z be a subscheme of T. We denote by $\pi : \mathbb{P}(E) \to T$ the canonical projection. For a subscheme W of $\mathbb{P}(E)$, let $\pi(W)$ be the subscheme (of T), of ideal sheaf ${\pi^{\#}}^{-1}(\pi_* I_W)$, where $\pi^{\#}$ is the canonical morphism from $\mathcal{O}_T$ to $\pi_*\mathcal{O}_{\mathbb{P}(E)}$.

Lemma 4.3. *One has:*

i) $\pi^{-1}(Z) \cong \mathbb{P}(E_{|Z})$, $\mathcal{O}_{\mathbb{P}(E)}(1)_{|\pi^{-1}(Z)} \cong \mathcal{O}_{\mathbb{P}(E_{|Z})}(1)$, $\pi_*(\mathcal{O}_{\mathbb{P}(E)}(1)) \cong E$,

ii) $\pi^* I_Z \cong I_{\pi^{-1}Z}$,

iii) if W *is a subscheme of* $\mathbb{P}(E)$*, then:* $\pi_* I_W \cong I_{\pi(W)}$.

Proof. i): see [9, p. 21] and [5, Proposition 7.11, p.162].
ii): π has smooth fibers so the functor π^* is exact. Thus, it suffices to apply it, on the exact sequence: $0 \to I_Z \to \mathcal{O}_T \to \mathcal{O}_Z \to 0$ and to consider the exact sequence: $0 \to I_{\pi^{-1}Z} \to \mathcal{O}_{\mathbb{P}(E)} \to \mathcal{O}_{\pi^{-1}Z} \to 0$.
iii): since π is proper and has connected fibers, $\pi^{\#}$ is an isomorphism. Therefore, $\pi_* I_W \cong {\pi^{\#}}^{-1}(\pi_* I_W) = I_{\pi(W)}$. □

Corollary 4.4. *One has, for* $n, a, b \in \mathbb{N}^*$ *and for any subscheme* C *of* Q*:*

$$h^0(L_n) = h^0(\Omega(n)),\ h^0(K_{a,b}) = h^0(\overline{\Omega}(a,b)),\ h^0(K_{a,b|\pi^{-1}(C)}) = h^0(\overline{\Omega}(a,b)_{|C}).$$

Proof. The projection formula (see [5], p. 124) and Lemma 4.3 give:

$$h^0(L_n) = h^0(\pi_*(L_n)) = h^0(\pi_*(\mathcal{O}_{\mathcal{X}^*}(1) \otimes \pi^*\mathcal{O}(n))) = h^0(\Omega(n)).$$

Similarly, we get: $h^0(K_{a,b}) = h^0(\overline{\Omega}(a,b))$ and $h^0(K_{a,b|\pi^{-1}(C)}) = h^0(\overline{\Omega}(a,b)_{|C})$. □

Lemma 4.5. (see [12, p. 8], [9, Section 3-1] and [4])
Let $n, a, b \in \mathbb{N}^*$ *and let* C *be a rational curve, of type* $(1, n)$ *on* Q*. Then*

i) $h^0(\Omega(n)) = \dfrac{(n^2-1)(n+2)}{2}$, $h^0(K_{a,b}) = h^0(\overline{\Omega}(a,b)) = 3ab - a - b - 1$.

ii) $\overline{\Omega}(a,b)_{|C} \cong 2\mathcal{O}_{\mathbb{P}^1}((a-1)n + b - 2) \oplus \mathcal{O}_{\mathbb{P}^1}((a-2)n + b)$.

Lemma 4.6. *i) If* H *is a plane in* $\mathbb{P}^3$*, then* $\Omega_{|H} \cong \Omega_{\mathbb{P}^2} \oplus \mathcal{O}_{\mathbb{P}^2}(-1)$.

ii) If D *is a line in* $\mathbb{P}^3$*, then* $\Omega_{|D} \cong \overline{\Omega}_{|D} \cong \mathcal{O}_{\mathbb{P}^1}(-2) \oplus 2\mathcal{O}_{\mathbb{P}^1}(-1)$.

Lemma 4.7. *Let $n, a, b, \tau, \delta, \epsilon \in \mathbb{N}$ and let D be a line on Q. Then*

i) $h^0(K_{a,b|D^*}) = h^0(\overline{\Omega}(a,b)_{|D}) = h^0(\overline{\Omega}_{|D}(b)) = 3b - 1$ *if D is of type $(1,0)$.*
ii) $h^0(K_{a,b|D^*}) = h^0(\overline{\Omega}(a,b)_{|D}) = h^0(\overline{\Omega}_{|D}(a)) = 3a - 1$ *if D is of type $(0,1)$.*
iii) *If $S^* \subset \mathcal{X}^*$ is a union of τ t-points, δ d-point and ϵ s-point, then* $h^0(L_{n|S^*}) = 3\tau + 2\delta + \epsilon = h^0(K_{a,b|S^*})$.

Lemma 4.8. *Let $a, b, a', b' \in \mathbb{N}$ and let C, C' be two distinct curves on Q, of type (a,b) and (a',b'). Then*

i) $\#(C \cap C') = ab' + a'b$.

ii) $\#(C \cap [P]) \leq 3,\ \#(C' \cap [P]) \leq 4,\ \#(\ell \cap [P]) \leq 2$, *if C is of type $(1,2)$, C' of type $(1,1)$, ℓ of type $(1,0)$ and $[P]$ a four-point.*

Corollary 4.4 and Lemma 4.5 imply:

Corollary 4.9. *Let $n, a, b \in \mathbb{N}^*$ and let C be a rational curve on Q. Then*

i) $h^0(K_{a,b|C^*}) = h^0(\overline{\Omega}(a,b)_{|C}) = 3(a+b-2)+1$ *if C is of type $(1,1)$.*
ii) $h^0(K_{a,b|C^*}) = h^0(\overline{\Omega}(a,b)_{|C}) = 3(2a+b-3)$ *if C is of type $(1,2)$.*
iii) $h^0(K_{a,b|C^*}) = h^0(\overline{\Omega}(a,b)_{|C}) = 3(2b+a-3)$ *if C is of type $(2,1)$.*

Lemma 4.10. *In one of the following cases, Z is (numerically) E-settled.*

a) $E = K_{a,b|C^*}$, *where C is a conic in Q and Z the generic union, in C^*, of $a+b-2$ t-points (counted with multiplicity) and one s-point;*

b) $E = K_{a,b|\ell^*}$, *where ℓ is a line in Q, of type $(1,0)$ and Z the generic union, in ℓ^*, of $b-1$ t-points (counted with multiplicity) and one d-point;*

c) $E = K_1$, $Z = \emptyset$;

d) $E = K_2$, *Z: a generic union of 2 t-points and 1 s-point;*

e) $E = K_3$, *Z: a generic union of 6 t-points and 1 d-point, such that at most 3 are cocyclic with the d-point.*

Proof. We see, from Lemma 4.7 and Corollary 4.9, that for each case, $h^0(E) = h^0(E_{|Z})$. So, Z is numerically E-settled. It remains to show that $H^0(E \otimes I_Z) = 0$.
a) and b): see [9, p. 23-24].
c): it follows from the fact: $h^0(E) = h^0(\overline{\Omega}(1,1)) = 0$.
d): Z specializes to a union of 2 t-points and 1 s-point lying on a t-conic C^*. We exploit $\Delta = C^*$. The residual scheme Z' is the empty scheme and $E(-\Delta) = K_1$. Hence we get the dègue from c). The trace Z'' consists of 2 t-points and 1 s-point. Moreover, we get by Lemma 4.3: $E_{|\Delta} \cong K_{2|C^*}$. Thus,

the dîme follows from a).
e): We may get a specialization of Z by putting the 6 t-points on a t-curve C^* (C of type $(1,2)$ on Q). We exploit C^*. The trace Z'' consists of 6 t-points and $E_{|C^*} \cong K_{3|C^*} \cong 3\mathcal{O}_{\mathbb{P}^1}(5)$. So the dîme is true.

The residual scheme Z' consists of 1 d-point and $E(-C^*) = K_{2,1}$. The dègue follows from b) and c), by exploiting a t-line ℓ^* of type $(1,0)$ passing through the d-point. □

Remark 4.11. Proposition 4.12 is crucial in the proof of the statement in Section 5: $H'(n-2) \Rightarrow H(n)$. The following notations will be useful to show it. For $f, h, i, \ell \in \mathbb{N}$ such that $1 \leq i \leq 3$, $f = i + 3\ell$, set:

$$a = f + h,\ \ b = f + 2h,\ \ \mu_{\max}(a,b) = \begin{cases} a+b-3 \text{ if } a+b \equiv 0 \mod 3 \\ a+b-2 \text{ if } a+b \not\equiv 0 \mod 3 \end{cases}$$

$$V_{\max}(f,h) = \sum_{k=1}^{\ell}(v_{1k}(f) + v_{2k}(f)) + \sum_{k=1}^{h} v_k^*(f,h),$$

$$M_{\max}(f,h) = \sum_{k=1}^{\ell}(m_{1k}(f) + m_{2k}(f)) + \sum_{k=1}^{h} m_k^*(f,h).$$

The choice of the integers $v_{1k}(f), m_{1k}(f), v_{2k}(f), m_{2k}(f), \ldots$ will allow us to exploit t-rational curves of type $(1,2)$ and $(2,1)$. We give below their different values.

- Case $f + h \leq 3$: $V_{\max}(f,h) = M_{\max}(f,h) = 0$.
- Case $4 \leq f \leq 6$ and $h = 0$: $v_{11} = i - 1, m_{11} = i, v_{21} = m_{21} = 0$.
- Case $f \geq 7$ and $h = 0$: $v_{11} = i - 1, m_{11} = i, v_{21} = m_{21} = 0$, and for $2 \leq k \leq \ell$, $v_{1k} = i + 3k - 4, m_{1k} = 3, v_{2k} = i + 3k - 6, m_{2k} = 3$, $V_{\max}(f,0) = i - 1 + (\ell-1)(f+i-4)$, $M_{\max}(f,0) = i + 6(\ell-1) = 2f - i - 6$.
- Case $f = 1$ and $h \geq 3$: $v_k^* = 0$, $m_k^* = 2k - 2$ for $1 \leq k \leq h$, $V_{\max}(1,h) = 0$, $M_{\max}(1,h) = h(h-1)$.
- Case ($f \geq 2$ and $h \geq 2$) or ($f \geq 3$ and $h = 1$): $v_k^* = f - 2$, $m_k^* = 2k$ for $1 \leq k \leq h$, $V_{\max}(f,h) = V_{\max}(f,0) + (f-2)h$, $M_{\max}(f,h) = M_{\max}(f,0) + h(h+1)$.

Proposition 4.12. *Let $f, h, a, b, v, m, u, \mu, \delta, \epsilon \in \mathbb{N}$ such that*

$$\begin{cases} 1 \leq f \leq a = f + h \leq b = f + 2h < 2a, \\ \delta + \epsilon \leq 1, \\ v \leq V_{\max}(f,h), \\ m \leq M_{\max}(f,h), \\ \mu \leq \mu_{\max}(a,b), \\ 3ab - a - b - 1 = 12v + 9m + 3u + 3\mu + 2\delta + \epsilon. \end{cases}$$

We consider the generic union $F(a,b) \subset Q^*$ of m t-triple-points, v t-four-points, u t-points and μ t-points, δ d-point et ϵ s-point which are cocyclic. Then $F(a,b)$ is $K_{a,b}$-settled.

Proof. By construction, $F(a,b)$ is numerically $K_{a,b}$-settled:

$$h^0(K_{a,b}) = 3ab - a - b - 1 = 12v + 9m + 3u + 3\mu + 2\delta + \epsilon = h^0(K_{a,b|F(a,b)}).$$

The proof is similar to that of Lemma 3.3.1 in [9]. Denote by $R(f,h)$ the statement: "the scheme $F(f+h, f+2h)$ is $K_{f+h,f+2h}$-settled". The main idea is as follows.

If $h \geq 1$, then pass from $R(f,h)$ to $R(f,0)$ by exploiting h times, a t-rational curve of type $(1,2)$. In other words, prove $R(f,k)$ by induction on k, for $0 \leq k \leq h$.

Now, we have to prove $R(f,0)$. If $f \geq 4$, then set $f = i + 3\ell$ where $\ell = [\frac{f-1}{3}]$, $i = f - 3\ell \in \{1,2,3\}$. Pass from $R(f,0)$ to $R(i,0)$ by exploiting alternately ℓ times, two t-rational curves of types $(1,2)$ and $(2,1)$. Here, we also use an inductive proof.

• *Proof of $R(i,0)$ (case $a = b = f = i \in \{1,2,3\}$):*
One has: $M_{\max}(i,0) = V_{\max}(i,0) = 0$ and thus $m = v = 0$.
- The case $i = 1$ follows from Lemma 4.10-c).
- If $i = 2$, then $\epsilon = 1, \delta = 0$, $u + \mu = 2$. So, we may suppose that $u = 2$ and $\mu = 0$. Lemma 4.10-d) gives our result.
- If $i = 3$, then $\epsilon = 0, \delta = 1, u + \mu = 6$, $\mu \leq \mu_{\max}(3,3) = 3$. $R(3,0)$ is true by Lemma 4.10-e).

• *Proof of $R(f,0)$, $f = i + 3\ell \geq 4$ (case $a = b = f \geq 4$):*
We denote by $\tilde{R}(i,k)$ the statement $R(i+3k, 0)$, for $0 \leq k \leq \ell$. We prove it by induction on k. The case $k = 0$ corresponds to $f \in \{1,2,3\}$ and is just treated. We refer to Notations in Remark 4.11.

We suppose that $k \geq 1$ and $\tilde{R}(i, k-1)$ is true. We denote by C' (resp. by Γ) a rational curve on Q, of type $(1,2)$ (resp. the conic passing through the cocyclic t-points). Put $\tilde{f} = \tilde{f}_k = f - 3(\ell - k) = i + 3k$. We take $\mu_1 = \min(\mu, 3)$, $v_1 = \min(v, v_{1k}(f))$ and $m_1 = \min(m, m_{1k}(f))$. Let $u_1 \in \mathbb{N}$ such that $u_1 \leq u$ and

$$3v_1 + 2m_1 + u_1 + \mu_1 = 3\tilde{f} - 3. \tag{3}$$

We define the two following subschemes of Q^*, $\tilde{F}_1$ and $\tilde{F}_2$ as follows.

$\tilde{F}_1$ is the union of $v - v_1$ t-four-points, $m - m_1$ t-triple-points, $u - u_1 + v_1$ t-points and $\mu - \mu_1$ t-points, δ d-point and ϵ s-point which are cocyclic.

$\tilde{F}_2$ consists of $u_1 + 3v_1$ t-points lying on C'^*, μ_1 t-points on $C'^* \cap \Gamma^*$ and the t-infinitesimal neighborhoods of m_1 points on C'.

The two subschemes $F(\tilde{f}, \tilde{f})$ and $\tilde{F}_1 \cup \tilde{F}_2$ have the same number of t-triple-points: $m = (m - m_1) + m_1$. Moreover, the v_1 t-points of $\tilde{F}_1$ together with the $3v_1$ t-points of $\tilde{F}_2$ form a specialization of v_1 t-four-points of $F(\tilde{f},\tilde{f})$. It follows that $F(\tilde{f},\tilde{f})$ generalizes $\tilde{F}_1 \cup \tilde{F}_2$.

We exploit C'^*. The trace Z'' consists of m_1 t-double-points and $3v_1+u_1+\mu_1$ t-points. Moreover, one has: $\pi_*(K_{\tilde{f}|C'^*}) \cong 3\mathcal{O}_{\mathbb{P}^1}(3\tilde{f}-4)$. By Corollary 4.9 and Equality (3), Z'' is numerically $K_{\tilde{f}|C'^*}$-settled:

$$h^0(K_{\tilde{f}|C'^*}) = h^0(\pi_*(K_{\tilde{f}|C'^*})) = 3(3\tilde{f}-3) = 9v_1+6m_1+3u_1+3\mu_1 = h^0(K_{\tilde{f}|Z''}).$$

Hence, we get the dîme.

Now, we prove the dègue. One has $K_{\tilde{f}}(-C'^*) \cong K_{\tilde{f}-1,\tilde{f}-2}$. The residual scheme Z' is exactly the disjoint union of $\tilde{F}_1$ with m_1 t-points. By Lemma 3.1, it is numerically $K_{\tilde{f}-1,\tilde{f}-2}$-settled:

$$3a'b'-a'-b'-1 = 12(v-v_1)+9(m-m_1)+3u^*+3(\mu-\mu_1)+2\delta+\epsilon,$$

where $a'=\tilde{f}-1,\ b'=\tilde{f}-2,\ u^*=u-u_1+m_1+v_1$. Take

$$\mu_2 = \min(\mu-\mu_1,3),\ v_2=\min(v-v_1,v_{2k}(f)),\ m_2=\min(m-m_1,m_{2k}(f)).$$

Let $u_2 \in \mathbb{N}$ such that $u_2 \le u^*$ and

$$3v_2+2m_2+u_2+\mu_2 = 3\tilde{f}-8. \tag{4}$$

Consider a rational curve C'' of type $(2,1)$ on Q. As above, Z' may specialize to the disjoint union of $F(\tilde{f}-3,\tilde{f}-3)$ with u_2+3v_2 t-points lying on C''^*, with μ_2 t-points on $C''^* \cap \Gamma^*$ and with the t-infinitesimal neighborhood of m_2 points on C''. We exploit C''^*. The trace consists of m_2 t-double-points and $3v_2+u_2+\mu_2$ t-points. Since $\pi_*(K_{\tilde{f}-1,\tilde{f}-2|C''^*}) \cong 3\mathcal{O}_{\mathbb{P}^1}(3\tilde{f}-9)$, Equality (4) implies the dîme.

The residual scheme is $F(\tilde{f}-3,\tilde{f}-3)$. By Lemma 3.1, it is numerically $K_{\tilde{f}-3,\tilde{f}-3}$-settled: $3\tilde{a}\tilde{b}-\tilde{a}-\tilde{b}-1 = 12\tilde{v}+9\tilde{m}+3\tilde{u}+3\tilde{\mu}+2\delta+\epsilon$, where

$$\begin{aligned}
&\tilde{v} = v-v_1-v_2 = \max(0, v-v_{1k}(f)-v_{2k}(f)) \le V_{\max}(\tilde{f}-3,0),\\
&\tilde{a}=\tilde{b}=\tilde{f}-3,\ \tilde{m}=m-m_1-m_2 \le M_{\max}(\tilde{f}-3,0),\\
&\tilde{\mu} = \mu-\mu_1-\mu_2 = \max(0,\mu-6) \le \mu_{\max}(\tilde{f},\tilde{f})-6 = \mu_{\max}(\tilde{f}-3,\tilde{f}-3),\\
&\tilde{u} = u^*-u_2+m_2+v_2 = u-u_1-u_2+m_1+v_1+m_2+v_2.
\end{aligned}$$

Therefore, $\tilde{a},\tilde{b},\tilde{m},\tilde{v},\tilde{u},\tilde{\mu},\delta$ and ϵ satisfy all the hypotheses of Proposition 4.12. The dègue is the statement $\tilde{R}(i,k-1)$. It is true by inductive assumption.

- *Proof of $R(f,h)$ (the general case):*

We necessarily have: $2 \le a < b \le 2a-1$. We recall that $a=f+h,\ b=f+2h$ where $f=2a-b \ge 1,\ h=b-a \ge 1$, and $R(f,k)$ is the statement:

"$F(f+k, f+2k)$ is $K_{f+k,f+2k}$-settled". We prove it by induction on k, for $0 \le k \le h$. The proof is similar to the previous one.

The case $k=0$ corresponds to $a=b=f$ and has been already done. We suppose that $k \ge 1$ and $R(f, k-1)$ is true. We denote by C' a rational curve on Q, of type $(1,2)$. Set

$$\mu_1 = \min(\mu, 3),\ v_1 = \min(v, v_k^*(f,h)),\ m_1 = \min(m, m_k^*(f,h)).$$

Let $u_1 \in \mathbb{N}$ such that $u_1 \le u$ and

$$3v_1 + 2m_1 + u_1 + \mu_1 = 3f + 4k - 3. \tag{5}$$

We consider the disjoint union $\tilde{F}$ of $F(f+k-1, f+2k-2)$ with $u_1 + 3v_1 + \mu_1$ t-points lying on C'^* and the t-infinitesimal neighborhoods of m_1 points on C'. We see that $\tilde{F}$ is a specialization of $F(f+k, f+2k)$. We exploit C'^*. The trace Z'' consists of m_1 t-double-points and $u_1 + 3v_1 + \mu_1$ t-points. Corollary 4.9 and Equality (5) give:

$$h^0(K_{f+k,f+2k|C'^*}) = 3(3f+4k-3) = 9v_1+6m_1+3u_1+3\mu_1 = h^0(K_{f+k,f+2k|Z''}).$$

Hence, Z'' is numerically $K_{f+k,f+2k|C'^*}$-settled and we get the dîme.

The residual scheme Z' is exactly $F(f+k-1, f+2k-2)$, $K_{f+k,f+2k}(-C'^*)$ is isomorphic to $K_{f+k-1,f+2k-2}$. Again, from Lemma 3.1, Z' is numerically $K_{f+k-1,f+2k-2}$-settled. As before, we see that all the hypotheses of Proposition 4.12 are satisfied. The dègue is then true, by inductive assumption. □

Corollary 4.13. *We consider the subscheme $F(a,b)$ of Proposition 4.12. Let $c, d_1, d_2, n \in \mathbb{N}^*$ and let G be the union, in Q^*, of c t-conics, d_1 t-lines of type $(1,0)$ and d_2 t-lines of type $(0,1)$, such that $G \cap F(a,b) = \emptyset$. We suppose that $J = G \cup F(a,b)$ is numerically K_n-settled and $a+c+d_1 = b+c+d_2 = n$. Then J is K_n-settled.*

Proof. Since any conic on Q is of type $(1,1)$, we see that the ideal sheaf $\mathcal{I}_G$ of G is isomorphic to $\pi^*\mathcal{O}_Q(-c-d_1, -c-d_2)$. Hence, we get: $H^0(K_n \otimes \mathcal{I}_J) = H^0(K_{a,b} \otimes \mathcal{I}_{F(a,b)}) = 0$ by Proposition 4.12. □

5. Proof of $H'(n-2) \Rightarrow H(n)$, $n \ge 5$

5.1. The subscheme $T^*(n)$

We define $T^*(n)$ as the generic union of $\lambda(n)$ disjoint t-conics, and $\tau(n)$ t-points, $\delta(n)$ d-points, $\epsilon(n)$ s-point which are cocyclic. We see that:

$T^*(n)$ is numerically L_n-settled $\iff$ $h^0(L_n) = h^0(L_{n|T^*(n)})$,
if S^* is a s-point (resp. d-point, t-point, t-line, t-conic, t-bamboo),
then $h^0(L_{n|S^*}) = 1$ (resp. 2, 3, $3n-1$, $6n-5$, $2(6n-5)-3 = 12n-13$).

It follows that: $h^0(L_{n|T^*(n)}) = \lambda(n)(6n-5) + 3\tau(n) + 2\delta(n) + \epsilon(n)$.

Thus, in order to get $T^*(n)$ numerically L_n-settled, we may take:
$\lambda(n) = [\dfrac{h^0(L_n)}{6n-5}]$, $\tau(n) = [\dfrac{s(n)}{3}]$, $2\delta(n) + \epsilon(n) = \{\dfrac{s(n)}{3}\}$, $\delta(n), \epsilon(n) \in \{0, 1\}$,
where: $h^0(L_n) = h^0(\Omega(n)) = \dfrac{(n^2-1)(n+2)}{2}$, and $s(n) = \{\dfrac{h^0(L_n)}{6n-5}\}$.

We must prove the statement $H(n) : H^0(L_n \otimes I_{T^*(n)}) = 0$ by the Horace method. We shall build a specialization $T_s(n)$ of $T^*(n)$ and show that $H^0(L_n \otimes I_{T_s(n)}) = 0$.

5.2. Specialization of $T^*(n)$ - The subscheme $T'^*(n-2)$

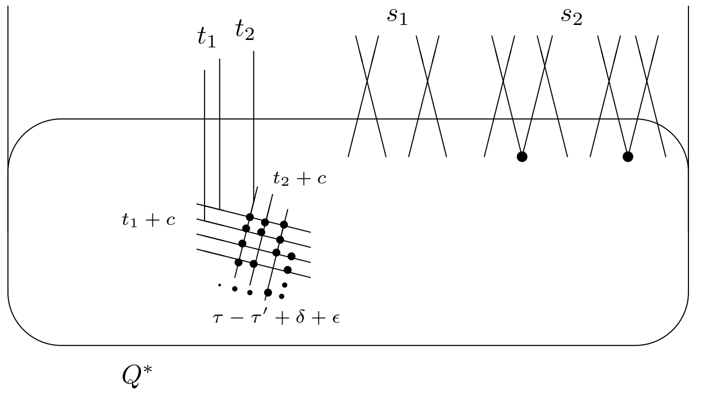
We define $T_s(n)$ as a union of:
- s_1 t-conics in general position,
- s_2 t-bamboos,
- t_1 degenerate t-conics: one of the lines of each of them is contained in Q and is of type $(1, 0)$,
- t_2 degenerate t-conics: one of the lines of each of them is contained in Q and is of type $(0, 1)$,
- c t-conics in Q^*;
- the t-first infinitesimal neighborhood (cf. 4.1) of $c^2 - c$ intersection points of c conics,
- the t-first infinitesimal neighborhood of s_2 triple-points, among the intersection points, with Q, of the s_2 bamboos,
- the t-first infinitesimal neighborhood of t_1t_2 intersection points of t_1+t_2 lines in Q,
- the t-first infinitesimal neighborhood of $(t_1+t_2)c$ intersection points, with c conics, of t_1+t_2 lines,
- the t-first infinitesimal neighborhood of τ' cocyclic t-points, where $\tau' \leq \tau(n)$ and $\tau' \leq (t_1+c)+(t_2+c) = t_1+t_2+2c$,
- $(\tau(n)-\tau')$ t-points, $\delta(n)$ d-point and $\epsilon(n)$ s-point lying on a t-conic in Q^*.

The integers $s_1, s_2, t_1, t_2, c, \tau', p_1, q_1$ are chosen in such a manner that the subscheme $T_s(n)$ is a (L_n, Q^*)-adjusted specialization of $T^*(n)$ (cf. Lemma 3.1). We may then use the Horace method by exploiting the divisor Q^*. In this case, we denote by $T'^*(n-2)$ the residual scheme of $T_s(n)$. It consists of:
- s_1 disjoint t-conics, s_2 disjoint t-bamboos, t_1+t_2 disjoint t-lines and
- $(t_1+c)(t_2+c) - c + \tau'$ t-points lying on a t-grille of type (p_1, q_1).

Since $L_n(-Q^*) = L_{n-2}$, the (L_n, Q^*)-adjusting condition gives:

$$h^0(L_{n-2}) = (6n-17)s_1 + (3n-7)(t_1+t_2) + (12n-37)s_2 + 3[(t_1+c)(t_2+c) - c + \tau'].$$

We prove $H^0(L_n \otimes I_{T_s(n)}) = 0$. We exploit Q^*. The dègue is the statement $H'(n-2) : H^0(L_{n-2} \otimes \mathcal{I}_{T'^*(n-2)}) = 0$, which is true by hypothesis.

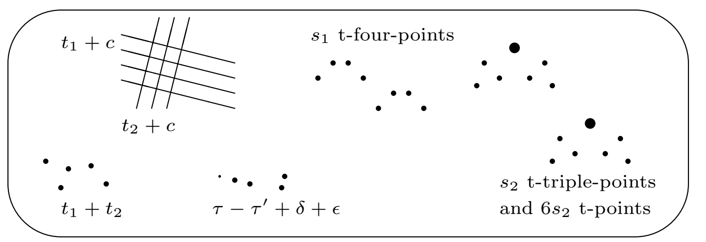


We prove now the dîme. We obtain the following facts:
- the s_1 t-conics of $T_s(n)$ meet Q^* in s_1 t-four-points,
- the s_2 t-bamboos meet Q^* in $6s_2$ t-points and in s_2 t-triple-points.

Thus, the trace $T_s(n) \cap Q^*$ is the subscheme J described in Corollary 4.13 with:

$$a = n-c-t_1,\ b = n-c-t_2,\ v = s_1,\ m = s_2,\ u = 6s_2+t_1+t_2,\ \mu = \tau(n)-\tau'.$$

$T_s(n) \cap Q^*$:



Furthermore, $L_{n|Q^*}$ is isomorphic to K_n and J is, by construction (see Lemma 3.1), numerically K_n-settled. One has: $h^0(K_n) = h^0(K_{n|J})$, which is equivalent to:

$$(E_1) : 3ab - a - b - 1 = 12s_1 + 27s_2 + 3(t_1 + t_2) + 3(\tau(n) - \tau') + 2\delta + \epsilon.$$

Lemma 5.1. *If $t_1, t_2, c, a = n - c - t_1$ and $b = n - c - t_2$ satisfy Equation (E_1), then $\tau(n) - \mu_{\max}(a, b) \leq t_1 + t_2 + 2c$.*

Proof. We know that $s(n) = \{\dfrac{h^0(L_n)}{6n-5}\} \leq 6n - 6$ and $\tau(n) = [\dfrac{s(n)}{3}] \leq 2n - 2$. Moreover, one has: $\tau(n) = 2n - 2 \Rightarrow (\delta = \epsilon = 0) \Rightarrow (a + b \equiv 2 \mod 3)$.

- If $a+b \equiv 0 \mod 3$, then

$$\tau(n) \leq 2n-3 \text{ and } \tau(n) - \mu_{\max}(a,b) = \tau(n) - (a+b-3) \leq t_1 + t_2 + 2c.$$

- If $a+b \not\equiv 0 \mod 3$, then

$$\tau(n) \leq 2n-2 \text{ and } \tau(n) - \mu_{\max}(a,b) = \tau(n) - (a+b-2) \leq t_1 + t_2 + 2c.$$

□

We suppose that $(a,b) \neq (1,1)$. According to the hypotheses of Proposition 4.12, the integers $s_1, s_2, t_1, t_2, c, \tau', p_1, q_1, a, b, f, h, u, \mu$ must satisfy:

$$(\star\star): \begin{cases} \lambda(n) = s_1 + 2s_2 + t_1 + t_2 + c,\ t_1 \geq t_2 \\ a = n - c - t_1,\ b = n - c - t_2 \\ 2 \leq a \leq b \leq 2a-1,\ h = b - a \geq 0, f = 2a - b \geq 1 \\ s_1 \leq V_{\max}(f,h),\ s_2 \leq M_{\max}(f,h) \\ p_1 = c + t_1 \text{ if } \tau' = 0,\ p_1 = c + t_1 + 1 \text{ otherwise} \\ q_1 = c + t_2 \text{ if } \tau' = 0,\ q_1 = c + t_2 + 1 \text{ otherwise} \\ \tau(n) - \tau' \leq \mu_{\max}(a,b),\ 0 \leq \tau' \leq \min(t_1 + t_2 + 2c, \tau(n)). \end{cases}$$

It remains then to prove the existence of $s_1, s_2, t_1, \ldots$ satisfying Equation (E_1) and Conditions $(\star\star)$ above.

5.3. Choice for the integers $s_1, s_2, t_1, \ldots$

We would like to know the orders of magnitude of integers involved in the definitions of $T^*(n)$, $T_s(n)$ and of $T'^*(n-2)$, for sufficiently large values of n. We shall prove (Proposition 5.5) that we may take $n \geq 25$ but $n \notin \Lambda = \{26, 27, 30, 31, 33, 34, 37, 38, 43, 45, 48, 51, 55, 72\}$. For $2 \leq n \leq 24$ or for $n \in \Lambda$, see Section 5.4.

In the subscheme $T^*(n)$, four integers occur: $\lambda(n)$, $\tau(n)$, $\delta(n)$ and $\epsilon(n)$. One has:

$\lambda(n) = [\dfrac{h^0(L_n)}{6n-5}]$ with $h^0(L_n) = \dfrac{n^3 + 2n^2 - n - 2}{2}$, so: $\lambda(n) \sim \dfrac{n^2}{12} + \dfrac{17n}{72}$,

$\tau(n) = [\dfrac{s(n)}{3}]$ with $s(n) = \{\dfrac{h^0(L_n)}{6n-5}\} < 6n-5$, so: $\tau(n) \leq 2n-2$,

$2\delta(n) + \epsilon(n) = \{\dfrac{s(n)}{3}\}$ with $0 \leq \delta(n) + \epsilon(n) \leq 1$.

In the subscheme $T_s(n)$, we must estimate five integers: $s_1,\ s_2,\ t_1,\ t_2$ and c. The adjusting condition gives:

$$3ab - a - b - 1 = 12s_1 + 27s_2 + \cdots \text{ with } a = n - c - t_1,\ b = n - c - t_2.$$

We take: $12s_1 \sim 3ab \sim n^2$, $c \sim 2t_1 \sim \dfrac{n}{3}$, $0 \leq t_2 \leq 2$, since $\lambda(n) \sim \dfrac{n^2}{12} + \dfrac{17n}{72}$. More precisely, we obtain

Proposition 5.2. *The following integers, if they exist, satisfy Equation* (E_1)*:*

$$\begin{array}{l} t_1 = [\frac{n}{6}] + \theta,\ c = [\frac{n}{3}],\ t_2 = \{\frac{2n+1+s(n)-t_1-2c}{3}\} \\ 3s_2 = \max(0, B(n,\theta), B(n,\theta) + 3(\tau(n) - \mu_{\max}(a,b))),\ 3\tau' = 3s_2 - B(n,\theta), \\ s_1 = \lambda(n) - t_1 - t_2 - c - 2s_2, \end{array}$$

where $a = n - c - t_1,\ b = n - c - t_2 \geq 5$ *and*

$$\begin{cases} \theta = 3\ [\frac{\theta_1}{3}],\ \theta_1 = \min([\frac{3\tau(n)+A(n)}{3b-10}], [\frac{3([\frac{n}{6}]+t_2+2c)+A(n)}{3b-13}]), \\ B(n,\theta) = A(n) - (3(n-t_2-c)-10)\theta = A(n) - (3b-10)\theta, \\ A(n) = 3(n-c-[\frac{n}{6}])b - 2n - 1 - s(n) - 12\lambda(n) + 14c + 10[\frac{n}{6}] + 10t_2. \end{cases}$$

Moreover, one has: $\tau(n) - \tau' \leq \mu_{\max}(a,b),\ 0 \leq \tau' \leq \min(t_1+t_2+2c, \tau(n))$.

Proof. By direct computations, since $a = n - t_1 - c$ and $b = n - t_2 - c$, Equation (E_1) may be written as $(E_2) : B(n,\theta) - 3s_2 + 3\tau' = 0$. The choice of t_2 is due to the fact:

$$a + b + 1 + s(n) \equiv a + b + 1 + 2\delta + \epsilon \equiv 0 \mod 3.$$

It follows that: $A(n) \equiv 0 \mod 3$ and $B(n,\theta) \equiv -(3b-10)\theta \equiv \theta \mod 3$. Conditions $(\star\star)$ and Equation (E_2) give:

$$\begin{array}{l} B(n,\theta) - 3s_2 = -3\tau' \leq 0,\ B(n,\theta) - 3s_2 = -3\tau' \geq -3\min(t_1+t_2+2c, \tau(n)), \\ B(n,\theta) - 3s_2 = -3\tau' \leq -3(\tau(n) - \mu_{\max}(a,b)). \end{array}$$

Thus, we must have:

$$\begin{array}{l} 3s_2 \geq B(n,\theta), B(n,\theta) \equiv 0 \mod 3,\ 0 \leq 3s_2 \leq 3\tau(n) + B(n,\theta), \\ 0 \leq 3s_2 \leq 3(t_1+t_2+2c) + B(n,\theta),\ 3s_2 \geq 3(\tau(n) - \mu_{\max}(a,b)) + B(n,\theta). \end{array}$$

Since $t_1 = [\frac{n}{6}] + \theta$, θ satisfies:

$$3\tau(n) + A(n) \geq (3b-10)\theta,\ 3([\frac{n}{6}] + t_2 + 2c) + A(n) \geq (3b-13)\theta.$$

It suffices then to take:

$$\begin{array}{l} \theta = 3\ [\frac{\theta_1}{3}],\ 3s_2 = \max(0, B(n,\theta), 3(\tau(n) - \mu_{\max}(a,b)) + B(n,\theta)), \\ \tau' = s_2 - \frac{1}{3}\ B(n,\theta), \\ \text{with } \theta_1 = \min([\frac{3\tau(n)+A(n)}{3b-10}], [\frac{3([\frac{n}{6}]+t_2+2c)+A(n)}{3b-13}]) \text{ and } b \geq 5. \end{array}$$

Now, we check that: $\tau' \geq 0, \tau(n) - \tau' \leq \mu_{\max}(a,b), \tau' \leq \min(t_1+t_2+2c, \tau(n))$. The first two inequalities follow from the facts:

$$3s_2 \geq B(n,\theta) \text{ and } 3s_2 \geq 3(\tau(n) - \mu_{\max}(a,b)) + B(n,\theta).$$

It remains to prove the third one. Since $3b-10 \geq 3b-13 \geq 1$ and $\theta \leq \theta_1$, one has: $(3b-10)\theta \leq 3\tau(n) + A(n), (3b-13)\theta \leq 3([\frac{n}{6}] + t_2 + 2c) + A(n)$. Therefore, $B(n,\theta)$ satisfies: $3\min(t_1+t_2+2c, \tau(n)) \geq -B(n,\theta)$, because

$$\begin{aligned}&3(t_1+t_2+2c) + A(n) = 3(\theta + [\tfrac{n}{6}] + t_2 + 2c) + A(n) \geq (3b-10)\theta,\\ &3\tau(n) + B(n,\theta) \geq 0 \text{ and } 3(t_1+t_2+2c) + B(n,\theta) \geq 0.\end{aligned}$$

- If $3s_2 = 0$, then $3\tau' = -B(n,\theta) \leq 3\min(\tau(n), t_1+t_2+2c)$.
- If $3s_2 = B(n,\theta)$, then $\tau' = 0 \leq \min(\tau(n), t_1+t_2+2c)$.
- If $3s_2 = B(n,\theta) + 3(\tau(n) - \mu_{\max}(a,b))$, then

$$\begin{aligned}&3\tau' = 3s_2 - B(n,\theta) = 3(\tau(n) - \mu_{\max}(a,b)) \leq 3\tau(n),\\ &3\tau' = 3(\tau(n) - \mu_{\max}(a,b)) \leq 3(t_1+t_2+2c) \text{ by Lemma 5.1.}\end{aligned}$$

□

Proposition 5.5 allows us to determine all values of n for which Equation (E_1) and Conditions $(\star\star)$ hold. We shall use the following results for its proof.

Lemma 5.3. *We consider the natural numbers: $b, A(n), \theta_1, \theta$ and s_2 defined in Proposition 5.2. One has for $n \geq 68$:*

$$b \geq 5,\ -8n \leq A(n) \leq 6n,\ -5 \leq \theta_1 < 5,\ \theta \in \{-3, 0\} \text{ and } s_2 \leq 2n.$$

Proof. By standard bounding, we obtain:

$$\begin{cases} \dfrac{n}{3} - 1 \leq c < \dfrac{n}{3},\ 0 \leq t_2 \leq 2,\ 0 \leq s(n) \leq 6n-6,\ 0 \leq \tau(n) \leq 2n-2,\\ \dfrac{2n}{3} - 2 \leq b = n - c - t_2 < \dfrac{2n}{3} + 1,\ \dfrac{n}{2} \leq a + \theta < \dfrac{n}{2} + 2,\\ \dfrac{n^2}{12} + \dfrac{17n}{72} - 2 \leq \lambda(n) \leq \lambda_{\max} = \dfrac{h^0(L_n)}{6n-5} \leq \dfrac{n^2}{12} + \dfrac{17n}{72} + 1. \end{cases} \tag{6}$$

So, $b \geq 5$ if $n \geq 11$. Set $n = 6\ell + w$ with $0 \leq w \leq 5$. Since $t_1 = [\frac{n}{6}] + \theta = \ell + \theta$, $t_2 \in \{0,1,2\}$ and $\lambda(n) = \dfrac{n^2}{12} + \dfrac{17n}{72} + \zeta$, for some $\zeta \in [-2, 1]$, simple calculations give: $-8n \leq A(n) \leq 6n$, because

$$A(n) = A_1(n) \text{ (resp. } A_1(n) - 21\ell - 6w + 3t_2 + 17) \text{ if } w \leq 2 \text{ (resp. if } w \geq 3),$$

where $A_1(n) = (9w + 9 - 9t_2)\ell - s(n) - 12\zeta + 10t_2 + 2w^2 - \dfrac{29}{6}w - 3t_2 - 1$.

By definition, $\theta = 3[\frac{\theta_1}{3}] \equiv 0 \mod 3$ and θ_1 satisfies:

$$\theta_1 \leq \frac{3([\frac{n}{6}] + t_2 + 2c) + A(n)}{3b - 13} \leq \frac{(\frac{5n}{2} + 6) + 6n}{3(\frac{2n}{3} - 2) - 13} = \frac{17n + 12}{4n - 38} < 5 \text{ if } n \geq 68,$$
$$\theta_1 \geq \frac{A(n)}{3b - 10} \geq \frac{-8n}{3b - 10} \geq \frac{-8n}{3(\frac{2n}{3} - 2) - 10} = \frac{-8n}{2n - 16} \geq -5 \text{ if } n \geq 40.$$

Hence, $\theta = 3[\frac{\theta_1}{3}] \leq \theta_1 < 5$, $\frac{\theta_1}{3} \geq -\frac{5}{3}$, and $\theta = 3[\frac{\theta_1}{3}] \geq 3 \times (-2) = -6$. We get $\theta \in \{-6, -3, 0, 3\}$. Now, we prove that $\theta \notin \{-6, 3\}$.

If $\theta = -6$, then $\theta_1 < -3$ so that $\theta_1 \leq -4$. Thus

$$\frac{3\tau(n) + A(n)}{3b - 10} < -3 \quad \text{or} \quad \frac{3([\frac{n}{6}] + t_2 + 2c) + A(n)}{3b - 13} < -3.$$

i.e., $(3\tau(n) + A(n) + 9b - 30 < 0)$ or $(3([\frac{n}{6}] + t_2 + 2c) + A(n) + 9b - 39 < 0)$. It is impossible, if $n \geq 56$, by taking into account the above expressions of $A(n)$ and by the facts: $-2 \leq -s(n) + 3\tau(n) = -2\delta - \epsilon \leq 0$ and $b \sim 4\ell$. If $\theta = 3$, then $\theta_1 \geq 3$ and $\frac{3\tau(n) + A(n)}{3b - 10} \geq 3$. So, $3\tau(n) + A(n) - 9b + 30 \geq 0$, which is also impossible because $3\tau(n) + A(n)$ is at most of order 33ℓ but $-9b + 30 \sim -36\ell$.

It remains to prove that $s_2 \leq 2n$.

- If $\theta = -3$, then $(3\tau(n) + A(n) < 0)$ or $(3([\frac{n}{6}] + t_2 + 2c) + A(n) < 0)$. Hence $A(n) < 0$ and $B(n, \theta) = A(n) + 3(3b - 10) < 9b - 30$.

- If $\tau(n) - \mu_{\max}(a, b) \leq 0$, then $3s_2 = \max(0, B(n, \theta)) \leq 9b - 30 \leq 6n - 21$.
- If $\tau(n) - \mu_{\max}(a, b) \geq 1$, then set $C(n, \theta) = B(n, \theta) + 3\tau(n) - 3\mu_{\max}(a, b)$. Note that $\mu_{\max}(a, b) \geq a + b - 3$ and from Inequalities (6), $a \geq \frac{n}{2} + 3, b \leq \frac{2n}{3} + 1$. If $A(n) + 3\tau(n) < 0$ then $C(n, \theta) \leq 9b - 30 - 3\mu_{\max}(a, b) \leq 6b - 21 \leq 4n$. If $A(n) + 3([\frac{n}{6}] + t_2 + 2c) < 0$ then

$$C(n, \theta) \leq -3([\frac{n}{6}] + t_2 + 2c) + 9b - 30 + 3\tau(n) - 3\mu_{\max}(a, b) \leq 6n.$$

Thus, $3s_2 = \max(0, C(n, \theta)) \leq 6n$.

- If $\theta = 0$, then $B(n, \theta) = A(n) \leq 6n$ and $\theta_1 < 3$.

- If $\tau(n) - \mu_{\max}(a, b) \leq 0$, then $3s_2 = \max(0, B(n, \theta)) \leq 6n$.
- Now, we suppose that $\tau(n) - \mu_{\max}(a, b) \geq 1$. Since $\theta_1 < 3$, one has

$$3\tau(n) + A(n) < 3(3b - 10) \quad \text{or} \quad 3([\frac{n}{6}] + t_2 + 2c) + A(n) < 3(3b - 13).$$

Therefore $(C(n, \theta) \leq 9b - 30 - 3\mu_{\max}(a, b) \leq 6b - 21 \leq 4n)$
or $(C(n, \theta) \leq -3([\frac{n}{6}] + t_2 + 2c) + 3(\tau(n) - \mu_{\max}(a, b)) + 9b - 39 \leq 6n)$.
So, $3s_2 = \max(0, C(n, \theta)) \leq 6n$. □

LEMMA 5.4. *Let a, b, c, t_1, t_2 be the integers defined in Proposition 5.2 and put $f = 2a - b = i + 3\ell$, $h = b - a, 1 \leq i \leq 3$. Then for $n \geq 347$, one has:*

$$2n \leq M_{\max}(f,h) \text{ and } \frac{n^2}{12} - \frac{n}{4} + 9 \leq V_{\max}(f,h).$$

Proof. We get from Inequalities (6): $f \geq \frac{n}{3} - 1$ and $h \geq \frac{n}{6} - 6$. Thus $f, h \geq 2$ if $n \geq 48$ and

$$\begin{aligned} M_{\max}(f,h) &= 2f - i - 6 + h^2 + h \geq \frac{n^2}{36} - \frac{7n}{6} + 19 \geq 2n \text{ if } n \geq 108, \\ V_{\max}(f,h) &= i - 1 + (f + i - 4)(\ell - 1) + (f-2)h \\ &\geq \frac{5n^2}{54} - \frac{7n}{2} + \frac{74}{3} \geq \frac{n^2}{12} - \frac{n}{4} + 9 \text{ if } n \geq 347. \end{aligned}$$

□

PROPOSITION 5.5. *If $n \geq 25$ and $n \notin \Lambda$, then the integers defined in Proposition 5.2 satisfy Equation (E_1) and Conditions $(\star\star)$.*

Proof. According to (the proof of) Proposition 5.2, it remains to prove, for such n, the existence of integers $s_1, s_2, t_1, t_2, c, \ldots$ satisfying:

$$\begin{aligned} &5 \leq a = f + h \leq b = f + 2h < 2a,\ s_1 \leq V_{\max}(f,h),\ s_2 \leq M_{\max}(f,h), \\ &\text{where } s_1 + 2s_2 + t_1 + t_2 + c = \lambda(n),\ t_1 \geq t_2,\ a = n - c - t_1,\ b = n - c - t_2. \end{aligned}$$

From Inequalities (6) and from Lemmas 5.3 and 5.4, one has for $n \geq 347$:

$$\begin{aligned} &\theta \in \{-3, 0\},\ \frac{n}{6} - 4 < t_1 \leq \frac{n}{6},\ h = b - a > \frac{n}{6} - 6 \geq 2,\ 5 \leq \frac{n}{2} \leq a < \frac{n}{2} + 5, \\ &2 \leq \frac{n}{3} - 1 = 2(\frac{n}{2}) - (\frac{2n}{3} + 1) < f = 2a - b \leq \frac{n}{3} + 12,\ s_2 \leq 2n \leq M_{\max}(f,h) \\ &s_1 \leq \lambda_{\max} - (\frac{n}{6} - 7) - 0 - (\frac{n}{3} - 1) \leq \frac{n^2}{12} - \frac{n}{4} + 9 \leq V_{\max}(f,h). \end{aligned}$$

Conditions $(\star\star)$ are then satisfied, for any $n \geq 347$. By direct computations in Section 6.2, those conditions hold too, for $25 \leq n \leq 346$, except for $n \in \Lambda$. □

5.4. Initial cases

We recall that Y denotes the generic union of r skew conics, Q a smooth quadric surface in $\mathbb{P}^3$, and Ω the cotangent bundle over $\mathbb{P}^3$. In this section, we prove that:
- the map $r_Y(n) : H^0(\Omega(n)) \to H^0(\Omega(n)_{|Y})$ has maximal rank if $2 \leq n \leq 4$,
- $H'(n-2) \Rightarrow H(n)$ if ($5 \leq n \leq 24$ or $n \in \Lambda$).

5.4.1. Case $2 \leq n \leq 4$

- $n = 2$

We prove that $r_Y(2)$ is injective if $r = 1$, i.e., $h^0(\Omega(2) \otimes I_Y) = 0$ if Y is a conic. We exploit a plane H containing Y. The dègue: $h^0(\Omega(1)) = 0$, is satisfied. We obtain also the dîme: $h^0(\Omega(2)_{|H} \otimes I_Y) = 0$, since $h^0(\Omega(2)_{|H} \otimes I_Y) = h^0(\Omega(2)_{|H} \otimes \mathcal{O}_H(-2)) = h^0(\Omega_{|H}) = h^0(\Omega_H \oplus \mathcal{O}_H(-1)) = 0$. It follows that $r_Y(2)$ is injective for any $r \geq 1$.

- $n = 3$

We prove that $r_Y(3)$ is injective if $r = 2$ and it is surjective if $r = 1$.

Injectivity of $r_Y(3)$: $H^0(\Omega(3) \otimes I_Y) = 0$ if Y is a union of two skew conics. By Lemma 4.1, Y specializes to a union of two (non-disjoint) conics in Q with the infinitesimal neighborhood (in $\mathbb{P}^3$) of their two intersection points. One exploits Q. The residual scheme Y'' is exactly two points. Hence, we get the dègue: $H^0((\Omega(1) \otimes I_{Y''}) = 0$. The trace Y' is a union of two conics (a curve of type $(2,2)$ in Q). So, the dîme: $H^0(\Omega(3)_{|Q} \otimes I_{Y'}) = 0$ is also satisfied because: $h^0(\Omega(3)_{|Q} \otimes I_{Y'}) = h^0(\overline{\Omega}(1)) = 0$.

Surjectivity of $r_Y(3)$: $H^1(\Omega(3) \otimes I_Y) = 0$ if Y is a conic. We may suppose that $Y \subset Q$ and we exploit Q. We obviously get the dègue: $H^1(\Omega(1)) = 0$. Now, to prove the dîme: $H^1(\Omega(3)_{|Q} \otimes I_Y) = 0$, we remark that the trace (Y itself) is a curve of type $(1,1)$ on Q. Thus, $h^1(\Omega(3)_{|Q} \otimes I_Y) = h^1(\overline{\Omega}(2)) = 0$.

- $n = 4$

We prove that $r_Y(4)$ is injective (resp. surjective) if $r = 3$ (resp. $r = 2$).

Injectivity of $r_Y(4)$: $H^0(\Omega(4) \otimes I_Y) = 0$ if Y is a union of 3 skew conics. Y specializes to a union of 2 (non disjoint) conics of Q, with the infinitesimal neighborhood (in $\mathbb{P}^3$) of their two intersection points, and one conic not contained in Q. One exploits Q. The residual scheme is a union of one conic and two points. Therefore, the dègue: $H^0(\Omega(2) \otimes I_{Y''}) = 0$ is verified (see case $n = 2$). The trace Y' consists of two conics and four points. The dîme: $H^0(\Omega(4)_{|Q} \otimes I_{Y'}) = 0$ is then equivalent to: $H^0(\overline{\Omega}(2) \otimes I_Z) = 0$, where Z is the union of those 4 points. In order to prove: $H^0(\overline{\Omega}(2) \otimes I_Z) = 0$, we exploit a conic C in Q, containing these 4 points: the dègue is trivial. We get the dîme since: $h^0(\overline{\Omega}(2)_{|C} \otimes I_Z) = h^0((2\mathcal{O}_{\mathbb{P}^1}(1) \oplus \mathcal{O}_{\mathbb{P}^1}(2)) \otimes I_Z) = h^0(2\mathcal{O}_{\mathbb{P}^1}(-3) \oplus \mathcal{O}_{\mathbb{P}^1}(-2)) = 0$.

Surjectivity of $r_Y(4)$: $H^1(\Omega(4) \otimes I_Y) = 0$ if Y is a union of 2 skew conics. One exploits a plane H containing one of the 2 conics. The residual schema Y'' is a conic and the dègue: $H^1(\Omega(3) \otimes I_{Y''}) = 0$ is satisfied (see case $n = 3$). The trace Y' is a union of one conic and 2 points. The dîme is equivalent to: $H^1(\Omega(2)_{|H} \otimes I_{Z'}) = 0$, where Z' consists of 2 points (of Y'). To prove the last equality, one exploits a line passing through those 2 points. The dîme and dègue are trivial.

n	λ	τ	$2\delta+\epsilon$	c	s_1	s_2	t_1	t_2	τ'
5	3	3	0	2	0	0	1	0	0
6	4	5	1	4	0	0	0	0	3
7	5	10	1	2	1	1	0	0	2
8	7	4	2	3	1	1	1	0	2
9	8	16	0	8	0	0	0	0	16
10	10	14	2	2	2	1	3	1	5
11	12	16	0	4	4	2	0	0	6
12	14	21	0	2	4	1	4	2	9
13	17	6	1	5	4	4	0	0	0
14	19	19	2	5	8	3	0	0	4
15	22	11	1	4	6	3	4	2	11
16	25	6	2	5	10	3	2	2	3
17	28	6	2	4	1	9	5	0	0
18	31	12	1	7	14	5	0	0	0
19	34	24	2	7	15	6	0	0	3
20	38	6	1	7	27	0	2	2	5
21	41	33	0	8	25	4	0	0	9
22	45	27	0	7	30	2	2	2	9
23	49	27	2	8	34	2	2	1	12
24	53	36	0	9	41	1	1	0	10

n	c	s_1	s_2	t_1	t_2	τ'
26	8	49	0	3	2	1
27	9	49	2	3	2	2
30	11	61	3	2	2	9
31	7	44	13	9	1	2
33	12	80	1	2	2	27
34	14	90	0	0	0	5
37	9	59	22	10	0	23
38	16	113	0	0	0	14
43	18	145	0	1	0	27
45	18	160	0	1	0	0
48	20	183	0	0	0	9
51	21	207	0	0	0	39
55	23	241	0	1	0	14
72	27	415	0	7	0	1

5.4.2. Proof of $H'(n-2) \Rightarrow H(n)$, $5 \leq n \leq 24$ or $n \in \Lambda$

We give here some tables of integers involved in $T^*(n), T_s^*(n)$ and in $T'^*(n-2)$. For $n \neq 20$, these integers are chosen (by Maple computations) in order to satisfy Equation (E_1) and Conditions $(\star\star)$. For example, the first row of the table means that $T'^*(3)$ does not contain any t-conic ($s_1 = 0$), any t-bamboo ($s_2 = 0$). It consists of $t_1+t_2 = 1$ t-line and $(c+t_1)(c+t_2)-c+\tau' = 3.2-2+0 = 4$ t-points on a t-grille of type (p_1, q_1), where $p_1 = c + t_1 = 3, q_1 = c + t_2 = 2$. Note that, for each n, the corresponding 6-tuple $(c, s_1, s_2, t_1, t_2, \tau')$ may not be unique.

If $n = 20$, then one has $a = b = f = 11 = i + 3\ell$, $i = 2, \ell = 3$, $s_2 = 0$, $\mu = 1$. We cannot apply Proposition 4.12, since $v = s_1 = 27 > 19 = V_{\max}(f, 0)$.

However, by exploiting alternately 3 times t-rational curves of type $(1,2)$ and $(2,1)$ as in the proof of Proposition 4.12, and by taking $v_{11} = 10, v_{21} = 8$, $v_{12} = 7, v_{22} = 2, v_{13} = v_{23} = 0$, and $m_{1k} = m_{2k} = 0$ for any $1 \leq k \leq 3$, we see that the corresponding subscheme $F(a,b)$ is $K_{a,b}$-settled.

6. Some Maple Programs

We give here the integers defined in Section 5.1 and in Remark 4.11:

$f(n) = h^0(L_n) = h^0(\Omega(n)),\ g(n) = h^0(L_{n|C^*}) = h^0(\Omega(n)_{|C})$ with C a conic, $\lambda(n),\ s(n),\ \tau(n), \Delta(n) = 2\delta(n) + \epsilon(n), \mu_{\max}(a,b), V_{\max}(f,h), M_{\max}(f,h)$.

```
restart:
f:=proc(n)(n**2-1)*(n+2)/2;end;
g:=proc(n) (6*n-5);end;

lambda:=proc(n) iquo(f(n),g(n));end;
s:=proc(n) irem(f(n),g(n));end;
tau:=proc(n) iquo(s(n),3);end;
Delta:=proc(n) irem(s(n),3);end;

mumax:=proc(a,b) if irem(a+b,3)=0 then a+b-3;else a+b-2;fi;end;

Vmax0:=proc(f) ell:=iquo(f-1,3):ii:=f-3*ell:if f<=3 then 0;
else if f<=6 then ii-1;else ii-1+(ell-1)*(f+ii-4);fi;fi;end;
Vmax:=proc(f,h) ell:=iquo(f-1,3):ii:=f-3*ell:if f+h<=3 then 0;
else if (f=1 and h>= 3) then 0; else Vmax0(f)+(f-2)*h; fi;fi;end;

Mmax0:=proc(f) ell:=iquo(f-1,3):ii:=f-3*ell: if f <= 3 then 0;
else if f<=6 then ii;else 2*f-ii-6;fi;fi;end;
Mmax:=proc(f,h)  ell:=iquo(f-1,3):ii:=f-3*ell:if f+h<=3 then 0;
else if (f=1 and h >= 3) then h*(h-1); else Mmax0(f)+h*(h+1);
fi;fi;end;
```

6.1. Program 1

The function $List1(n)$ returns the list of $n, \lambda(n), \tau(n), \Delta(n), c, s_1, s_2, \ldots$ if they satisfy Conditions $(\star\star)$ and Equation (E_1) in Section 5.2. It returns "impossible" if they do not. Note also that EQUA1 is exactly Equation (E_1).

```
List1:=proc(n) c:=iquo(n,3):t2:=irem(2*n+1+s(n)-iquo(n,6)-2*c,3):
b:=n-c-t2:lamb:=lambda(n):
A:=3*b*(n-c-iquo(n,6))-2*n-1-s(n)-12*lamb+14*c
                                             +10*iquo(n,6)+10*t2:
```

```
theta1:=min(floor((3*tau(n)+A)/(3*b-10)),
floor((3*(iquo(n,6)+t2+2*c)+A)/(3*b-13))):
theta:=3*floor(theta1/3):
t1:=iquo(n,6)+theta:
a:=n-c-t1:
ef:=2*a-b:hh:=b-a:
iji:=ef-3*iquo(ef-1,3):
MUMAX:=mumax(a,b):
Bntheta:=A-(3*b-10)*theta:
troissdeux:=max(0,Bntheta,Bntheta+3*(tau(n)-MUMAX)):
s2:=troissdeux/3:s1:=lamb-2*s2-t1-t2-c:tauprim:=s2-Bntheta/3:
uu:=6*s2+t1+t2:muu:=tau(n)-tauprim:
EQUA1:=3*a*b-a-b-1-(12*s1+9*s2+3*uu+3*muu+Delta(n)):
VEmax:=Vmax(ef,hh):EMmax:=Mmax(ef,hh):
if EQUA1 = 0 and a <= b and b < 2*a and muu >= 0 and s1>=0 and
tauprim <= t1+t2+2*c and muu <= MUMAX and s2<=EMmax  and
s1 <= VEmax then [ene=n,lambdaa=lamb,TAU=tau(n),Deltaa=Delta(n),
C=c,es1=s1,es2=s2,te1=t1,te2=t2,Tauprime=tauprim,THeta=theta];
else impossible;fi;end;
```

6.2. Program 2

*List*2 returns the list of integers $n \in \{5, \ldots, 346\}$ for which Equation (E_1) and Conditions $(\star\star)$ are not satisfied. We see that it contains only integers n such that $5 \leq n \leq 24$ or $n \in \Lambda$.

```
ll:={}:for n from 5 to 346 do if evalb(List1(n)=impossible) then
ll:={op(ll),n};fi;od:List2=ll;
```

```
List2 = {5,6,7,8,9,10,12,15,16,17,18,19,20,21,22,23,24,
         26,27,30,31,33,34,37,38,43,45,48,51,55,72}
```

Acknowledgments

We are indebted to the referees for their remarks and suggestions which considerably improve this paper.

References

[1] J. Alexander and A. Hirschowitz, *Un lemme d'Horace différentiel: applications aux singularités hyperquartiques de* $\mathbb{P}^5$, J. Algebraic Geom. **1** (1992), no. 3, 411–426.

[2] E. Ballico, *On the postulation of disjoint rational curves in a projective space*, Rend. Sem. Mat. Univ. Politec. Torino **44** (1986), no. 2, 207–249.

[3] E. Ballico, *Generators for the homogeneous ideal of s general points in* $\mathbb{P}^3$, J. Algebra **106** (1987), no. 1, 46–52.

[4] D. Eisenbud and A. Van de Ven, *On the normal bundles of smooth rational space curves*, Math. Ann. **256** (1981), no. 4, 453–463.

[5] R. Hartshorne, *Algebraic geometry*, Springer Verlag, New York-Heidelberg, 1977.

[6] R. Hartshorne and A. Hirschowitz, *Cohomology of a general instanton bundle*, Ann. Sci. École Norm. Sup. (4) **15** (1982), no. 2, 365–390.

[7] R. Hartshorne and A. Hirschowitz, *Droites en position générale dans l'espace projectif*, Algebraic geometry (La Rábida, 1981), Lecture Notes in Math., vol. 961, Springer, Berlin, 1982, pp. 169–188.

[8] A. Hirschowitz, *La méthode d'Horace pour l'interpolation à plusieurs variables*, Manuscripta Math. **50** (1985), 337–388.

[9] M. Idà, *On the homogeneous ideal of the generic union of lines in* $\mathbb{P}^3$, J. Reine Angew. Math. **403** (1990), 67–153.

[10] M. Idà, *Postulation problems for vector bundles*, Rend. Istit. Mat. Univ. Trieste **39** (2007), 231–243.

[11] D. Mumford, *Lectures on curves on an algebraic surface*, With a section by G. M. Bergman. Annals of Mathematics Studies, No. 59, Princeton University Press, Princeton, N.J., 1966.

[12] C. Okonek, M. Schneider, and H. Spindler, *Vector bundles on complex projective spaces*, Progress in Mathematics, vol. 3, Birkhäuser, Boston, Mass., 1980.

[13] O. Rahavandrainy, *Sur la résolution des fibrés instantons généraux*, Ph.D. thesis, Université de Nice, 1991.

[14] O. Rahavandrainy, *Résolution des fibrés instantons généraux*, C. R. Acad. Sci. Paris Sér. I Math. **325** (1997), no. 2, 189–192.

[15] O. Rahavandrainy, *Résolution des fibrés généraux stables de rang* 2 *sur* $\mathbb{P}^3$*, de classes de Chern* $c_1 = -1, c_2 = 2p \geq 6$. *I*, Ann. Fac. Sci. Toulouse Math. (6) **19** (2010), no. 2, 231–267.

Author's address:

Olivier Rahavandrainy
Mathematics (LMBA - UMR 6205), University of Brest
6, Avenue Le Gorgeu, C.S. 93837, 29238 Brest Cedex 3, France.
E-mail: `Olivier.Rahavandrainy@univ-brest.fr`

Received January 27, 2014
Revised October 6, 2014

Rend. Istit. Mat. Univ. Trieste
Volume 46 (2014), 231–235

On an inequality from Information Theory

Horst Alzer

Abstract. *We prove that the inequalities*

$$\sum_{j=1}^{n} \frac{q_j(q_j-p_j)^2}{q_j^2+m_j^{\alpha}M_j^{1-\alpha}} \leq \sum_{j=1}^{n} p_j \log\frac{p_j}{q_j} \leq \sum_{j=1}^{n} \frac{q_j(q_j-p_j)^2}{q_j^2+m_j^{\beta}M_j^{1-\beta}} \quad (\alpha,\beta\in\mathbb{R}),$$

where

$$m_j=\min(p_j^2,q_j^2) \quad \textit{and} \quad M_j=\max(p_j^2,q_j^2) \quad (j=1,...,n),$$

hold for all positive real numbers p_j, q_j $(j=1,...,n; n\geq 2)$ *with* $\sum_{j=1}^{n} p_j = \sum_{j=1}^{n} q_j$ *if and only if* $\alpha \leq 1/3$ *and* $\beta \geq 2/3$. *This refines a result of Halliwell and Mercer, who showed that the inequalities are valid with* $\alpha=0$ *and* $\beta=1$.

Keywords: Gibbs' inequality, Kullback-Leibler divergence, information theory, log-function.
MS Classification 2010: 26D15, 94A15.

1. Introduction

If p_j and q_j $(j=1,...,n)$ are positive real numbers with $\sum_{j=1}^{n} p_j = \sum_{j=1}^{n} q_j$, then

$$0 \leq \sum_{j=1}^{n} p_j \log\frac{p_j}{q_j}. \tag{1}$$

The sign of equality holds in (1) if and only if $p_j=q_j$ $(j=1,...,n)$. This inequality is known in the literature as Gibbs' inequality, named after the American scientist Josiah Willard Gibbs (1839-1903). A proof of (1) can be found, for instance, in [5, p. 382].

The expression on the right-hand side of (1) is called the Kullback-Leibler divergence. It is a measure of the difference between the probability distributions $P=\{p_1,...,p_n\}$ and $Q=\{q_1,...,q_n\}$. Gibbs' inequality has many applications in information theory and also in mathematical statistics. It attracted the attention of numerous researchers, who discovered remarkable extensions, improvements and related results. For details we refer to [1, 2, 4] and the references therein.

The work on this note has been inspired by an interesting paper published by Halliwell and Mercer [3] in 2004. They presented the following elegant refinement and converse of (1).

Proposition 1.1. *Let p_j, q_j $(j = 1, ..., n)$ be positive real numbers satisfying $\sum_{j=1}^n p_j = \sum_{j=1}^n q_j$. Then,*

$$\sum_{j=1}^n \frac{q_j(q_j-p_j)^2}{q_j^2+M_j} \le \sum_{j=1}^n p_j \log \frac{p_j}{q_j} \le \sum_{j=1}^n \frac{q_j(q_j-p_j)^2}{q_j^2+m_j}, \tag{2}$$

where

$$m_j = \min(p_j^2, q_j^2) \quad \textit{and} \quad M_j = \max(p_j^2, q_j^2) \quad (j = 1, ..., n).$$

Double-inequality (2) can be written as

$$\sum_{j=1}^n \frac{q_j(q_j-p_j)^2}{q_j^2+m_j^{\alpha}M_j^{1-\alpha}} \le \sum_{j=1}^n p_j \log \frac{p_j}{q_j} \le \sum_{j=1}^n \frac{q_j(q_j-p_j)^2}{q_j^2+m_j^{\beta}M_j^{1-\beta}} \tag{3}$$

with $\alpha = 0$ and $\beta = 1$. With regard to this result it is natural to ask for all real parameters α and β such that (3) holds. In the next section, we establish that (3) is valid if and only if $\alpha \le 1/3$ and $\beta \ge 2/3$. In particular, setting $\alpha = 1/3$ and $\beta = 2/3$ leads to an improvement of both sides of (2).

2. Result

We need certain upper and lower bounds for the log-function.

Lemma 2.1. *(i) If $0 < x \le 1$, then*

$$x-1-\frac{(x-1)^2}{x+x^{1/3}} \le \log x \le x-1-\frac{(x-1)^2}{x+1} \tag{4}$$

with equality if and only if $x = 1$.

(ii) If $x > 1$, then

$$x-1-\frac{(x-1)^2}{x+1} < \log x < x-1-\frac{(x-1)^2}{x+x^{1/3}}. \tag{5}$$

Proof. Let

$$f(x) = \log x - x + 1 + \frac{(x-1)^2}{x+1} \quad \text{and} \quad g(x) = -\log x + x - 1 - \frac{(x-1)^2}{x + x^{1/3}}.$$

Then,

$$f'(x) = \frac{(x-1)^2}{x(x+1)^2} \quad \text{and} \quad g'(x) = \frac{(t-1)^4(t^2+t+1)}{3t^4(t^2+1)^2} \quad (t = x^{1/3}).$$

It follows that f and g are strictly increasing on $(0, \infty)$. Since $f(1) = g(1) = 0$, we conclude that (4) and (5) are valid. □

We are now in a position to prove the following refinement of (2).

THEOREM 2.2. *Let* $\alpha, \beta \in \mathbb{R}$. *The inequalities* (3) *hold for all positive real numbers* p_j, q_j $(j = 1, ..., n; n \geq 2)$ *with* $\sum_{j=1}^n p_j = \sum_{j=1}^n q_j$ *if and only if* $\alpha \leq 1/3$ *and* $\beta \geq 2/3$.

Proof. First, we show that if $\alpha \leq 1/3$ and $\beta \geq 2/3$, then (3) is valid for all $p_j, q_j > 0$ $(j = 1, ..., n)$ with $\sum_{j=1}^n p_j = \sum_{j=1}^n q_j$. Since the sums on the left-hand side and on the right-hand side of (3) are increasing with respect to α and β, respectively, it suffices to prove (3) for $\alpha = 1/3$ and $\beta = 2/3$.

First, let $q_j \leq p_j$. Applying (4) gives

$$\frac{q_j}{p_j} - 1 - \frac{(q_j/p_j - 1)^2}{q_j/p_j + (q_j/p_j)^{1/3}} \leq \log \frac{q_j}{p_j} \leq \frac{q_j}{p_j} - 1 - \frac{(q_j/p_j - 1)^2}{q_j/p_j + 1}.$$

We multiply by p_j and sum up. This yields

$$\begin{aligned}
\sum_{q_j \leq p_j} q_j - \sum_{q_j \leq p_j} p_j - \sum_{q_j \leq p_j} \frac{q_j(q_j - p_j)^2}{q_j^2 + m_j^{2/3} M_j^{1/3}} \\
&= \sum_{q_j \leq p_j} q_j - \sum_{q_j \leq p_j} p_j - \sum_{q_j \leq p_j} \frac{p_j(q_j/p_j - 1)^2}{q_j/p_j + (q_j/p_j)^{1/3}} \\
&\leq \sum_{q_j \leq p_j} p_j \log \frac{q_j}{p_j} \\
&\leq \sum_{q_j \leq p_j} q_j - \sum_{q_j \leq p_j} p_j - \sum_{q_j \leq p_j} \frac{p_j(q_j/p_j - 1)^2}{q_j/p_j + 1} \\
&\leq \sum_{q_j \leq p_j} q_j - \sum_{q_j \leq p_j} p_j - \sum_{q_j \leq p_j} \frac{q_j(q_j - p_j)^2}{q_j^2 + m_j^{1/3} M_j^{2/3}}.
\end{aligned} \tag{6}$$

Next, let $q_j > p_j$. Using (5) leads to

$$\frac{q_j}{p_j} - 1 - \frac{(q_j/p_j - 1)^2}{q_j/p_j + 1} < \log \frac{q_j}{p_j} < \frac{q_j}{p_j} - 1 - \frac{(q_j/p_j - 1)^2}{q_j/p_j + (q_j/p_j)^{1/3}}.$$

Again we multiply by p_j and sum up. Then we obtain

$$\begin{aligned}
&\sum_{q_j>p_j} q_j - \sum_{q_j>p_j} p_j - \sum_{q_j>p_j} \frac{q_j(q_j-p_j)^2}{q_j^2+m_j^{2/3}M_j^{1/3}} \\
&\qquad < \sum_{q_j>p_j} q_j - \sum_{q_j>p_j} p_j - \sum_{q_j>p_j} \frac{p_j(q_j/p_j-1)^2}{q_j/p_j+1} \\
&\qquad < \sum_{q_j>p_j} p_j \log\frac{q_j}{p_j} \\
&\qquad < \sum_{q_j>p_j} q_j - \sum_{q_j>p_j} p_j - \sum_{q_j>p_j} \frac{p_j(q_j/p_j-1)^2}{q_j/p_j+(q_j/p_j)^{1/3}} \\
&\qquad = \sum_{q_j>p_j} q_j - \sum_{q_j>p_j} p_j - \sum_{q_j>p_j} \frac{q_j(q_j-p_j)^2}{q_j^2+m_j^{1/3}M_j^{2/3}}.
\end{aligned} \tag{7}$$

Combining (6) and (7) gives

$$\begin{aligned}
&\sum_{j=1}^n q_j - \sum_{j=1}^n p_j - \sum_{j=1}^n \frac{q_j(q_j-p_j)^2}{q_j^2+m_j^{2/3}M_j^{1/3}} \\
&\qquad \le \sum_{j=1}^n p_j \log\frac{q_j}{p_j} \le \sum_{j=1}^n q_j - \sum_{j=1}^n p_j - \sum_{j=1}^n \frac{q_j(q_j-p_j)^2}{q_j^2+m_j^{1/3}M_j^{2/3}}.
\end{aligned} \tag{8}$$

Since $\sum_{j=1}^n p_j = \sum_{j=1}^n q_j$, we conclude from (8) that (3) is valid with $\alpha = 1/3$ and $\beta = 2/3$.

It remains to prove that if (3) holds for all $p_j, q_j > 0$ $(j = 1, ..., n)$ with $\sum_{j=1}^n p_j = \sum_{j=1}^n q_j$, then $\alpha \le 1/3$ and $\beta \ge 2/3$.

Let $s, t \in \mathbb{R}$ with $1 < t < s+1$. We set

$$p_1 = \frac{s}{t}, \quad p_2 = \frac{1}{t}, \quad q_1 = \frac{s+1}{t} - 1, \quad q_2 = 1, \quad p_j = q_j \quad (j = 3, ..., n).$$

Then we have

$$\sum_{j=1}^n p_j = \sum_{j=1}^n q_j, \quad m_1 = q_1^2, \quad M_1 = p_1^2, \quad m_2 = p_2^2, \quad M_2 = q_2^2.$$

A short calculation reveals that (3) is equivalent to

$$F_\alpha(s,t) \le s\log\frac{s}{s+1-t} - \log t \le F_\beta(s,t),$$

where

$$F_c(s,t) = \frac{(t-1)^2}{s+1-t+s^{2(1-c)}(s+1-t)^{2c-1}} + \frac{(t-1)^2}{t+t^{1-2c}}.$$

We define
$$G_c(s,t) = s\log\frac{s}{s+1-t} - \log t - F_c(s,t).$$
Then,
$$G_c(s,1) = \frac{\partial}{\partial t}G_c(s,t)\Big|_{t=1} = \frac{\partial^2}{\partial t^2}G_c(s,t)\Big|_{t=1} = 0$$
and
$$\frac{s^2}{3(s^2+1)}\frac{\partial^3}{\partial t^3}G_c(s,t)\Big|_{t=1} = \frac{s^2+2}{3(s^2+1)} - c.$$
Since
$$\lim_{s\to 0}\frac{s^2+2}{3(s^2+1)} = \frac{2}{3} \quad \text{and} \quad \lim_{s\to\infty}\frac{s^2+2}{3(s^2+1)} = \frac{1}{3},$$
we conclude from $G_\alpha(s,t) \geq 0$ that $\alpha \leq 1/3$ and from $G_\beta(s,t) \leq 0$ that $\beta \geq 2/3$. □

REMARK 2.3. *The proof of the Theorem reveals that if $\alpha \leq 1/3$ and $\beta \geq 2/3$, then the sign of equality holds in each inequality of* (3) *if and only if $p_j = q_j$ $(j = 1, ..., n)$.*

Acknowledgements

I thank the referee for helpful comments.

REFERENCES

[1] A. CLIM AND P. COZMA, *A new inequality of Shannon type*, Proc. Rom. Acad. Ser A **11** (2010), 294–302.
[2] T. M. COVER AND J. A. THOMAS, *Elements of Information Theory*, Wiley, 2006.
[3] G. T. HALLIWELL AND P. R. MERCER, *A refinement of an inequality from information theory*, J. Inequal. Pure Appl. Math. **5(1)** (2004), article 3, 3 pages.
[4] PL. KANNAPPAN AND P. K. SAHOO, *On a generalization of the Shannon functional inequality*, J. Math. Anal. Appl. **140** (1989), 341–350.
[5] D. S. MITRINOVIĆ, *Analytic Inequalities*, Springer, New York, 1970.

Author's address:

Horst Alzer
Morsbacher Str. 10
51545 Waldbröl, Germany
E-mail: h.alzer@gmx.de

Received September 26, 2014
Revised October 14, 2014

Rend. Istit. Mat. Univ. Trieste
Volume 46 (2014), 237–254

On Grothendieck's counterexample to the Generalized Hodge Conjecture

Dario Portelli

Communicated by Emilia Mezzetti

Abstract. *For a smooth complex projective variety X, let N^p and F^p denote respectively the coniveau filtration on $H^i(X,\mathbb{Q})$ and the Hodge filtration on $H^i(X,\mathbb{C})$. Hodge proved that $N^pH^i(X,\mathbb{Q}) \subset F^pH^i(X,\mathbb{C}) \cap H^i(X,\mathbb{Q})$, and conjectured that equality holds. Grothendieck exhibited a threefold X for which the dimensions of $N^1H^3(X,\mathbb{Q})$ and $F^1H^3(X,\mathbb{C}) \cap H^3(X,\mathbb{Q})$ differ by one. Recently the point of view of Hodge was somewhat refined (Portelli, 2014), and we aimed to use this refinement to revisit Grothendieck's example. We explicitly compute the classes in this second space which are not in $N^1H^3(X,\mathbb{Q})$. We also get a complete clarification that the representation of the homology customarily used for complex tori does not allow to apply the methods of (Portelli, 2014) to give a different proof of $N^1H^3(X,\mathbb{Q}) \subsetneq F^1H^3(X,\mathbb{C}) \cap H^3(X,\mathbb{Q})$.*

Keywords: Cohomology classes, supports, generalized Hodge conjecture.
MS Classification 2010: 14C30, 14C25.

1. Introduction

First of all, let us quickly recall the Generalized Hodge Conjecture.

Let X be a projective n-dimensional variety over $\mathbb{C}$, smooth and connected. To understand the algebraic geometry of X it is certainly of great interest the knowledge of the cohomology classes $\xi \in H^i(X,\mathbb{Q})$ for which there exists an algebraic subvariety $Y \subset X$ such that the image of ξ in the map $H^i(X,\mathbb{Q}) \to H^i(X-Y,\mathbb{Q})$ induced by the inclusion $X - Y \subset X$, is zero. We will say that ξ is *supported* by Y, or that Y is a *support* for ξ.

For any fixed integer $p \geq 0$ we can then consider the subspace of $H^i(X,\mathbb{Q})$

$$N^pH^i(X,\mathbb{Q}) := \sum_{\substack{Y \subset X \ \text{Zariski closed} \\ \text{codim}\, Y \geq p}} Ker\big(H^i(X,\mathbb{Q}) \to H^i(X-Y,\mathbb{Q}) \big) .$$

When p varies the $N^pH^i(X,\mathbb{Q})$ form a decreasing filtration of $H^i(X,\mathbb{Q})$, the so called coniveau filtration. Working with homology instead of cohomology, this filtration was introduced by Hodge ([3, p. 213]). He always deal with single classes instead of spaces [1].

In particular, Hodge gave the following necessary condition for a class $\xi \in H^i(X,\mathbb{Q})$ to be supported by an algebraic subvariety $Y \subset X$, of codimension $\geq p$ (for details and the proof, see § 2, Proposition 2.1). Assume that a singular $(2n-i)$-cycle Γ is homologous to a $(2n-i)$-cycle, contained into a subvariety Y as above. Let α be any *closed* $(2n-i)$-form on X, which contains in any of its local expressions at least $n-p+1$ dz's. Then $\int_\Gamma \alpha = 0$ (under suitable smoothness conditions on Γ). The translation between homology and cohomology is made by means of the Poincaré Duality, and it is implicit in the above statement that $[\Gamma]$ and ξ are Poincaré duals each other.

As we will see, in modern terms this amounts to say that $N^pH^i(X,\mathbb{Q})$ is contained into $F^pH^i(X,\mathbb{C})$, where $F^pH^i(X,\mathbb{C}) = \oplus_{a\geq p} H^{a\,,\,i-a}(X)$ is a space of the Hodge filtration on $H^i(X,\mathbb{C})$. Therefore we can conclude

$$N^pH^i(X,\mathbb{Q}) \subset F^pH^i(X,\mathbb{C}) \cap H^i(X,\mathbb{Q}) \ . \tag{1}$$

After this, Hodge raised a "problem" ([3, p. 214]; see also [4], where these contents of [3] have been presented to a large audience) of whether the above inclusion is, actually, an equality. Over the years this problem has become known as the Generalized Hodge Conjecture (from now on GHC, for short). If $i = 2p$ in (1), then the conjectured equality is the ordinary Hodge Conjecture.

Twenty eight years after [3], Grothendieck exhibited in [2] a particular abelian threefold X for which (1) is a strict inclusion, thus answering Hodge's question in the negative.

However, Grothendieck also showed that it is possible to correct the GHC simply by asking whether $N^pH^i(X,\mathbb{Q})$ (instead of being equal to $F^pH^i(X,\mathbb{C}) \cap H^i(X,\mathbb{Q})$) is the *maximal* rational sub-Hodge structure of $H^i(X,\mathbb{Q})$ which is contained into $F^pH^i(X,\mathbb{C})$. Of course, the abelian threefold X satisfies this amended GHC.

Let us remark here that, although Grothendieck also gives a variant of his amended GHC valid for a single class [2], he puts the major emphasis in a direct comparison between the *spaces* at the left and right hand sides of (1).

---

[1] However, it seems to us that his explicit concern to allow Y reducible (loc. cit.) is an indication that he was aware of the fact that these classes could be added together to obtain still classes of the same type.

[2] He wrote in [2, p. 300], "... *, an element of* $H^i(X,\mathbb{C})$ *should belong to* $N^pH^i(X,\mathbb{Q})$ (he certainly meant here: to the complexification of this space, N.d.A.) *if and only if all its bihomogeneous components* (i.e. the components with respect to the Hodge decomposition of $H^i(X,\mathbb{C})$, N.d.A.) *belong to the* $\mathbb{C}$*-vector space spanned by the right hand side of* (1)."

The ordinary Hodge Conjecture does not require to be corrected.

Finally, to put the present paper in the right perspective we have to spend a few words about the content of [6], where we assumed a point of view which is close to that of Hodge. The starting point of that paper was the remark that, if we assume from the beginning that Γ is contained into the subvariety Y, then $\int_\Gamma \alpha = 0$ holds true for *any* r-form α on X, *not necessarily closed*, which contains in its local expressions at least $n-p+1\ \ dz$'s.

It is not clear to us whether Hodge was aware or not of this fact. On the one hand, he considered (limits of) integrals of a given form on suitably "small" sets, see e.g. [3, pp. 113–115], to express punctual properties of the form. On the other hand, he was looking for a handy criterion to check whether a homology class on X is algebraic or not. And the model for such a criterion was, undoubtely, Lefschetz theorem on $(1,1)$-classes.

However, starting from the above remark, in [6] we proved that, if Γ is a suitable r-cycle such that $\int_\Gamma \alpha = 0$ for *every* r-form containing at least $n-p+1$ dz's, then Γ is contained in an algebraic subvariety Y of X, of codimension $\geq p$.

To prove this, the main ingredients in the proof are the following. First of all, our complex projective variety X can be thought as a smooth, compact *real* algebraic variety, of real dimension $2n$. It is rather well known that these varieties can be triangulated into simplexes which are real-analytic and semi-algebraic. Moreover, it is necessary to use only certain peculiar systems of local holomorphic coordinates, of essentially (complex) algebraic-geometric nature.

All this may give rise to the feeling that, perhaps, it is rather difficult to apply the results of [6] to deal with some concrete case of the GHC.

This paper is a first attempt to make such an application. More precisely, we analized if it is possible to check that the GHC fails for the Grothendieck's example, by an argument exclusively based on homology, in the spirit of [6]. It turns out that the customary representation of the homology classes in the case of complex tori, which we used throughout in the paper, is completely inadequate for this purpose. In a certain sense, this confirms the above feeling.

The content of the paper is as follows.

In the first section we examine in detail the two main steps which lead to the Generalized Hodge Conjecture as amended by Grothendieck, namely Hodge's necessary condition for a cohomology class to belong to $N^pH^i(X,\mathbb{Q})$ and the translation from homology to cohomology of the whole stuff. Grothendieck's example X is briefly introduced at the end of the section. In the next three sections we undertake a thorough analysis of this example. More precisely, in § 2, the homology and cohomology of X are quickly recalled for the reader convenience, ant to fix notations. In § 3, we determine a basis for the vector space $H^2(X,\mathbb{Q}) \cap H^{1,1}(X)$ over $\mathbb{Q}$. Moreover, for every element of such a basis, we determine a smooth, integral surface in X representing such class. In this way we obtain the surfaces $S_1, S_2, S_3, T_1, T_2, T_3$ of X. The contribution of all these surface to $N^1H^3(X,\mathbb{Q})$ is computed in § 4. In § 5 we compute $F^1H^3(X,\mathbb{C}) \cap H^3(X,\mathbb{Q})$, thus completing the examination of Grothendieck's counterexample to the GHC. In the last section we compare the Poincaré duals of two classes inside $F^1H^3(X,\mathbb{C}) \cap H^3(X,\mathbb{Q})$, only one of which belongs to $N^1H^3(X,\mathbb{Q})$, while the other does not. The representation of the homology classes used in the paper reveals to be completely inefficient to detect the difference between the two.

2. The Generalized Hodge Conjecture, from Hodge to Grothendieck

From now on we will set $r = 2n - i$.

Let us start with Hodge's result, which is the following:

PROPOSITION 2.1. *Let $Y \subset X$ be an algebraic subvariety, of codimension $\geq p$. Consider a class $[\Gamma] \in H_r(Y,\mathbb{Q})$. We can assume without loss of generality that the singular r-cycle Γ is a linear combination of $\mathscr{C}^\infty$ singular r-simplexes. Let α be any closed r-form on X such that every term of the expression of α in any system of local holomorphic coordinates contains at least $n-p+1$ dz's. Then*

$$\int_\Gamma \alpha = 0 . \tag{2}$$

Proof. First of all, notice that the image of $[\Gamma]$ in the canonical map $H_r(Y,\mathbb{Q}) \to H_r(X,\mathbb{Q})$ induced by the inclusion $Y \subset X$ can be represented by the same cycle Γ.

Notice moreover that the form which is actually integrated here is the pull-back of α to the various singular simplexes of Γ. An intermediate step in this pull-back procedure is the pull-back of α to Y_{sm}. But α contains too many dz's to be supported by Y_{sm}, hence

$$\alpha_{|_{Y_{sm}}} \equiv 0$$

and (2) is proved.

A possible doubt here is that some simplex S of Γ can be contained, actually, into the singular locus of Y. But the codimension of $Sing(Y)$ is $\geq p+1$, hence we can use the above argument with $Sing(Y)$ at the place of Y, to conclude that $\int_S \alpha = 0$. $\square$

Remark 2.2. *The argument given above to prove (2) does not use the assumption that α is a closed form. Only the dimension of Y and the type of the form α are relevant.*

To deal with the Generalized Hodge Conjecture, Grothendieck had the idea to translate everything from homology to cohomology. The device for this is the Poincaré duality isomorphism (recall that $i = 2n - r$)

$$PD : H_r(X, \mathbb{C}) \to H^i(X, \mathbb{C}) \ ,$$

which works as follows. Fix $[\Gamma] \in H_r(X, \mathbb{C})$, and let $j : \Gamma \to X$ denote the inclusion (*cum grano salis*, because Γ is a cycle). Then we have a well defined $\mathbb{C}$-linear map

$$\lambda_{[\Gamma]} : H^r(X, \mathbb{C}) \to \mathbb{C} \qquad \text{given by} \qquad [\omega] \mapsto \int_\Gamma j^*\omega \ .$$

Therefore, thanks to the canonical perfect pairing

$$\Psi : H^r(X, \mathbb{C}) \times H^i(X, \mathbb{C}) \to \mathbb{C} \ , \qquad ([\omega], [\omega']) \mapsto \int_X \omega \wedge \omega' \ , \tag{3}$$

there is one and only one $[\xi] \in H^i(X, \mathbb{C})$ such that $\lambda_{[\Gamma]} = \Psi(\_, [\xi])$. In more down-to-earth terms

$$\int_\Gamma j^*\omega = \int_X \omega \wedge \xi \tag{4}$$

for any closed r-form ω.

The class $[\xi] \in H^i(X, \mathbb{C})$ is called the Poincaré dual of $[\Gamma] \in H_r(X, \mathbb{Q})$.

The attentive reader had certainly noticed that, to introduce Poincaré duality as above, one represents cohomology classes à la de Rham, i.e. by mean of closed forms. This forces us to use cohomology with *complex* coefficients. How to deal with *rational* cohomology classes in this set-up?

Recall that, if η is a closed differential s-form on X, then a *period* of η is any complex number of the form

$$\int_\Gamma \eta \ ,$$

where Γ is a s-cycle with *integral* coefficients. Then, η represents a class in $H^s(X, \mathbb{Q})$ ($\subset H^s(X, \mathbb{C})$) if and only if all its periods are in $\mathbb{Q}$ ([8, pp. 34–35]).

With these last preparations, we have at hand everything we need to translate in cohomological terms Hodge's necessary condition.

Given a proper map $f : Z \to X$, where Z is a smooth complex projective variety, equidimensional of dimension $t \leq n - p$, Poincaré duality allows us to define the Gysin map $f_* : H^{2t-r}(Z, \mathbb{Q}) \to H^i(X, \mathbb{Q})$ as the composition

$$H^{2t-r}(Z, \mathbb{Q}) \xrightarrow{PD} H_r(Z, \mathbb{Q}) \xrightarrow{f_* = can.} H_r(X, \mathbb{Q}) \xrightarrow{PD} H^i(X, \mathbb{Q}) \ .$$

For this Gysin map we have the *exact* sequence ([1, Coroll. (8.2.8)])

$$H^{2t-r}(Z, \mathbb{Q}) \xrightarrow{f_*} H^i(X, \mathbb{Q}) \longrightarrow H^i(X - f(Z), \mathbb{Q}) \ .$$

Let us remark that $Y := f(Z)$ is Zariski closed inside X because f is proper. Moreover, the map f can be thought of as the composition of a resolution of the singularities $Z \to f(Z)$ of $f(Z) =: Y$ with the inclusion $Y \subset X$. All this shows that the coniveau filtration on $H^i(X, \mathbb{Q})$ is also given by

$$N^p H^i(X, \mathbb{Q}) \ = \sum_{f \ as \ above} Im(f_*) \ .$$

Finally, a simple weight argument shows that (for more details the reader is referred to [6])

$$Im\big(f_* : H^{2t-r}(Z, \mathbb{Q}) \to H^i(X, \mathbb{Q})\big) = PD\Big(Im\big(H_r(Y, \mathbb{Q}) \to H_r(X, \mathbb{Q})\big)\Big) \ . \quad (5)$$

Of course, Hodge's was concerned to characterize the classes in $H_r(X, \mathbb{Q})$ which were in the image of some map $H_r(Y, \mathbb{Q}) \to H_r(X, \mathbb{Q})$.

The assumption on α in the statement of Proposition 2.1 can be simply rephrased saying that $[\alpha] \in F^{n-p+1} H^r(X, \mathbb{C})$. The relation (2) then becomes, thanks to (4),

$$\int_X \alpha \wedge \xi \ = \ 0$$

for every $[\alpha] \in F^{n-p+1} H^r(X, \mathbb{C})$, i.e.

$$PD([\Gamma]) \ = \ [\xi] \ \in \ \big(F^{n-p+1} H^r(X, \mathbb{C}) \big)^{\perp} \ ,$$

where the orthogonal subspace is taken with respect to the canonical perfect pairing (3). But it is easily computed that

$$\big(F^{n-p+1} H^r(X, \mathbb{C}) \big)^{\perp} \ = \ F^p H^i(X, \mathbb{C}) \ . \qquad (6)$$

In fact, if $[\omega] \in H^{a\,,\,i-a}(X)$ with $a \geq p$, then $n - p + 1 + a > n$ and therefore we have trivially $[\omega] \wedge F^{n-p+1} H^r(X, \mathbb{C}) \ = \ 0$. On the other hand, if $[\omega] \in$

$H^{a,i-a}(X)$ with $a < p$, then $*\overline{[\omega]} \in F^{n-p+1}H^r(X,\mathbb{C})$ and it is well known that

$$\int_X [\omega] \wedge *\overline{[\omega]} > 0 \ .$$

Therefore the complete translation into cohomology of Proposition 2.1 amounts to the inclusion (1).

An advantage of the coomological translation is that the Gysin maps f_* *are maps of rational Hodge structures* ([7, 7.3.2]). Hence, for any proper map $f : Z \to X$, where Z is a smooth complex projective variety, equidimensional of dimension $t \leq n-p$, we have that $Im(f_*)$ is a rational sub-Hodge structure of $H^i(X,\mathbb{Q})$. Then, by general facts on the category of (pure) rational Hodge structures, the space $N^pH^i(X,\mathbb{Q})$ is also a rational sub-Hodge structure of $H^i(X,\mathbb{Q})$.

But then, from $N^1 H^3(X,\mathbb{Q}) \subset F^1 H^3(X,\mathbb{C})$ it follows that the Hodge decomposition of $N^1 H^3(X,\mathbb{Q}) \otimes_{\mathbb{Q}} \mathbb{C}$ has necessarily the form

$$N^1 H^3(X,\mathbb{Q}) \otimes_{\mathbb{Q}} \mathbb{C} = U^{1,2} \oplus U^{2,1} \ ,$$

for suitable complex subspaces $U^{1,2}$ and $U^{2,1}$ of $H^3(X,\mathbb{C})$. Finally, since we have $\overline{U^{1,2}} = U^{2,1}$, we conclude that *the dimension of $N^1 H^3(X,\mathbb{Q})$ is even.*

On the other hand, in Grothendieck's example the right hand side of (1) is odd-dimensional. The example is constructed as follows.

Let E be an elliptic curve over the field of complex numbers. The projective manifolds we are interested in are the abelian threefolds

$$X := E \times E \times E = E^3 \ .$$

More precisely, X can be defined as follows. Let e_1, e_2, e_3 denote the standard basis of $\mathbb{C}^3$, and let z_1, z_2, z_3 denote the corresponding complex coordinates. Fix a complex number $\tau = u + iv$, where $u, v \in \mathbb{R}$, with $v > 0$. Then $e_1, e_2, e_3, \tau e_1, \tau e_2, \tau e_3$ is an (ordered) basis of a lattice $\Lambda \simeq \mathbb{Z}^6$ contained into $\mathbb{C}^3$, and we set

$$X := \mathbb{C}^3/\Lambda \ .$$

So everything depends on the choice of τ. To avoid some minor difficulties, we will assume from now on that $[\mathbb{Q}(\tau) : \mathbb{Q}] \geq 3$. Of course, we do not exclude that τ may be a transcendental number, but if τ is algebraic, then its degree is not 2.

3. The topology of X

The homology and cohomology of complex tori is a completely standard topic. Hence this section is just for the reader convenience, and to fix the notations.

We will denote by $u_1, u_2, u_3, u_4, u_5, u_6$ the real coordinates in $\mathbb{C}^3$ with respect to the basis of Λ fixed above, namely

$$z_h = u_h + \tau\, u_{h+3} , \qquad h = 1,2,3 . \tag{7}$$

Concerning the topology of X, let us consider integral homology first.

Let $I = [0,1] \subset \mathbb{R}$, and define maps $\gamma_i : I \to \mathbb{C}^3$ by setting

$$\gamma_i(t) = t\, e_i \quad \text{for} \quad i = 1,2,3 \qquad \text{and} \qquad \gamma_i(t) = t\, \tau\, e_{i-3} \quad \text{for} \quad i = 4,5,6 .$$

If we compose these γ_i with the canonical map $\pi : \mathbb{C}^3 \to X$ we get six singular 1-cycles of X, whose classes are a basis for the free abelian group $H_1(X, \mathbb{Z})$. So inside X there are six copies of S^1, the images of the $\pi \circ \gamma_i$; we will denote them by $C_1, \ldots, C_6$. It is well known that a basis for $H_r(X, \mathbb{Z})$ is then given by the classes of all the r-cycles

$$C_{i_1} \times C_{i_2} \times \ldots \times C_{i_r} \qquad \text{where} \quad 1 \leq i_1 < i_2 < \ldots < i_r \leq 6 \ . \tag{8}$$

Now we turn to the cohomology with complex coefficients of X.

A basis for $H^r(X, \mathbb{C})$ is given by the classes of the closed r-forms

$$d_{H} = du_{h_1 h_2 \ldots h_r} = du_{h_1} \wedge du_{h_2} \wedge \ldots \wedge du_{h_r} , \tag{9}$$

where $H = (\, h_1 h_2 \ldots h_r \,)$ is a multi-index, and $1 \leq h_1 < h_2 < \ldots < h_r \leq 6$. A straightforward computation then shows that

$$\int_{C_{i_1} \times C_{i_2} \times \ldots \times C_{i_r}} du_{h_1} \wedge du_{h_2} \wedge \ldots \wedge du_{h_r} = \delta^{h_1}_{i_1}\, \delta^{h_2}_{i_2} \ldots \delta^{h_r}_{i_r} , \tag{10}$$

where the δ's are Kronecker's. As remarked in the previous section, then *the classes of the forms* (9) *are also a basis for* $H^r(X, \mathbb{Q})$ *over* $\mathbb{Q}$.

EXAMPLE 3.1. *For future use, let us show how goes the computation (10), at least in a particular case, for* $r = 3$. *Parametrize* $C_2 \times C_4 \times C_5$ *by first defining*

$$\varphi : [0,1]^3 \to \mathbb{C}^3 \qquad \varphi : (t_1, t_2, t_3) \mapsto t_1\, e_2 + t_2\, \tau\, e_1 + t_3\, \tau\, e_2$$

and then composing with the canonical map $\pi : \mathbb{C}^3 \to X$. *Namely we have*

$$u_1 = 0 \qquad u_2 = t_1 \qquad u_3 = 0 \qquad u_4 = t_2 \qquad u_5 = t_3 \qquad u_6 = 0 .$$

Actually, we can consider φ *defined in an open neighborhood of* $[0,1]^3$ *inside* $\mathbb{R}^3$, *and therefore*

$$\varphi^* \big(du_2 \wedge du_4 \wedge du_5 \big) = dt_1 \wedge dt_2 \wedge dt_3 ,$$

which yields the result.

We turn now to the Hodge decomposition of the spaces $H^r(X,\mathbb{C})$, and to their relations with $H^r(X,\mathbb{Q})$.

For our purposes we have to use also the basis $dz_1, \dots, d\bar{z}_3$ of $H^1(X,\mathbb{C})$. Because of (7), the simple relations between the dz_h, $d\bar{z}_k$ and the du_j are

$$dz_h = du_h + \tau\, du_{h+3} \qquad d\bar{z}_h = du_h + \bar{\tau}\, du_{h+3} \tag{11}$$

and

$$\begin{aligned} du_h &= \left(\tfrac{1}{2} + i\,\tfrac{u}{2v}\right) dz_h + \left(\tfrac{1}{2} - i\,\tfrac{u}{2v}\right) d\bar{z}_h \quad , \\ du_{h+3} &= -\tfrac{i}{2v}\, dz_h + \tfrac{i}{2v}\, d\bar{z}_h \end{aligned} \tag{12}$$

for any $h = 1,2,3$.

Finally, we can compute the various classes $PD([C_i \times C_j \times C_k])$ with respect to the basis (9) by means of formula (4). To be precise, assume that $\{i,j,k,l,m,n\} = \{1,2,3,4,5,6\}$, that $1 \le i < j < k \le 6$ and that $1 \le l < m < n \le 6$. Moreover, denote by σ the permutation

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ i & j & k & l & m & n \end{pmatrix} .$$

Then it is easily checked that

$$PD([C_i \times C_j \times C_k]) = (-1)^{sign(\sigma)+1}\, du_{lmn} . \tag{13}$$

4. Divisors on X

To test Hodge's and Grothendieck's guesses on X we have to produce elements of $N^1 H^3(X,\mathbb{Q})$. This requires a rather detailed knowledge of the surfaces on X.

PROPOSITION 4.1. *A basis of the $\mathbb{Q}$-module $H^2(X,\mathbb{Q}) \cap H^{1,1}(X)$ is given by the classes*

$$\frac{i}{2v}\, dz_h \wedge d\bar{z}_h = du_h \wedge du_{h+3} \quad , \qquad h = 1,2,3 \tag{14}$$

and

$$\frac{i}{2v}\left(dz_h \wedge d\bar{z}_k + dz_k \wedge d\bar{z}_h\right) = du_h \wedge du_{k+3} + du_k \wedge du_{h+3} \quad , \qquad h = 1,2,3 . \tag{15}$$

Hence

$$dim_{\mathbb{Q}}\left(H^2(X,\mathbb{Q}) \cap H^{1,1}(X)\right) = 6 .$$

Proof. Consider the closed 2-form

$$F \;=\; \sum_{1\le h,k\le 3} a_{hk}\, dz_h \wedge d\bar z_k \;=\; \sum_{1\le s<t\le 6} b_{st}\, du_s \wedge du_t \;, \tag{16}$$

where the a_{hk} and b_{st} are all in $\mathbb{C}$. By (10), we have that $[F]\in H^2(X,\mathbb{Q})$ if and only if all the b_{st} are in $\mathbb{Q}$. Murasaki's idea in [5] is to write

$$F \;=\; F_1 + F_2 + F_3 + F_{12} + F_{13} + F_{23} \;, \tag{17}$$

where, for every $h = 1,2,3$,

$$F_h \;:=\; a_{hh}\, dz_h \wedge d\bar z_h$$

and for every $1\le h<k\le 3$,

$$F_{hk} \;:=\; a_{hk}\, dz_h \wedge d\bar z_k + a_{kh}\, dz_k \wedge d\bar z_h \;.$$

Lemma 4.2. *F represents a rational cohomology class if and only if all the F_h and the F_{hk} represent rational cohomology classes.*

Proof. One direction is obvious, so assume that F represents a rational cohomology class. From (11) we get the relations

$$F_h \;=\; -\,2iv\, a_{hh}\; du_h \wedge du_{h+3}$$

$$\text{and} \tag{18}$$

$$\begin{aligned} F_{hk} \;\;=\;\; & (\, a_{hk} - a_{kh}\,)\, du_h \wedge du_k \;+\; (\, a_{hk}\,\bar\tau - a_{kh}\,\tau\,)\, du_h \wedge du_{k+3} + \\ & + (\, a_{kh}\,\bar\tau - a_{hk}\,\tau\,)\, du_k \wedge du_{h+3} \;+\; (\, a_{hk} - a_{kh}\,)\,\tau\,\bar\tau\, du_{h+3}\wedge du_{k+3} \;. \end{aligned}$$

This shows that each of the six terms in (17) involves different elements of the basis $du_i \wedge du_j$ of $H^2(X,\mathbb{C})$, hence the lemma is completely proved. □

Now, the first equation of (18) yields by (16) that

$$-\,2iv\, a_{hh} \;=\; b_{h,h+3} \;.$$

Therefore $b_{h,h+3}\in\mathbb{Q}$ if and only if

$$a_{hh} \;=\; \frac{i}{2\,v}\, r \qquad \text{where} \quad r\in\mathbb{Q} \;.$$

In other words, all the classes (14) are in $H^2(X,\mathbb{Q})\cap H^{1,1}(X)$, and they are independent over $\mathbb{Q}$.

Concerning the class F_{hk}, for any fixed $1 \le h < k \le 3$, the second relation in (18) implies that all the following numbers are rational:

$$a_{hk} - a_{kh} \quad , \qquad (\, a_{hk} - a_{kh}\,)\, \tau\, \bar{\tau} \quad , \qquad a_{hk}\, \bar{\tau} - a_{kh}\, \tau \quad , \qquad a_{kh}\, \bar{\tau} - a_{hk}\, \tau \ .$$

From this we get, in particular

$$(\, a_{hk} - a_{kh}\,)\, (\, \tau + \bar{\tau}\,) \,\in\, \mathbb{Q} \ .$$

Therefore, *if* $a_{hk} - a_{kh} \neq 0$, *then necessarily* τ *is an algebraic number over* $\mathbb{Q}$, *of degree* $[\mathbb{Q}(\tau) : \mathbb{Q}] \le 2$. We ruled out this possibility at the end of § 2.

Hence if F is rational, then necessarily $a_{hk} = a_{kh} = a$ and

$$F_{hk} \,=\, a\, (\, \bar{\tau} - \tau\,) \left(du_h \wedge du_{k+3} + du_k \wedge du_{h+3} \right) \,=$$

$$=\, -\, 2\, i\, a\, v \left(du_h \wedge du_{k+3} + du_k \wedge du_{h+3} \right) \ .$$

We get in this way the classes (15) of $H^2(X, \mathbb{Q}) \cap H^{1,1}(X)$, which are linearly independent over $\mathbb{Q}$, and are also independent of the classes (14). □

Let us consider the divisors now, i.e. the surfaces on X.

First of all, the abelian surface $E \times E$ can be embedded in X in a trivial way by setting, for an arbitrary $P \in E$:

$$S_3 \,:=\, E \times E \times P\, .$$

The family $\{\, E \times E \times P\,\}_{P \in E}$ is a fibration of X. Moreover, if $P, Q \in E$, then $E \times E \times P$ and $E \times E \times Q$ are algebraically equivalent, hence they are homologically equivalent.

To determine the cohomology class of S_3, we remark that as a singular 4-cycle inside X we have $S_3 = C_1 \times C_2 \times C_4 \times C_5$. Then, for any closed 4-form

$$\alpha \,=\, \sum_{\#I = 4} b_{_I}\, du_{_I} \quad , \qquad b_{_I} \in \mathbb{Q} \quad \text{ for any } \quad I \ ,$$

the relation (10) implies

$$\int_{S_3} \alpha \,=\, \int_{C_1 \times C_2 \times C_4 \times C_5} \alpha \,=\, b_{(1,2,4,5)} \ .$$

Therefore, (4) will be satisfied for every form α if we take

$$\omega_3 \,:=\, du_3 \wedge du_6 \,=\, \frac{i}{2\, v}\, dz_3 \wedge d\bar{z}_3 \ . \tag{19}$$

In other words
$$PD(\,[S_3]\,) \;=\; \frac{i}{2\,v}\,dz_3 \wedge d\bar{z}_3 \;. \tag{20}$$

On X we have also two other obvious families of surfaces, given respectively by
$$S_1 \;:=\; P \times E \times E \quad , \qquad\qquad S_2 \;:=\; E \times P \times E \;.$$

The corresponding cohomology classes, computed as above, are
$$PD(\,[S_1]\,) \;=\; \omega_1 \;:=\; \frac{i}{2\,v}\,dz_1 \wedge d\bar{z}_1 \quad , \qquad PD(\,[S_2]\,) \;=\; \omega_2 \;:=\; \frac{i}{2\,v}\,dz_2 \wedge d\bar{z}_2 \;.$$

To produce divisors not equivalent to the S_i's, we have to embed $E \times E$ inside X by using the diagonal map $\Delta : E \to E \times E$ like in
$$f : E \times E \xhookrightarrow{\Delta \times id_E} E \times E \times E \quad . \tag{21}$$

We will denote by T_3 the image of $E \times E$ in the proper map f. Two other surfaces T_1 and T_2 can be defined inside X as the images of the (proper) map
$$E \times E \xhookrightarrow{id_E \times \Delta} E \times E \times E \quad ,$$
and similarly for T_2.

We will determine now the cohomology class of the divisor T_3.

To compute the pull-back of forms it is better to describe the map f in (21) as follows. If $\varepsilon_1, \varepsilon_2$ is the standard basis of $\mathbb{C}^2$, we have the real basis $\varepsilon_1, \varepsilon_2, \tau\,\varepsilon_1, \tau\,\varepsilon_2$ of this space. It generates over the integers a lattice $L \subset \mathbb{C}^2$, and of course $E \times E = \mathbb{C}^2/L$. Moreover, let v_1, v_2, v_3, v_4 denote real coordinates in $\mathbb{C}^2$ with respect to $\varepsilon_1, \varepsilon_2, \tau\,\varepsilon_1, \tau\,\varepsilon_2$. Then the map f is induced by the map $\mathbb{C}^2 \to \mathbb{C}^3$ given in real coordinates by
$$u_1 = v_1 \;,\quad u_2 = v_1 \;,\quad u_3 = v_2 \;,\quad u_4 = v_3 \;,\quad u_5 = v_3 \;,\quad u_6 = v_4 \,. \tag{22}$$

First consequences of these relations are
$$f^*\,(\,du_1 \wedge du_2\,) \;=\; 0 \quad , \qquad f^*\,(\,du_4 \wedge du_5\,) \;=\; 0 \;.$$

Hence, given any rational closed 4-form $\alpha = b_{1234}\,du_1 \wedge du_2 \wedge du_3 \wedge du_4 + \ldots$ on X, we have
$$f^*\,\alpha \;=\; -\,(\,b_{1346} + b_{1356} + b_{2346} + b_{2356}\,)\,dv_1 \wedge dv_2 \wedge dv_3 \wedge dv_4 \;.$$

If we set
$$\eta_3 \;:=\; du_{15} + du_{24} - du_{14} - du_{25} \;, \tag{23}$$

it is easily checked that

$$\eta_3 \wedge \alpha = -(b_{1346} + b_{1356} + b_{2346} + b_{2356})\, du_1 \wedge du_2 \wedge \ldots \wedge du_6 \ ,$$

and we can conclude that

$$\int_{T_3} f^* \alpha = \int_X \eta_3 \wedge \alpha$$

for any rational closed 4-form α on X. Namely, we have that the Poincaré dual σ_3 of T_3 is represented by the closed form η_3, which can also be written as

$$\eta_3 = \frac{i}{2v}\left(dz_1 \wedge d\bar{z}_2 + dz_2 \wedge d\bar{z}_1 - dz_1 \wedge d\bar{z}_1 - dz_2 \wedge d\bar{z}_2 \right) , \tag{24}$$

thanks to (14) and (15).

5. Classes of $N^1H^3(X,\mathbb{Q})$

The purpose of this section is to compute the contribution to $N^1H^3(X,\mathbb{Q})$ of the surfaces $S_1, S_2, S_3, T_1, T_2, T_3$ on X, introduced in the previous section. More concretely, we will prove the

Proposition 5.1. *Consider the abelian threefold $X = E \times E \times E$, where E is the elliptic curve $\mathbb{C}/\mathbb{Z} + \tau\mathbb{Z}$. Here we assume that τ is a complex number with $\Im(\tau) > 0$, and such that $[\mathbb{Q}(\tau) : \mathbb{Q}] \geq 3$. Then, for every such threefold $N^1H^3(X,\mathbb{Q})$ contains a rational sub-Hodge structure M of $H^3(X,\mathbb{Q})$, with $dim_{\mathbb{Q}} M = 16$.*

Proof. Let i denote the inclusion $S_1 \subset X$. We start by computing the image of the Gysin map $i_* : H^1(S_1,\mathbb{Q}) \to H^3(X,\mathbb{Q})$. A basis for $H_3(S_1,\mathbb{Q})$ is given by the classes

$$[C_2 \times C_3 \times C_5] \ , \quad [C_2 \times C_3 \times C_6] \ , \quad [C_2 \times C_5 \times C_6] \ , \quad [C_3 \times C_5 \times C_6] \ .$$

They are sent by i_* into the same classes, viewed as elements of $H_3(X,\mathbb{Q})$. Finally, by (13) we conclude that $i_*(H^1(S_1,\mathbb{Q}))$ is generated by

$$du_{146} \ , \ du_{145} \ , \ du_{134} \ , \ du_{124} \ .$$

Similarly, bases for the images of the Gysin maps for the surfaces S_2 and S_3 are given respectively by

$$du_{256} \ , \ du_{245} \ , \ du_{235} \ , \ du_{125} \qquad \text{and} \qquad du_{356} \ , \ du_{346} \ , \ du_{236} \ , \ du_{136} \ .$$

These twelve 3-forms are distinct elements of the basis (9) of $H^3(X,\mathbb{Q})$; let M_1 denote the subspace they generate.

The contribution of the surfaces T_1, T_2 and T_3 to $N^1H^3(X,\mathbb{Q})$ is not easily determined in homology, so we switch directly to cohomology.

Consider the map f defined in (21) (or (22), in coordinates). The image of f was denoted by T_3; for simplicity, we will still denote by f the inclusion of T_3 into X.

LEMMA 5.2. *Recall that the closed form η_3 given in (23) represents the class $\sigma_3 = PD([T_3])$. The following diagram is then commutative*

$$\begin{tikzcd} H^1(X,\mathbb{Q}) \arrow[rr, "\sigma_3 \cup \_"] \arrow[dr, "f^*"'] & & H^3(X,\mathbb{Q}) \\ & H^1(T_3,\mathbb{Q}) \arrow[ur, "f_*"'] & \end{tikzcd} \quad . \tag{25}$$

Proof. of Lemma 5.2.

In fact, consider the following commutative diagram, where $[X] \in H_6(X,\mathbb{Q})$ and $[T_3] \in H_4(T_3,\mathbb{Q})$ are the fundamental classes of X and T_3 respectively

$$\begin{tikzcd} H^1(X,\mathbb{Q}) \arrow[r, "f^*"] & H^1(T_3,\mathbb{Q}) \arrow[r, "f_*"] \arrow[d, "\_ \cap [T_3]"', "\wr"] & H^3(X,\mathbb{Q}) \arrow[d, "\wr"', "\_ \cap [X]"] \\ & H_3(T_3,\mathbb{Q}) \arrow[r, "f_*"'] & H_3(X,\mathbb{Q}) \ . \end{tikzcd}$$

The classes σ_3 and $[T_3]$ are related by Poincaré duality in the following way

$$f_*[T_3] = \sigma_3 \cap [X] \ .$$

Then, for every $x \in H^1(X,\mathbb{Q})$ we have by the "projection formula" and the above relation

$$(f_* f^* x) \cap [X] = f_*(f^* x \cap [T_3]) = x \cap f_*[T_3] =$$
$$x \cap (\sigma_3 \cap [X]) = (x \cup \sigma_3) \cap [X] = (\sigma_3 \cup x) \cap [X] \ ,$$

where the last equality is true because the degree of σ_3 is 2. Since the Poincaré duality map is an isomorphism, the commutativity of (25) is completely proved. □

Now, the space $Im(\sigma_3 \cup \_)$ is generated by the classes of

$$\begin{aligned} \eta_3 \wedge du_1 \ = \ \eta_3 \wedge du_2 &= du_{124} - du_{125} \in M_1 \\ \eta_3 \wedge du_3 &= \underbrace{du_{134} + du_{235}}_{\in M_1} - du_{135} - du_{234} \ , \end{aligned}$$

$$\begin{aligned} \eta_3 \wedge du_4 \ = \ \eta_3 \wedge du_5 &= du_{245} - du_{145} \in M_1 \\ \eta_3 \wedge du_6 &= \underbrace{-du_{256} - du_{146}}_{\in M_1} + du_{246} + du_{156} \ . \end{aligned}$$

These relations shows that $Im(\sigma_3 \cup \_)$ has dimension four. But this forces f^* to be onto because $dim_{\mathbb{Q}} H^1(T_3, \mathbb{Q}) = 4$, hence $Im(f_*) = Im(\sigma_3 \cup \_)$.

To summarize, the contribution of $Im(f_*)$ to the generation of the space $N^1 H^3(X, \mathbb{Q})$ is given by the classes

$$du_{135} + du_{234} \quad , \qquad du_{246} + du_{156} \ .$$

Similar computations can be performed for the surfaces T_1 and T_2, which contribute to the generation of $N^1 H^3(X, \mathbb{Q})$ respectively with the classes

$$du_{135}+du_{126} \ , \quad du_{246}+du_{345} \qquad \text{and} \qquad du_{126}-du_{234} \ , \quad du_{156}-du_{345} \ .$$

Finally, denote by M_2 the subspace of $H^3(X, \mathbb{Q})$ generated by the six classes above. It is easily seen that a basis for M_2 is given by

$$du_{126} - du_{234} \ , \quad du_{156} - du_{345} \ , \quad du_{246} + du_{345} \ , \quad du_{135} + du_{234} \tag{26}$$

and that $M_1 \cap M_2 = 0$. Then $M := M_1 \oplus M_2$ is a rational sub-Hodge structure of $H^3(X, \mathbb{Q})$, contained into $N^1 H^3(X, \mathbb{Q})$. Since $dim_{\mathbb{Q}} M = 16$, the proof of Proposition 5.1 is complete. □

6. Computation of $F^1 H^3(X, \mathbb{C}) \cap H^3(X, \mathbb{Q})$

For this computation we will exploit (6) and the fact that $H^{3,0}(X)$ is isomorphic to $\mathbb{C}$, generated by the class of the following closed form α (which we give also in terms of the base (9))

$$\begin{aligned} dz_1 \wedge dz_2 \wedge dz_3 \quad = \quad & du_{123} + \tau\,(du_{126} - du_{135} + du_{234}) + \\ & + \tau^2\,(du_{156} - du_{246} + du_{345}) + \tau^3\, du_{456} \ . \end{aligned} \tag{27}$$

Then, for an arbitrary $\omega = \sum_{1 \le i < j < k \le 6} r_{ijk}\, du_{ijk}$ where $r_{ijk} \in \mathbb{Q}$ for any i, j, k, we have that

$$\omega \wedge \alpha = \big(r_{123}\,\tau^3 - (\, r_{234} - r_{135} + r_{126}\,)\,\tau^2 + (\, r_{345} - r_{246} + r_{156}\,)\,\tau - r_{456} \big)\, du_{123456} \ .$$

Hence $[\omega]$ is orthogonal to $[\alpha]$ with respect to (3) if and only if

$$r_{123}\,\tau^3 - (\, r_{234} - r_{135} + r_{126}\,)\,\tau^2 + (\, r_{345} - r_{246} + r_{156}\,)\,\tau - r_{456} \ = \ 0 \ . \tag{28}$$

At the end of § 2 we made the assumption $[\mathbb{Q}(\tau) : \mathbb{Q}] \geq 3$. Therefore, if τ is not algebraic over $\mathbb{Q}$, of degree 3, the above relation is satisfied only if all the coefficients in it vanish. Since the linear system

$$\left\{ \begin{array}{rcl} r_{123} & = & 0 \\ r_{234} - r_{135} + r_{126} & = & 0 \\ r_{345} - r_{246} + r_{156} & = & 0 \\ r_{456} & = & 0 \end{array} \right.$$

has rank four, we conclude $dim_{\mathbb{Q}}\left(H^3(X,\mathbb{Q}) \cap F^1 H^3(X,\mathbb{C})\right) = 16$, and we have the following straightforward consequence of Proposition 5.1:

PROPOSITION 6.1. *If* $[\mathbb{Q}(\tau):\mathbb{Q}] > 3$ *(in particular, if τ is transcendental over $\mathbb{Q}$), then*

$$N^1H^3(X,\mathbb{Q}) = F^1H^3(X,\mathbb{C}) \cap H^3(X,\mathbb{Q}) ,$$

i.e. the Generalized Hodge conjecture is true in its original form for X.

On the other hand, Grothendieck considered the case when τ is algebraic over $\mathbb{Q}$, of degree 3. Let $f := X^3 + \mu_1 X^2 + \mu_2 X + \mu_3$ be the minimal polynomial of τ over $\mathbb{Q}$. Then relation (28) can be rewritten as

$$(r_{234}-r_{135}+r_{126}+\mu_1 r_{123})\,\tau^2 - (r_{345}-r_{246}+r_{156}-\mu_2 r_{123})\,\tau + r_{456}+\mu_3 r_{123} = 0 .$$

Since $[\mathbb{Q}(\tau):\mathbb{Q}] = 3$, we have necessarily

$$\left\{\begin{array}{rcl} r_{234} - r_{135} + r_{126} + \mu_1 r_{123} & = & 0 \\ r_{345} - r_{246} + r_{156} - \mu_2 r_{123} & = & 0 \\ r_{456} + \mu_3 r_{123} & = & 0 \end{array}\right. .$$

This linear system has rank 3, hence

$$dim_{\mathbb{Q}}\Big(H^3(X,\mathbb{Q}) \cap F^1 H^3(X,\mathbb{C})\Big) = 17 , \tag{29}$$

and *the Generalized Hodge Conjecture fails in its original form.*

But, since we know a priori that the dimension of $N^1H^3(X,\mathbb{Q})$ is even, (29) forces $N^1H^3(X,\mathbb{Q}) = M$, the space introduced in Proposition 5.1, and this is also the maximal rational sub-Hodge structure of $F^1H^3(X,\mathbb{C})$. Hence, the Generalized Hodge Conjecture as amended by Grothendieck is true for such threefolds X.

7. Final remarks

Let us set $\varphi := d_{123} - \mu_1 d_{234} + \mu_2 d_{345} - \mu_3 d_{456}$. Then the detailed computations performed in the last two sections show that

$$F^1H^3(X,\mathbb{C}) \cap H^3(X,\mathbb{Q}) = \mathbb{Q}\,[\varphi] \oplus N^1H^3(X,\mathbb{Q}) .$$

We would like to be able to conclude that $[\varphi] \notin N^1H^3(X,\mathbb{Q})$ *by a direct examination of* $[\varphi]$, or, in a spirit close to Hodge's, by a direct examination of the Poincaré dual of $[\varphi]$. If we set

$$\Gamma_1 := C_4\times C_5\times C_6 \ , \ \Gamma_2 := C_1\times C_5\times C_6 \ , \ \Gamma_3 := C_1\times C_2\times C_6 \ , \ \Gamma_4 := C_1\times C_2\times C_3$$

then $PD([\varphi])$ is represented by the cycle $\Gamma := \Gamma_1 + \mu_1\,\Gamma_2 + \mu_2\,\Gamma_3 + \mu_3\,\Gamma_4$, where the rational numbers μ_i are the coefficients of the minimal polynomial of τ . Now, (10) and (27) together imply

$$\int_{\Gamma_1} \alpha \,=\, \tau^3 \quad , \qquad \int_{\Gamma_2} \alpha \,=\, \tau^2 \quad , \qquad \int_{\Gamma_3} \alpha \,=\, \tau \quad , \qquad \int_{\Gamma_4} \alpha \,=\, 1 \; . \tag{30}$$

Therefore

$$\int_{\Gamma} \alpha \,=\, \tau^3 + \mu_1\tau^2 + \mu_2\tau + \mu_3 \,=\, 0 \; ,$$

which simply means that $[\varphi] \in F^1H^3(X,\mathbb{C})$, as we already know.

The integrals (30) imply, in particular, that no one of the 3-cycles $\Gamma_1, \ldots, \Gamma_4$ can be contained into an algebraic surface $Y \subset X$.

Consider now the cycles $\Gamma_5 := C_2 \times C_3 \times C_4$ and $\Gamma_6 := C_1 \times C_3 \times C_5$. Then, it is easily checked that

$$\int_{\Gamma_5+\Gamma_6} \alpha \,=\, \int_{\Gamma_5} \alpha \,+\, \int_{\Gamma_6} \alpha \,=\, \tau - \tau \,=\, 0 \; .$$

In particular, the computation shows also that neither of the two cycles Γ_5 and Γ_6 can be contained into an effective divisor on X. But

$$PD(\,[\Gamma_5] + [\Gamma_6]\,) \,=\, [d_{156} + d_{246}] \,\in\, N^1H^3(X,\mathbb{Q}) \; ,$$

as we have seen.

Let us add another remark on the above cycles, in the spirit of [6]. Let Γ denote any of the cycles $\Gamma_1, \ldots, \Gamma_4, \Gamma_5, \Gamma_6$. Since Γ is smooth, for any point $P \in \Gamma$ we have $T_P\Gamma \subset T_PX$, where X denotes here the differentiable manifold underlying the projective, smooth variety. Moreover, if $J : T_PX \to T_PX$ denotes the complex structure on the *real* vector space T_PX, then the non vanishing of the integral of the form α on Γ implies that

$$T_P\Gamma \,+\, J(T_P\Gamma) \,=\, T_PX \; .$$

This relation can also be easily checked by an easy, direct computation.

To summarize, thanks to the detailed computations performed in §§ 5 and 6, we know that $[d_{156} + d_{246}]$ is supported by an algebraic surface $Y \subset X$, whereas $[\varphi]$ is not.

On the other hand, our aim would be to be able to directly determine this different nature of these two classes, by examining their respective Poincaré duals, to get a direct proof that $N^1H^3(X,\mathbb{Q}) \subsetneq F^1H^3(X,\mathbb{C}) \cap H^3(X,\mathbb{Q})$.

Throughout the paper we used the representation of the homology classes which is customary for complex tori. The above remarks show that this particular representation is completely inadequate to reach our goal.

All this seems to indicate that to apply the results of [6] to the computation of some concrete case of the Generalized Hodge Conjecture, it is necessary to use bases for the homology spaces, which are induced by some suitable real-analytic semi-algebraic triangulation of X.

Acknowledgements

The author was supported by *Università di Trieste - Finanziamento di Ateneo per progetti di ricerca scientifica - FRA 2013.*

References

[1] P. Deligne, *Théorie de Hodge. III*, Inst. Hautes Études Sci. Publ. Math. (1974), no. 44, 5–77.

[2] A. Grothendieck, *Hodge's general conjecture is false for trivial reasons*, Topology **8** (1969), 299–303.

[3] W. V. D. Hodge, *The Theory and Applications of Harmonic Integrals*, Cambridge University Press, Cambridge, England; Macmillan Company, New York, 1941.

[4] W. V. D. Hodge, *The topological invariants of algebraic varieties*, Proceedings of the International Congress of Mathematicians, Cambridge, Mass., 1950, vol. 1, Amer. Math. Soc., Providence, R. I., 1952, pp. 182–192.

[5] T. Murasaki, *A calculation of rational cohomology classes on some complex tori*, Science Reports of the Faculty of Education, Gunma University **28** (1979), 13–16.

[6] D. Portelli, *A remark on the generalized Hodge conjecture*, Math. Z. **278** (2014), no. 1-2, 1–17.

[7] C. Voisin, *Théorie de Hodge et géométrie algébrique complexe*, Cours Spécialisés [Specialized Courses], vol. 10, Société Mathématique de France, Paris, 2002.

[8] A. Weil, *Sur les théorèmes de de Rham*, Comment. Math. Helv. **26** (1952), 119–145.

Author's address:

Dario Portelli
Università di Trieste
Dipartimento di Matematica e Geoscienze
via A. Valerio 12/1
34127 Trieste, Italy
E-mail: `porteda@units.it`

Received September 26, 2014
Revised November 7, 2014

Rend. Istit. Mat. Univ. Trieste
Volume 46 (2014), 255–270

Nonresonance conditions for radial solutions of nonlinear Neumann elliptic problems on annuli

Andrea Sfecci

Abstract. *An existence result to some nonlinear Neumann elliptic problems defined on balls has been provided recently by the author in [21]. We investigate, in this paper, the possibility of extending such a result to annuli.*

Keywords: Neumann problem, radial solutions, nonresonance, time-map, lower and upper solutions.
MS Classification 2010: 35J25, 34B15.

1. Introduction

In a previous paper [21], in order to obtain an existence result, the author introduced a *liminf-limsup* type of nonresonance condition *below* the first positive eigenvalue for Neumann problems defined on the ball $B_R = \{x \in \mathbb{R}^N\,,\ |x| < R\}$. As an example, the following problem

$$\begin{cases} -\Delta u = g(u) + e(|x|) & \text{in } B_R \\ \dfrac{\partial u}{\partial \nu} = 0 & \text{on } \partial B_R\,, \end{cases}$$

where the functions $g : \mathbb{R} \to \mathbb{R}$ and $e : [0, R] \to \mathbb{R}$ are continuous, has a radial solution if

$$\liminf_{u\to+\infty} \frac{2G(u)}{u^2} < \left(\frac{\pi}{2R}\right)^2 \quad \text{and} \quad \limsup_{u\to-\infty} \frac{g(u)}{u} < \left(\frac{\pi}{2R}\right)^2$$

(here G is a primitive of g), and assuming the existence of a positive d such that

$$\big(g(u) + \bar{e}\big)\,\mathrm{sgn}\, u > 0 \qquad \text{when } |u| \geq d\,,$$

where $\bar{e} = \dfrac{N}{R^N} \displaystyle\int_0^R s^{N-1} e(s)\, ds$.

In this paper, we treat Neumann problems defined on annuli and, in the spirit of the above quoted paper, we will provide in Theorem 2.2 some sufficient condition for the existence of radial solutions. Moreover, we will provide with Theorem 2.6 a different result in presence of *not well-ordered* constant upper and lower solutions.

Let us briefly introduce some notations. We denote by $\mathcal{A}^N(R_1,R_2) \subset \mathbb{R}^N$, with $R_2 > R_1 > 0$, the open annulus of internal radius R_1 and external radius R_2:

$$\mathcal{A}^N(R_1,R_2) = B_{R_2} \setminus \overline{B_{R_1}}\,,$$

where $B_r \subset \mathbb{R}^N$ is the open ball of radius r centered at the origin. As usual, we denote the boundary of $\mathcal{A}^N(R_1,R_2)$ with $\partial\mathcal{A}^N(R_1,R_2)$, and the euclidean norm with $|\cdot|$. The problem we are going to study is of the type

$$\begin{cases} -\Delta u = g(|x|,u) + e(|x|) & \text{in } \mathcal{A}^N(R_1,R_2) \\ \dfrac{\partial u}{\partial \nu} = 0 & \text{on } \partial\mathcal{A}^N(R_1,R_2)\,, \end{cases}$$

where $g : [R_1,R_2] \times \mathbb{R} \to \mathbb{R}$ and $e : [R_1,R_2] \to \mathbb{R}$ are continuous functions.

The spirit of this paper follows the idea presented by Fonda, Gossez and Zanolin in [9]. In that paper, the authors deal with a Dirichlet problem defined in a smooth domain $\Omega \subset \mathbb{R}^N$ contained in a ball B_ρ of a certain radius ρ:

$$\begin{cases} -\Delta u = g(u) + h(x) & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega\,. \end{cases}$$

They replace the classical *limsup* nonresonance condition with respect to the first eigenvalue λ_1 provided by Hammerstein in [15],

$$\limsup_{|u|\to\infty} \frac{2G(u)}{u^2} < \lambda_1\,,$$

with a *double liminf* condition like the following one

$$\liminf_{u\to-\infty} \frac{2G(u)}{u^2} < \frac{\pi^2}{4\rho^2}, \quad \text{and} \quad \liminf_{u\to+\infty} \frac{2G(u)}{u^2} < \frac{\pi^2}{4\rho^2} \tag{1}$$

(here, again, G is a primitive of g). Notice that, one has $\pi^2/4\rho^2 < \lambda_1$, except to the case $\Omega = (-2\rho, 2\rho) \subset \mathbb{R}$ where the equality holds. Condition (1) has been first introduced in the frame of the one-dimensional Dirichlet problem in $(-2\rho, 2\rho)$ in [7].

In the case of a Neumann problem, a condition of *liminf* type was studied by Gossez and Omari in [12, 13]. In Neumann problems, a nonresonance condition with respect to the *zero* eigenvalue must be introduced, so that the *liminf* condition will be related to the *first positive* eigenvalue (see also [3, 16] for related problems). Such a situation occurs also when dealing with periodic problems (see for example [8]).

The paper is organized as follows: in Section 2 we will state all the results, the proofs of them are postponed to Section 3.

2. Main results

In this paper we are concerned with the following class of problems defined on an annulus $\mathcal{A}^N(R_1, R_2)$ with $R_2 > R_1 > 0$:

$$\begin{cases} -\Delta u = g(|x|, u) + e(|x|) & \text{in } \mathcal{A}^N(R_1, R_2) \\ \dfrac{\partial u}{\partial \nu} = 0 & \text{on } \partial\mathcal{A}^N(R_1, R_2)\,, \end{cases} \tag{2}$$

where $g : [R_1, R_2] \times \mathbb{R} \to \mathbb{R}$ and $e : [R_1, R_2] \to \mathbb{R}$ are continuous functions. In particular, consider a radial solution $u(x) = v(|x|)$ to (2). Setting $r = |x|$, and denoting with $'$ the derivative with respect to r, we have the equivalent system

$$\begin{cases} -v'' - \dfrac{N-1}{r}\, v' = g(r, v) + e(r) & r \in [R_1, R_2] \\ v'(R_1) = 0 = v'(R_2)\,. \end{cases} \tag{3}$$

Notice that the differential equation in (3) does not present a singularity, being $R_1 > 0$. The case $R_1 = 0$ has been treated by the author in [21]. It will be useful to consider the mean value of the function e

$$\bar{e} = \frac{N}{R_2^N - R_1^N} \int_{R_1}^{R_2} s^{N-1} e(s)\, ds$$

and $\tilde{e}(t) = e(t) - \bar{e}$, so that $\int_{R_1}^{R_2} s^{N-1} \tilde{e}(s)\, ds = 0$.

In the proof of the theorem, we will use the so-called *time-map* function. Let us spend a few words about it. Consider the scalar second order differential equation $x'' + \psi(x) = 0$. It is possible to write the associated system in the plane

$$x' = y\,, \qquad -y' = \psi(x)\,. \tag{4}$$

Suppose $\psi(x)x > 0$ for every $x \neq 0$, and consider the primitive $\Psi(x) = \int_0^x \psi(\xi)\, d\xi$. The function

$$\tau_\psi(x) = \operatorname{sgn}(x)\sqrt{2} \int_0^x \frac{d\xi}{\sqrt{\Psi(x) - \Psi(\xi)}}\,,$$

is defined as the time-map associated to the planar system (4), and gives an estimate of the time between two subsequent zeroes t_1 and t_2 of the function

$x = x(t)$. In particular, if the function x reaches its maximum $x(t_0) = x_M$ at $t_0 \in (t_1, t_2)$, then $t_2 - t_1 = \tau_\psi(x_M)$ and $t_2 - t_0 = t_0 - t_1 = \tau_\psi(x_M)/2$. See, e.g., [6, 11, 18, 19] for details and their applications to periodic scalar problems.

In view of this, let us define the *half-valued* time-map

$$\mathcal{T}_\psi(x) = \operatorname{sgn}(x) \frac{1}{\sqrt{2}} \int_0^x \frac{d\xi}{\sqrt{\Psi(x) - \Psi(\xi)}} \,,$$

and the following limits

$$\mathcal{T}_\psi^\pm = \limsup_{x \to \pm\infty} \mathcal{T}_\psi(x) \,, \qquad \mathcal{T}_\pm^\psi = \liminf_{x \to \pm\infty} \mathcal{T}_\psi(x) \,.$$

In [11], Fonda and Zanolin provided some estimates on these values, some of which we collect in the following proposition.

Proposition 2.1 ([11]). *Assume that ψ is a continuous function, with primitive Ψ, and ℓ_+, ℓ_- are positive constants. If ψ satisfies at $+\infty$ or $-\infty$ some of the following limits on the left, then the correspondent estimate on the right holds.*

$$\liminf_{x \to \pm\infty} \frac{2\Psi(x)}{x^2} \le \ell_\pm \quad \Rightarrow \quad \mathcal{T}_\psi^\pm \ge \frac{\pi}{2\sqrt{\ell_\pm}},$$

$$\limsup_{x \to \pm\infty} \frac{\psi(x)}{x} \le \ell_\pm \quad \Rightarrow \quad \mathcal{T}_\pm^\psi \ge \frac{\pi}{2\sqrt{\ell_\pm}},$$

$$\exists \lim_{x \to \pm\infty} \frac{2\Psi(x)}{x^2} \le \ell_\pm \quad \Rightarrow \quad \mathcal{T}_\pm^\psi \ge \frac{\pi}{2\sqrt{\ell_\pm}}.$$

We can now state our main result.

Theorem 2.2. *Assume the existence of a continuous function $\phi : \mathbb{R} \to \mathbb{R}$, and of a constant $d > 0$ such that*

$$-\bar{e} < g(r, v) \le \phi(v) \qquad \text{for every } r \in [R_1, R_2] \text{ and every } v \ge d \,, \tag{5}$$

$$\phi(v) \le g(r, v) < -\bar{e} \qquad \text{for every } r \in [R_1, R_2] \text{ and every } v \le -d \,. \tag{6}$$

Moreover assume the existence of a constant $\eta > 0$ such that

$$\phi(v)v \ge \eta\, v^2 \qquad \text{for every } |v| \ge d \,. \tag{7}$$

Suppose that the function $\mathcal{T}_\phi$ is well-defined for $|v| > d$ and its limits satisfy either

$$\mathcal{T}_\phi^+ > R_2 - R_1 \quad \text{and} \quad \mathcal{T}_-^\phi > R_2 - R_1 \,, \tag{8}$$

or

$$\mathcal{T}_+^\phi > R_2 - R_1 \quad \text{and} \quad \mathcal{T}_\phi^- > R_2 - R_1 \,. \tag{9}$$

Then (2) has at least one radial solution.

Remark 2.3. Assuming (5) and (6) we implicitly have that the function $\mathcal{T}_\phi = \mathcal{T}_\phi(v)$ is well-defined for v large enough. The assumptions of the theorem require only that the value d is chosen large enough to guarantee that the domain of $\mathcal{T}_\phi$ contains the set $(-\infty, -d) \cup (d, +\infty)$.

We will prove this theorem in Section 3.1. We will give now, as an example of application, some possible corollaries to Theorem 2.2 using the estimates in Proposition 2.1. In order to simplify the statement, in the setting of problem (2), we assume g not depending by x.

Corollary 2.4. *Let be $g(|x|, u) = g(u)$ with primitive G. Assume*

$$\liminf_{u\to+\infty} \frac{2G(u)}{u^2} < \left(\frac{\pi}{2(R_2 - R_1)}\right)^2 \quad \text{and} \quad \limsup_{u\to-\infty} \frac{g(u)}{u} < \left(\frac{\pi}{2(R_2 - R_1)}\right)^2$$

and that there exists $d > 0$ such that

$$(g(u) + \bar{e}) \operatorname{sgn} u > 0 \quad \text{when } |u| > d\,.$$

Then, problem (2) has at least one radial solution.

Corollary 2.5. *Let be $g(|x|, u) = g(u)$ with primitive G. Assume*

$$\liminf_{u\to+\infty} \frac{2G(u)}{u^2} < \left(\frac{\pi}{2(R_2 - R_1)}\right)^2 \quad \text{and} \quad \exists \lim_{u\to-\infty} \frac{2G(u)}{u^2} < \left(\frac{\pi}{2(R_2 - R_1)}\right)^2$$

and that there exists $d > 0$ such that

$$(g(u) + \bar{e}) \operatorname{sgn} u > 0 \quad \text{when } |u| > d\,.$$

Then, problem (2) has at least one radial solution.

The proof is obtained by defining, for $\eta > 0$ sufficiently small, the function

$$\phi(v) = \begin{cases} \max\{g(v)\,,\, \eta\, v\} & \text{if } v \geq d \\ \min\{g(v)\,,\, \eta\, v\} & \text{if } v \leq -d\,, \end{cases} \tag{10}$$

enlarging d if necessary, and extending its domain to the whole $\mathbb{R}$.

Another existence result can be obtained assuming the existence of constant lower and upper solutions which are not *well-ordered*, as the following theorem states. For further results on non-well-ordered lower and upper solutions, see, e.g., [1, 5, 10, 14, 17].

Theorem 2.6. *Assume the existence of a continuous function $\phi : \mathbb{R} \to \mathbb{R}$ and of some positive constants d, χ, η such that*

$$-\chi < g(r, v) \leq \phi(v) \qquad \text{for every } r \in [R_1, R_2] \text{ and every } v \geq d\,,$$

$$\phi(v) \le g(r,v) < \chi \qquad \textit{for every } r \in [R_1, R_2] \textit{ and every } v \le -d\,,$$
$$\phi(v)v \ge \eta v^2 \qquad \textit{for every } |v| \ge d\,.$$

Moreover, assume that there exist some constants $\beta < \alpha$ such that

$$g(r,\beta) + e(r) < 0 < g(r,\alpha) + e(r)\,, \qquad \textit{for every } r \in [R_1, R_2]\,. \tag{11}$$

Suppose that the function $\mathcal{T}_\phi$ is well-defined for $|v| > d$ and its limits satisfy either

$$\mathcal{T}_\phi^+ > R_2 - R_1 \quad \textit{and} \quad \mathcal{T}_-^\phi > R_2 - R_1\,,$$

or

$$\mathcal{T}_+^\phi > R_2 - R_1 \quad \textit{and} \quad \mathcal{T}_\phi^- > R_2 - R_1\,.$$

Then (2) *has at least one radial solution.*

Such a statement has been inspired by a result obtained by Gossez and Omari in [12], and the following results follow as a direct consequence of the previous theorem. We will refer also to [13] for comparison.

Theorem 2.7. *Assume g to be a continuous function, with primitive G, such that*

$$\liminf_{u\to+\infty} \frac{2G(u)}{u^2} < \left(\frac{\pi}{2(R_2-R_1)}\right)^2 \quad \textit{and} \quad \limsup_{u\to-\infty} \frac{g(u)}{u} < \left(\frac{\pi}{2(R_2-R_1)}\right)^2.$$

Then,

$$\begin{cases} -\Delta u = g(u) + e(|x|) & \textit{in } \mathcal{A}^N(R_1,R_2) \\ \dfrac{\partial u}{\partial \nu} = 0 & \textit{on } \partial\mathcal{A}^N(R_1,R_2)\,, \end{cases}$$

has a solution for every continuous function e if and only if $g(\mathbb{R}) = \mathbb{R}$.

Theorem 2.8. *Assume g to be a continuous function, with primitive G, such that*

$$\liminf_{u\to+\infty} \frac{2G(u)}{u^2} < \left(\frac{\pi}{2(R_2-R_1)}\right)^2 \quad \textit{and} \quad \exists \lim_{u\to-\infty} \frac{2G(u)}{u^2} < \left(\frac{\pi}{2(R_2-R_1)}\right)^2.$$

Then,

$$\begin{cases} -\Delta u = g(u) + e(|x|) & \textit{in } \mathcal{A}^N(R_1,R_2) \\ \dfrac{\partial u}{\partial \nu} = 0 & \textit{on } \partial\mathcal{A}^N(R_1,R_2)\,, \end{cases}$$

has a solution for every continuous function e if and only if $g(\mathbb{R}) = \mathbb{R}$.

For comparison, let us quote here a possible application of [13, Theorem 1.1]. The quoted theorem is a more general application to asymmetric nonlinearities.

THEOREM 2.9 ([13]). *Let λ_2 be the second eigenvalue of $-\Delta$ on $\mathcal{A}^N(R_1, R_2)$ with Neumann boundary condition. Assume g to be a continuous function, with primitive G, such that*

$$\limsup_{u\to+\infty} \frac{g(u)}{u} \le \lambda_2 \quad \text{and} \quad \limsup_{u\to-\infty} \frac{g(u)}{u} \le \lambda_2\,,$$

with moreover $\liminf\limits_{u\to+\infty} \dfrac{2G(u)}{u^2} < \lambda_2$. *Then,*

$$\begin{cases} -\Delta u = g(u) + e(x) & \text{in } \mathcal{A}^N(R_1, R_2) \\ \dfrac{\partial u}{\partial \nu} = 0 & \text{on } \partial\mathcal{A}^N(R_1, R_2)\,, \end{cases}$$

has a solution for every continuous function e if and only if $g(\mathbb{R}) = \mathbb{R}$.

Notice that Theorem 2.7 does not require any *limsup* type of condition at $+\infty$, but it requires that all the limits are *below* the second eigenvalue.

REMARK 2.10. It is known that the Fučik spectrum for the radial elliptic Neumann problem on an annulus presents two monotone curves departing from the point (λ_2, λ_2), where λ_2 is the first positive eigenvalue (cf. [2, 20]). Such curves have two different asymptotes, call $a > 0$ the smaller one. A natural question arises about the order of a and of the constants $k = \frac{\pi^2}{4(R_2-R_1)^2}$ involved in the previous theorems. The value a is strictly related to the zeroes of Bessel functions of index ν and $\nu + 1 = N/2$ and also, in particular, to the choice of R_1 and R_2. It is possible to find suitable values for them thus obtaining both the cases $a < k$ and $a > k$, so that the *liminf* condition in Corollary 2.4 and Theorem 2.7 is sometimes not necessary, but it is hard to verify if this situation occurs when R_1 and R_2 are arbitrarily fixed.

REMARK 2.11. Similar results can be obtained by assuming the existence of non-constant lower and upper solutions which are not well-ordered, following the main ideas of the paper by Alif and Omari [1]. We do not enter in such details for briefness.

REMARK 2.12. Several other theorems can be formulated using the estimates in Proposition 2.1 and the other ones contained in [11]. As a trivial example, the asymptotic behaviour of the nonlinearities at $+\infty$ and $-\infty$ can be switched in all the previous theorems. For briefness we do not enter in such details.

3. Proofs

3.1. Proof of Theorem 2.2

We will prove the theorem under assumption (8). The proof of the other case is specular.

Define the function $T : (-\infty, -d] \cup [d, +\infty) \to \mathbb{R}$ as

$$T(v) = \frac{1}{\sqrt{2}} \int_d^v \frac{d\xi}{\sqrt{\Phi(v) - \Phi(\xi) + \|e\|_\infty (v - \xi)}}\,, \qquad \text{for } v \geq d\,,$$

$$T(v) = \frac{1}{\sqrt{2}} \int_v^{-d} \frac{d\xi}{\sqrt{\Phi(v) - \Phi(\xi) - \|e\|_\infty (v - \xi)}}\,, \qquad \text{for } v \leq -d\,.$$

The following proposition was proved in [21, Lemma 3.1].

PROPOSITION 3.1. *For every $\epsilon > 0$ there exists $v_\epsilon > d$ such that the following inequalities hold*

$$T(v) \leq \mathcal{T}_\phi(v) \leq (1 + \epsilon)\, T(v) + \epsilon\,,$$

for every v with $|v| > v_\epsilon$.

By (8), it is possible to find a sufficiently small $\epsilon > 0$ such that there exist an increasing sequence of positive real values $(\omega_n)_n$, with $\lim_n \omega_n = +\infty$, and $\bar{\omega} > 0$ with the following property:

$$\mathcal{T}_\phi(\omega_n) > (R_2 - R_1)(1 + \epsilon) + \epsilon \qquad \text{for every } n \in \mathbb{N}\,,$$

$$\mathcal{T}_\phi(v) > (R_2 - R_1)(1 + \epsilon) + \epsilon \qquad \text{for every } v < -\bar{\omega}\,.$$

We can assume $\bar{\omega}$ and ω_0 to be greater than $d + 1$ and v_ϵ, where v_ϵ is given by Proposition 3.1, thus permitting to have the following estimates

$$T(\omega_n) \geq \frac{\mathcal{T}_\phi(\omega_n) - \epsilon}{1 + \epsilon} > R_2 - R_1 \qquad \text{for every } n \in \mathbb{N}\,, \tag{12}$$

$$T(v) \geq \frac{\mathcal{T}_\phi(v) - \epsilon}{1 + \epsilon} > R_2 - R_1 \qquad \text{for every } v < -\bar{\omega}\,. \tag{13}$$

We introduce the following family of problems, for $\lambda \in [0, 1]$,

$$\begin{cases} -v'' - \dfrac{N-1}{r}\, v' = \lambda\big(g(r, v) + e(r)\big) + (1 - \lambda)\eta v\,, \quad r \in [R_1, R_2]\,, \\ v'(R_1) = 0 = v'(R_2)\,, \end{cases} \tag{14}$$

where η was introduced in (7). We define the following sets

$$C^k_N = \{v \in C^k([R_1, R_2]) \,:\, v'(R_1) = 0 = v'(R_2)\}\,, \quad k = 1, 2\,.$$

It is not restrictive to assume that the constant η introduced in (7) is smaller than the first positive eigenvalue, so to have the existence of a unique solution of the problem

$$\begin{cases} -v'' - \dfrac{N-1}{r}\, v' = \eta v + f(r)\,, \quad r \in [R_1, R_2]\,, \\ v'(R_1) = 0 = v'(R_2)\,, \end{cases}$$

for every continuous function f. Call $T : C^0 \to C^1_N$ the operator that sends the continuous function f to the unique solution v. Then (14) is equivalent to the fixed point problem

$$v = \mathcal{G}_\lambda(v) := \lambda T\big(-\eta v + g(r,v) + e(r)\big)\,.$$

where $\mathcal{G}_\lambda : C^1_N \to C^1_N$ is a completely continuous operator. Moreover, any fixed point of $\mathcal{G}_\lambda$ is a function belonging to C^2_N and $d_{LS}(I - \mathcal{G}_0, \Omega, 0) = 1$ for every open bounded set $\Omega \subset C^1_N$ such that $0 \in \Omega$. Hence, by Leray-Schauder degree theory, it will be sufficient to find an open bounded set $\Omega \subset C^1_N$, containing 0, such that there are no solutions of (14) on $\partial\Omega$, for every $\lambda \in [0,1]$, in order to prove the existence of a solution of (3).

We are going to prove the existence of such a set, looking for some positive constants A, B, M defining Ω as follows:

$$\Omega = \{v \in C^1_N : -A < v(r) < B \text{ and } |v'(r)| < M\,,\ \forall r \in [R_1,R_2]\}\,. \tag{15}$$

First of all, we show now that all the solutions of (14) cannot *remain large* because of assumptions (5) and (6). We will prove the following claim.

Claim. Every solution v of (14) satisfies $|v(r)| < d$ for some $r \in [R_1,R_2]$.

Consider a solution v of (14) such that $v(r) > d$ for every $r \in [R_1,R_2]$. It satisfies also the following differential equation for every $r \in [R_1,R_2]$:

$$\frac{d}{dr}\big(r^{N-1}v'(r)\big) = -r^{N-1}\left[\lambda\big(g(r,v(r)) + e(r)\big) + (1-\lambda)\eta v(r)\right]\,. \tag{16}$$

Integrating it in the interval $[R_1,R_2]$, we get

$$0 = -\int_{R_1}^{R_2} r^{N-1}\left[\lambda\big(g(r,v(r)) + \bar{e}\big) + (1-\lambda)\eta v(r)\right]\,dr\,.$$

Notice that, by (5), the integral must be negative, providing a contradiction. A similar computation proves also the impossibility of having a solution v of (14) such that $v(r) < -d$ for every $r \in [R_1,R_2]$. We have so proved the claim.

The proof of Theorem 2.2 consists of three steps: in each one we provide one of the needed constants A, B, M which appear in (15).

► *Step 1 (Find the constant B).* The positive constant B can be chosen in the set of the values of the previously introduced sequence $(\omega_n)_n$, taking n sufficiently large. In fact, suppose by contradiction that there exist a sequence $(\lambda_n)_n$, with $\lambda_n \in [0,1]$ for every n, a subsequence of $(\omega_n)_n$, still denoted $(\omega_n)_n$, and a sequence of solutions v_n to (14), with $\lambda = \lambda_n$, such that $\max_{[R_1,R_2]} v_n = \omega_n$. The maximum is reached at the instant

$$r^n_M = \max\{r \in [R_1,R_2] : v_n(r) = \omega_n\}\,.$$

The Claim permits us to define also

$$r_d^n = \max\{r \in [R_1, R_2] : v_n(r) = d\}\,.$$

We consider two different situations, up to subsequences.

$\diamond$ *Case 1:* $r_M^n < r_d^n$.
In this situation, the solution reaches its maximum at r_M^n and then *becomes small* reaching the value d at

$$\tilde r_n = \min\{r \in [r_M^n, R_2] : v_n(r) = d\}\,.$$

For every $r \in [r_M^n, \tilde r_n]$ such that $v_n'(r) < 0$, it is possible to find a value $s(r) \in [r_M^n, r)$ such that $v_n'(s(r)) = 0$ and $v_n'(r) < 0$ for every $s \in (s(r), r]$. Consider the differential equation in (14) with $v = v_n$ and $\lambda = \lambda_n$. Using (5) and (7), we have

$$-v_n''(s) \le \phi(v_n(s)) + \|e\|_\infty \qquad \text{for every } s \in [s(r), r]\,,$$

and multiplying by $v_n'(s) \le 0$ and integrating in the interval $[s(r), r]$, we obtain

$$-\frac{1}{2} v_n'(r)^2 \ge \Phi(v_n(r)) - \Phi(v_n(s(r))) + \|e\|_\infty \big(v_n(r) - v_n(s(r))\big)\,.$$

Using the monotonicity of Φ in the interval $[d, +\infty)$, we get

$$1 \ge \frac{1}{\sqrt{2}} \frac{-v_n'(r)}{\sqrt{\Phi(\omega_n) - \Phi(v_n(r)) + \|e\|_\infty(\omega_n - v_n(r))}}\,,$$

for every $r \in [r_M^n, \tilde r_n]$ such that $v_n'(r) < 0$. The previous inequality holds also when $v_n' > 0$, so we can obtain the following contradiction using (12):

$$\begin{aligned}\tilde r_n - r_M^n &\ge \frac{1}{\sqrt{2}} \int_{r_M^n}^{\tilde r_n} \frac{-v_n'(r)}{\sqrt{\Phi(\omega_n) - \Phi(v_n(r)) + \|e\|_\infty(\omega_n - v_n(r))}}\, dr \\ &= T(\omega_n) > R_2 - R_1\,.\end{aligned}$$

$\diamond$ *Case 2:* $r_M^n > r_d^n$.
We want to show that, in this situation, the solutions v_n must reach a negative minimum $m_n = \min_{[R_1,R_2]} v_n$ which is large in absolute value, in particular we will prove that

$$\lim_n m_n = -\infty\,. \tag{17}$$

We consider the last point of minimum

$$r_m^n = \max\{r \in [R_1, R_2] : v_n(r) = m_n\} < r_M^n.$$

Arguing by contradiction, we suppose that there exists a constant $C > 0$ such that, up to a subsequence, $v_n(r) \geq -C$ for every $r \in [R_1, R_2]$ and every $n \in \mathbb{N}$. Defining

$$\tilde{g}_n(r) = -r^{N-1}\left[\lambda_n\big(g(r, v_n(r)) + \bar{e}\big) + (1-\lambda_n)\eta v_n(r)\right], \tag{18}$$

we verify, using (16), that

$$\int_{R_1}^{R_2} \tilde{g}_n(r)dr = 0\,.$$

Hence, being $\tilde{g}_n$ negative when $v_n > d$,

$$\begin{aligned}\int_{R_1}^{R_2} |\tilde{g}_n(r)|dr &= \int\limits_{v_n > d} -\tilde{g}_n(r)dr + \int\limits_{-C \leq v_n \leq d} |\tilde{g}_n(r)|dr \\ &\leq 2 \int\limits_{-C \leq v_n \leq d} |\tilde{g}_n(r)|dr\,,\end{aligned}$$

which is bounded. Form (16) and the previous computation, we obtain

$$\left\| \frac{d}{dr}(r^{N-1}v_n') \right\|_{L^1} \leq D\,, \tag{19}$$

for a suitable constant D, independent of n. Thus, for every $r \in (r_m^n, R_2]$,

$$r^{N-1}v_n'(r) = (r_m^n)^{N-1}v_n'(r_m^n) + \int_{r_m^n}^{r} \big(s^{N-1}v_n'(s)\big)'ds \leq D\,.$$

Hence, $v_n'(r) < D/R_1^{N-1}$ for every $r > r_m^n$, for every n. The following computation gives us a contradiction with the assumption $\omega_n \to +\infty$ giving us the proof of the limit in (17):

$$\omega_n = v_n(r_M^n) = v_n(r_m^n) + \int_{r_m^n}^{r_M^n} v_n'(s)\,ds \leq d + \frac{D}{R_1^{N-1}}(R_2 - R_1)\,.$$

Being (17) valid, we can assume $m_n < -d$ for every n. Consider

$$\hat{r}_n = \min\{r \in (r_m^n, r_d^n) : v_n(r) = -d\}\,.$$

Arguing as in Case 1, we can find, for every $r \in [r_m^n, \hat{r}_n]$ such that $v_n'(r) > 0$, a value $s(r) \in [r_m^n, r)$ such that $v_n'(s(r)) = 0$ and $v_n'(s) > 0$ for every $s \in (s(r), r]$. Considering the differential equation in (14) with $v = v_n$ and $\lambda = \lambda_n$, we can write, using (6) and (7),

$$-v_n''(s) \geq \phi(v_n(s)) - \|e\|_\infty \qquad \text{for every } s \in [s(r), r]\,.$$

Multiplying it by $v_n'(s) \geq 0$ and integrating in the interval $[s(r), r]$, using the monotonicity of Φ in $(-\infty, -d]$, we obtain, arguing as above,

$$1 \geq \frac{1}{\sqrt{2}} \frac{v_n'(r)}{\sqrt{\Phi(m_n) - \Phi(v_n(r)) - \|e\|_\infty (m_n - v_n(r))}} \,.$$

for every $r \in [r_m^n, \hat{r}_n]$, thus giving us the following contradiction when n is large enough, using (13):

$$\begin{aligned} \hat{r}_n - r_m^n &\geq \frac{1}{\sqrt{2}} \int_{r_m^n}^{\hat{r}_n} \frac{v_n'(r)}{\sqrt{\Phi(m_n) - \Phi(v_n(r)) - \|e\|_\infty (m_n - v_n(r))}}\, dr \\ &= T(m_n) > R_2 - R_1 \,. \end{aligned}$$

We have just proved that there cannot exist solutions to (14) such that $\max_{[R_1,R_2]} v_n = \omega_n$ if n is large enough. So, we can choose B among such values.

► *Step 2 (Find the constant A).* When B is fixed, it is possible to prove that there cannot exist solutions to (14), for a certain λ, having $\max_{[R_1,R_2]} v < B$ with a large (in absolute value) negative minimum.

Suppose by contradiction that, for every $m \in \mathbb{N}$, there exists a solution v_m to (14), for a certain λ, with $\max_{[R_1,R_2]} v_m < B$, such that $\min_{[R_1,R_2]} v_m < -m$. By the Claim, if $-m < -d$ then $\max_{[R_1,R_2]} v_m > -d$.

Arguing as above, we can define the function $\tilde{g}_m$ as in (18) and, being $\tilde{g}_m$ positive when $v_m < -d$, with a similar procedure, we can find a constant D' (independent of m) such that $v_m' < D'/R_1^{N-1}$ so to obtain

$$\begin{aligned} -d < \max_{[R_1,R_2]} v_m &\leq \min_{[R_1,R_2]} v_m + \frac{D'}{R_1^{N-1}}(R_2 - R_1) \\ &< -m + \frac{D'}{R_1^{N-1}}(R_2 - R_1) \,, \end{aligned}$$

which gives us a contradiction when m is large enough. Hence, we can find a positive constant A, such that every solution v to (14) satisfying $\max_{[R_1,R_2]} v < B$ must also satisfy $\min_{[R_1,R_2]} v > -A$.

► *Step 3 (Find the constant M).* Consider a solution v of (14) with $-A < v < B$, then by (16) it is easy to see that

$$r^{N-1}|v'(r)| \leq \int_{R_1}^{R_2} \left| \frac{d}{ds}(s^{N-1} v'(s)) \right| ds \leq K r^N \,,$$

for a suitable positive constant K. So, we get $|v'| \leq K R_2$ and setting, for example, $M = K R_2 + 1$ also the third step of the proof is completed.

We have just found the three constants A, B, M describing a set Ω suitable to apply the Leray-Schauder degree theory, completing the proof of Theorem 2.2.

3.2. Proof of Theorem 2.6

The proof is rather similar to the one of Theorem 2.2. Proposition 3.1 remains valid also under the assumptions of Theorem 2.6.

It is not restrictive to assume $d > \max\{-\beta, \alpha\}$. Let us consider first the case $\beta < 0 < \alpha$. As above, we can find a sequence of values $(\omega_n)_n$ and a constant $\bar{\omega}$ satisfying (12) and (13). We can introduce problem (14) and the operator $\mathcal{G}_\lambda$, but we are now going to look for a different kind of set Ω. In particular it will be of the form

$$\Omega = \{v \in C^1_N : -A < v(r) < B\,,\ |v'(r)| < M \text{ for every } r \in [R_1, R_2] \\ \text{and } \exists r_0 \in [R_1, R_2]\,:\, \beta < v(r_0) < \alpha\}\,.$$

The impossibility of having solutions v to (14) satisfying $\max_{[R_1,R_2]} v = \beta$ or $\min_{[R_1,R_2]} v = \alpha$ is given by assumption (11). The proof of this theorem follows the main procedure of the one of Theorem 2.2 and consists of three steps, too.

► *Step 1 (Find the constant B).* It is possible to find the constant B in the set of the values of the sequence $(\omega_n)_n$. If n is chosen large enough, one can prove that any solution v to (14), with $\beta < v(r_0) < \alpha$ for a certain $r_0 \in [R_1, R_2]$, must satisfy $\max_{[R_1,R_2]} v \neq \omega_n$.

► *Step 2 (Find the constant A).* It is possible to find the constant A choosing it sufficiently large in order to obtain that any solution v to (14), with $\beta < v(r_0) < \alpha$ for a certain $r_0 \in [R_1, R_2]$ and satisfying $\max_{[R_1,R_2]} v < B$, must also satisfy $\min_{[R_1,R_2]} v > -A$.

► *Step 3 (Find the constant M).* It is possible to find a constant M, sufficiently large, in order to guarantee that any solution v to (14), with $-A < v(r) < B$ for every $r \in [R_1, R_2]$, must satisfy $\max_{[R_1,R_2]} |v'| < M$.

We emphasize that, in order to adapt the proof of Theorem 2.2 to the assumptions of Theorem 2.6, we have to rewrite the part involving the estimate in (19). In this part, in fact, we have used the property $\operatorname{sgn}(v)(g(r,v) + \bar{e}) > 0$ which does not hold necessarily under the assumptions of Theorem 2.6. So, we are going to rewrite this part.

We can assume $\chi > \bar{e}$ and rename the function $\tilde{g}_n$, appearing in (18), as

$$\tilde{g}_n(r) = -r^{N-1}\left[\lambda_n\big(g(r, v_n(r)) + \chi\big) + (1 - \lambda_n)\eta v_n(r)\right],$$

so that

$$\frac{d}{dr}(r^{N-1}v_n'(r))' = \tilde{g}_n(r) + \lambda_n r^{N-1}(\chi - e(r))\,.$$

Integrating the equation in the interval $[R_1, R_2]$, we obtain

$$\int_{R_1}^{R_2} \tilde{g}_n(s)ds \geq -\frac{\chi - \bar{e}}{N}(R_2^N - R_1^N)\,.$$

Let us define $H = (\chi - \bar{e})(R_2^N - R_1^N)/N$. Being $\tilde{g}_n < 0$ when $v_n > d$, we have

$$\begin{aligned}\int_{R_1}^{R_2} |\tilde{g}_n(s)|ds &= \int\limits_{v_n > d} -\tilde{g}_n(s)ds + \int\limits_{-C \le v_n \le d} |\tilde{g}_n(s)|ds \\ &\le H + \int\limits_{-C \le v_n \le d} \tilde{g}_n(s)ds + \int\limits_{-C \le v_n \le d} |\tilde{g}_n(s)|ds \\ &\le H + \quad 2 \int\limits_{-C \le v_n \le d} |\tilde{g}_n(s)|ds\,,\end{aligned}$$

which is bounded, independently of n. Then,

$$\begin{aligned}\int_{R_1}^{R_2} \left| \frac{d}{ds}(s^{N-1}v_n'(s)) \right| ds \le \int_{R_1}^{R_2} |\tilde{g}_n(s)|ds + \frac{\chi}{N}(R_2^N - R_1^N) \\ + \int_{R_1}^{R_2} s^{N-1}|e(s)|\, ds \le D'\,,\end{aligned}$$

for a suitable constant D', independent of n. We thus obtain

$$\left\| \frac{d}{dr}(r^{N-1}v_n') \right\|_{L^1} \le D'\,,$$

for every n. Then the proof works as the one of Theorem 2.2.

The proof of Theorem 2.6, is now completed only in the case $\beta < 0 < \alpha$. Suppose now that this condition is not fulfilled. Choosing $\xi \in (\beta, \alpha)$ and defining $h(r, v) = g(r, v+\xi)$, we can verify that the assumptions of Theorem 2.6 are also satisfied for $\beta_1 = \beta - \xi < 0 < \alpha - \xi = \alpha_1$ and $\phi_1 = \phi(\,\cdot\, + \xi)$, even slightly modifying the other values. Thus, we can find a solution z of

$$\begin{cases} -\Delta z = h(|x|, z) + e(|x|) & \text{in } \mathcal{A}^N(R_1, R_2) \\ \dfrac{\partial z}{\partial \nu} = 0 & \text{on } \partial\mathcal{A}^N(R_1, R_2)\,, \end{cases}$$

so that the function $u = z + \xi$ is a solution to (2), thus completing the proof.

3.3. Proof of Theorems 2.7 and 2.8

It is easy to prove that the requirement $g(\mathbb{R}) = \mathbb{R}$ is a necessary condition. So, we will prove only that it is also sufficient. By hypothesis, for every continuous function e, it is possible to find $\alpha, \beta \in \mathbb{R}$ such that $g(\alpha) \ge \|e\|_\infty$ and $g(\beta) \le -\|e\|_\infty$, thus having the property of being respectively a lower and an upper constant solution to (2). The case $\alpha \le \beta$ follows from classical results (see, e.g., [4]), so let us assume $\beta < \alpha$. Suppose that $\liminf_{u\to+\infty} g(u) = -\infty$, then it is possible to find a constant upper solution $\beta_2 > \alpha$, thus concluding. Similarly, if $\limsup_{u\to-\infty} g(u) = +\infty$, we can find $\alpha_2 < \beta$ being a constant lower solution. The interesting case is the remaining one: there exists a positive constant $\chi > 0$ such that $g(v)\operatorname{sgn}(v) \geq -\chi$. Defining ϕ as in (10), the assumption of Theorem 2.6 are fulfilled using the estimates in Proposition 2.1. Applying Theorem 2.6 we complete the proof.

Acknowledgements

I thank Pierpaolo Omari for the useful discussions.

References

[1] M. Alif and P. Omari, *On a p-Laplace Neumann problem with asymptotically asymmetric perturbations*, Nonlinear Anal. **51** (2002), 369–389.
[2] M. Arias and J. Campos, *Radial Fučik spectrum of the Laplace operator*, J. Math. Anal. Appl. **190** (1995), 654–666.
[3] C. Bereanu, P. Jebelean and J. Mawhin, *Radial solutions for Neumann problems involving mean curvature operators in Euclidean and Minkowski spaces*, Math. Nachr. **3** (2010), 379–391.
[4] C. De Coster and P. Habets, *Two-point boundary value problems. Lower and upper solutions*, Elsevier, Amsterdam, 2006.
[5] C. De Coster and M. Henrard, *Existence and localization of solution for elliptic problem in presence of lower and upper solutions without any order*, J. Differential Equations **145** (1998), 420–452.
[6] T. Ding, R. Iannacci and F. Zanolin, *Existence and multiplicity results for periodic solutions of semilinear Duffing equations*, J. Differential Equations **105** (1993), 364–409.
[7] L. Fernandes, P. Omari and F. Zanolin, *On the solvability of a semilinear two-point BVP around the first eigenvalue*, Differential Integral Equations **2** (1989), 63–79.
[8] L. Fernandes and F. Zanolin, *Periodic solutions of a second order differential equation with one-sided growth restriction on the restoring term.* Arch. Math. **51** (1988), 151–163.
[9] A. Fonda, J.P. Gossez and F. Zanolin, *On a nonresonance condition for a semilinear elliptic problem.* Differential Integral Equations **4** (1991), 945–951.
[10] A. Fonda and R. Toader, *Lower and upper solutions to semilinear boundary value problems: an abstract approach*, Top. Meth. Nonlin. Anal. **38** (2011), 59–93.
[11] A. Fonda and F. Zanolin, *On the use of time-maps for the solvability of nonlinear boundary value problems.* Arch. Math. **59** (1992), 245–259.
[12] J.P. Gossez and P. Omari, *A necessary and sufficient condition of nonresonance for a semilinear Neumann problem*, Proc. Amer. Math. Soc. **114** (1992), 433–442.

[13] J.P. Gossez and P. Omari, *On a semilinear elliptic Neumann problem with asymmetric nonlinearities*, Trans. Amer. Math. Soc. **347** (1995), 2553–2562.

[14] P. Habets and P. Omari, *Existence and localization of solutions of second order elliptic problems using lower and upper solutions in the reversed order*, Top. Meth. Nonlin. Anal. **8** (1996), 25–56.

[15] A. Hammerstein, *Nichtlineare Integralgleichungen nebst Anwendungen*, Acta Math. **54** (1930), 117–176.

[16] J. Mawhin, *Radial solutions of Neumann problem for periodic perturbations of the mean extrinsic curvature operator*. Milan J. Math. **79** (2011), 95–112.

[17] P. Omari, *Non-ordered lower and upper solutions and solvability of the periodic problem for the Liénard and the Rayleigh equations*, Rend. Ist. Mat. Univ. Trieste **20** (1988), 54–64.

[18] Z. Opial, *Sur les périodes des solutions de l'équation différentielle* $x''+g(x)=0$, Bull. Acad. Polon. Sci. Sér. Sci. Math., Astronom. Phys. **8** (1960), 151–156.

[19] Z. Opial, *Sur l'existence des solutions périodiques de l'équation différentielle* $x''+f(x,x')x'+g(x)=p(t)$, Ann. Polon. Math. **11** (1961), 149–159.

[20] W. Reichel and W. Walter, *Sturm-Liouville type problems for the p-Laplacian under asymptotic non-resonance conditions*, J. Differential Equations **156** (1999), 50–70.

[21] A. Sfecci, *A nonresonance condition for radial solutions of a nonlinear Neumann elliptic problem*, Nonlinear Anal. **75** (2012), 6191–6202.

Author's address:

Andrea Sfecci
Università Politecnica delle Marche
Dipartimento di Ingegneria Industriale e Scienze Matematiche
Via Brecce Bianche 12
60131, Ancona, Italy
E-mail: `sfecci@dipmat.univpm.it`

Received November 13, 2014
Revised November 26, 2014

Rend. Istit. Mat. Univ. Trieste
Volume 46 (2014), 271–299

On quotient orbifolds of hyperbolic 3-manifolds of genus two

Annalisa Bruno and Mattia Mecchia

Abstract. *We analyse the orbifolds that can be obtained as quotients of genus two hyperbolic 3-manifolds by their orientation preserving isometry groups. The genus two hyperbolic 3-manifolds are exactly the hyperbolic 2-fold branched coverings of 3-bridge links. If the 3-bridge link is a knot, we prove that the underlying topological space of the quotient orbifold is either the 3-sphere or a lens space and we describe the combinatorial setting of the singular set for each possible isometry group. In the case of 3-bridge links with two or three components, the situation is more complicated and we show that the underlying topological space is the 3-sphere, a lens space or a prism manifold. Finally we present an infinite family of hyperbolic 3-manifolds that are simultaneously the 2-fold branched covering of three inequivalent knots, two with bridge number three and the third one with bridge number strictly greater than three.*

Keywords: Genus two 3-manifold, 3-bridge knot, 2-fold branched covering, quotient orbifold.
MS Classification 2010: 57M25, 57M10, 57M60

1. Introduction

In this paper we consider quotient orbifolds obtained from the smooth action of finite groups on hyperbolic 3-manifolds admitting a Heegaard splitting of genus two.

A genus n *Heegaard splitting* of a closed orientable 3-manifold M is a decomposition of M into a union $V_1 \cup V_2$ of two handlebodies of genus n intersecting in their common boundary (the Heegaard surface of the splitting). The *genus of* M is the lowest genus for which M admits a Heegaard splitting. The only 3-manifold of genus 0 is the 3-sphere while the genus one 3-manifolds are the lens spaces and $S^1 \times S^2$. We remark that two is the lowest possible genus for a hyperbolic manifold.

Quotient orbifolds of 3-manifolds admitting a Heegaard splitting of genus 2 were also studied by J. Kim by using different methods (see [9]). In his paper J. Kim considered only groups leaving invariant the Heegaard splitting of genus 2. Here we do not make this assumption. On the other hand the results in [9] include also non hyperbolic 3-manifolds.

We recall that 3-bridge knots are strictly related to genus two 3-manifolds. An *m-bridge presentation* of a knot K in the 3-sphere S^3 is a decomposition of the pair (S^3, K) into a union $(B_1, a_1) \cup (B_2, a_2)$ where B_i for $i = 1, 2$ is a 3-ball and a_i is a set of m arcs which are embedded in the standard way in B_i. We shall say that K is an m-bridge knot if m is the minimal number for which K admits an m-bridge presentation (see [17]).

Now a genus two closed orientable surface admits a *hyperelliptic involution*, (i.e. the quotient of the surface by the involution is S^2). This involution extends in a standard way to a handlebody of genus two and has the property that, up to isotopy, any diffeomorphism of the surface commutes with it. So for any genus two Heegaard splitting of M there exists an orientation preserving involution of M, which we shall also call *hyperelliptic*, which leaves invariant the Heegaard splitting and induces the hyperelliptic involution on the Heegaard surface (in contrast with the two-dimensional case, a genus two 3-manifold admits, in general, more than one hyperelliptic involution, even up to isotopy).

The quotient of M by a hyperelliptic involution is topologically S^3 and its singular set is a link L. In this case the Heegaard splitting of M naturally induces a 3-bridge presentation $(B_1, a_1) \cup (B_2, a_2)$ of L where each (B_i, a_i) is the quotient of a handlebody of the Heegaard splitting .

Conversely, a sphere that induces an m-bridge presentation of L lifts to a Heegaard surface of genus $m - 1$ of the 2-fold branched covering of L. In particular a 3-bridge presentation induces a genus two Heegaard splitting of the 2-fold branched covering and the covering involution is hyperelliptic.

We can conclude that the hyperbolic 3-manifolds of genus two are exactly the hyperbolic 3-manifolds that are the 2-fold branched covering of a 3-bridge link. This representation is not unique, in fact there exist examples of three inequivalent 3-bridge knots with the same hyperbolic 2-fold branched covering (see [10, Section 5]). In [16] it is proved that a hyperbolic 3-manifold of genus two is the 2-fold branched covering of at most three 3-bridge links. The representation of 3-manifolds as 2-fold branched coverings of knots and links have been extensively studied (see for example the survey by L. Paoluzzi [15] and the recent results by J.E. Greene [7])

In this paper we prove the following theorem about the structure of the quotient orbifolds of hyperbolic 3-manifolds of genus two. We remark that by the Thurston orbifold geometrization theorem (see [4]), any periodic diffeomorphism of a hyperbolic 3-manifold M is conjugate to an isometry of M, so we can suppose that the covering transformation of a 3-bridge link is an isometry. We recall also that a prism manifold is a Seifert 3-manifold such that its base orbifold is a 2-sphere with three singular points, two of them of singularity index 2 (see [14]).

Theorem 1.1. *Let L be a 3-bridge link and let M be the 3-manifold of genus 2 that is the 2-fold branched covering of L Suppose that M is hyperbolic and denote by G a group of orientation preserving isometries of M containing the covering involution of L.*

1. *If L is a knot, then the underlying topological space of M/G is either S^3 or a lens space and the combinatorial setting of the singular set of M/G is represented in Figure 1. If the underlying topological space is a lens space, then the covering transformation of L is central in G.*

2. *If L has two or three components, the underlying topological space of M/G is S^3, a lens space or a prism manifold. If the underlying topological space is a prism manifold, then*

    - *the symmetries of L that are projections of elements of G and fix setwise each component of L form a non-trivial cyclic group acting freely on S^3;*
    - *the symmetries of L preserving the orientation of S^3 induce a group acting faithfully on the set of the components of L that is isomorphic to the symmetry group $\mathcal{S}_n$ where n is the number of the components.*

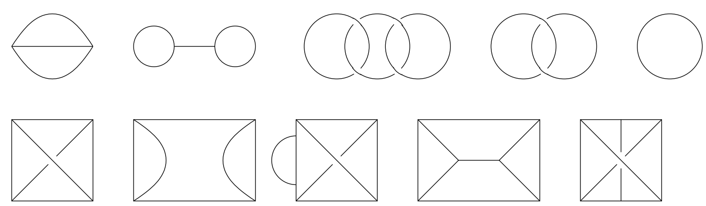


Figure 1: Admissible singular sets for M/G if L is a knot.

The proof of the theorem is based on the characterization of the isometry group of the hyperbolic 3-manifolds of genus two given in [10].

By Thurston orbifold geometrization theorem (see [4]) any finite group of diffeomorphisms of M is conjugate to a group of isometries, so the theorem holds also for finite groups of diffeomorphisms containing a hyperelliptic involution.

The underlying topological spaces are analysed both for knots and links with more than one component. In the knot case, if the underlying topological space is not S^3, then the hyperelliptic involution of L is central and each element of G projects to a symmetry of L (the 2-fold branched covering has no " hidden symmetries"). We remark also that the situation of quotient orbifolds with underlying topological space that is neither the 3-sphere nor a lens space is very special.

More details about the quotients are contained in Section 3 and 4. In particular in Figure 20 the case of knots is summarized, distinguishing for each group, that can occur as G, the possible combinatorial settings of the singular set of M/G. In principle a similar analysis could be done when L is not a knot but in the link case the number of possible graphs for the singular set of M/G is very large and we obtain a very long and complicated list.

In this paper we define a hyperelliptic involution of a genus two 3-manifold as an extension of a hyperelliptic involution of the Heegaard surface of genus two. Often in literature an involution t acting on a 3-manifold M is defined to be hyperelliptic if it gives S^3 as underlying topological of the quotient M/t. We know that a hyperelliptic involution in our sense is hyperelliptic in this broader sense. One might ask if for genus two 3-manifolds the two definitions coincide. The answer is no, in fact, in the last section of the paper, we present an infinite family of genus two 3-manifolds that are the 2-fold branched coverings of knots with bridge number strictly greater than three. Since the bridge number of the knots is not three, the covering involutions of these branched coverings are not hyperelliptic in our sense but they give S^3 as underlying topological space of the quotient. Each of these manifolds is also the 2-fold branched covering of two inequivalent 3-bridge knots. This family gives also examples of 2-fold branched coverings where a sphere giving a minimal bridge presentation of the knot does not lift to a Heegaard surface of minimal genus. A different method to obtain examples of this phenomenon and some comments about it can be found in [8].

2. Preliminaries

In this section we present some preliminary results about finite group actions on 3-manifolds, and in particular on the 3-sphere.

Proposition 2.1. *Let G be a finite group of orientation preserving diffeomorphisms of a closed orientable 3-manifold and let h be an element in G with nonempty connected fixed point set. Then the normalizer $N_G h$ of the subgroup generated by h in G is isomorphic to a subgroup of a semidirect product $\mathbb{Z}_2 \ltimes (\mathbb{Z}_a \times \mathbb{Z}_b)$, for some positive integers a and b, where a generator of $\mathbb{Z}_2$ (a h-reflection) acts on the normal subgroup $\mathbb{Z}_a \times \mathbb{Z}_b$ of h-rotations by sending each element to its inverse.*

Proof. The fixed point set of h is a simple closed curve K, which is invariant under the action of $N_G h$. By a result of Newman (see [5, Theorem 9.5]), a periodic transformation of a manifold which is the identity on an open subset is the identity. Thus the action of an element of $N_G h$ is determined by its action on a regular neighborhood of K where it is a standard action on a solid torus. Every element of $N_G h$ restricts to a reflection (strong inversion) or to a (possibly trivial) rotation on K. The subgroup of h-rotations has index one or two in $N_G h$ and is abelian. It has a cyclic subgroup (the elements fixing K pointwise) with cyclic quotient group, so it is abelian of rank at most two. $\square$

We consider now the finite subgroups of SO(4) and their action on the unit sphere $S^3 = \{(x, y, z, w) \in \mathbb{R}^4 | x^2 + y^2 + z^2 + w^2 = 1\}$. We recall that a nontrivial element of prime order in SO(4) either acts freely or fixes pointwise a simple closed curve in S^3. The finite subgroups of SO(4) are classified by Seifert and Threlfall ([19] and [20]). In Lemma 2.2 we collect some properties of these groups which we need in this paper. The results in Lemma 2.2 can be obtained from the classification and the results contained in [11], but here a direct proof seems to be more suitable. Point 2 of Lemma 2.2 is taken from [3]; the paper is no longer available, since it was a previous version of the paper [2], where this statement is not considered.

LEMMA 2.2. *Let G be a finite subgroup of* SO(4).

1. *Suppose that G is abelian, then either it has rank at most two or it is an elementary 2-group of rank three. If G acts freely on S^3, it is cyclic. If G has rank at most two, then either at most two simple closed curves of S^3 are fixed pointwise by some nontrivial element of G or G is an elementary 2-group of rank two and the whole G fixes two points (where the fixed-point sets of the three involutions meet).*

2. *If G is generalized dihedral (i.e. G is a semidirect product of an abelian subgroup of index two with a subgroup of order two whose generator acts dihedrally on the abelian subgroup of index two), then the underlying topological space of S^3/G is S^3.*

3. *If G has a cyclic normal subgroup H such that G/H is cyclic of odd order, then G is abelian.*

Proof. 1) Since G is abelian, the elements of the group can be simultaneously conjugate to block-diagonal matrices, i.e. G can be conjugate to a group such that each element has the following form:

$$\left(\begin{array}{c|c} A & 0_2 \\ \hline 0_2 & B \end{array}\right)$$

where $A, B \in \mathrm{O}(2)$ and 0_2 is the 2×2 matrix whose entries are all zero. Then, by using standard arguments from linear algebra, Point 1 can be proved.

2) In this case the group G has an abelian subgroup R of rank at most two and index two. Let ϕ and ψ be two elements in SO(4) that generate R. Since R is abelian, the action of R leaves setwise invariant a 2-dimensional plane P in $\mathbb{R}^4$, which corresponds to a simple closed curve in S^3. The group R leaves invariant also $P^\perp$, and $P \oplus P^\perp$ is a R-invariant decomposition of $\mathbb{R}^4$. We will prove that also the action of G on $\mathbb{R}^4$ leaves invariant setwise a 2-dimensional plane, that may be different from P. Let σ be an involution not in R. We have: $\psi(\sigma(P)) = \sigma(\psi^{-1}(P)) = \sigma(P)$ and $\phi(\sigma(P)) = \sigma(\phi^{-1}(P)) = \sigma(P)$, this implies that the set $Q = P \cap \sigma(P)$ is G-invariant.

If $Q = P$, then P is a 2-dimensional plane left setwise invariant by P.

If Q is a subspace of dimension 1, then we can construct explicitly another plane that is G-invariant. Let $\{v, w\}$ be an orthonormal basis of P such that $v \in Q$. Since $\psi(v) = \pm v$ and $\phi(v) = \pm v$ and P is both ψ- and ϕ-invariant, we must have that $\psi(w) = \pm w$ and $\phi(w) = \pm w$. Therefore the plane spanned by the vectors v and $w + \sigma(w)$ is G-invariant.

If $Q = \{0\}$, we fix again an orthonormal basis $\{v, w\}$ of P. If ψ or ϕ acts as a reflection on P, then a normal subgroup of G leaves pointwise invariant a 2-dimensional plane which is left setwise invariant by G. We can suppose that ψ and ϕ act as rotations on the plane P and we will prove that the plane spanned by the couple of vectors $v + \sigma(v)$ and $w - \sigma(w)$ is a G-invariant plane. In fact we have that $\psi(v + \sigma(v)) = \psi(v) + \psi(\sigma(v)) = \psi(v) + \sigma(\psi^{-1}(v))$. Supposing ψ acting on the basis in the following way: $\psi(v) = \alpha v + \beta w$, $\psi(w) = -\beta v + \alpha w$, we have that $\psi^{-1}(v) = \alpha v - \beta w$ and $\psi^{-1}(w) = \beta v + \alpha w$. Then $\psi(v + \sigma(v)) = \alpha v + \beta w + \sigma(\alpha v - \beta w) = \alpha(v + \sigma(v)) + \beta(w - \sigma(w))$ and $\psi(w - \sigma(w)) = -\beta v + \alpha w - \sigma(\beta v + \alpha w) = -\beta(v + \sigma(v)) + \alpha(w - \sigma(w))$. The same argument works with ϕ, since it is a rotation on P too, moreover $\sigma(v + \sigma(v)) = v + \sigma(v)$ and $\sigma(w - \sigma(w)) = -(w - \sigma(w))$. This completes the proof of the fact that G leaves invariant a 2- plane in $\mathbb{R}^4$.

At this point we can suppose that $\sigma(x, y, z, w) = (x, -y, z, -w)$, up to conjugacy. The whole isometry group G respects the Heegaard splitting $S^3 = T_1 \cup T_2$, where $T_1 = \{(x, y, z, w) \in S_3 : x^2 + y^2 \geq 1/2\}$ and $T_2 = \{(x, y, z, w) \in S^3 : x^2 + y^2 \leq 1/2\}$. We obtain that G acts on the solid tori T_1 and T_2 in such a way that their quotients by G are two solid balls B_1 and B_2; then S_3/G is given by the gluing of a couple of solid balls, that is known to be a 3-sphere S_3.

3) Let ψ be a generator of H and σ be an element of G such that σH is a generator of G/H. We denote by h the natural number smaller than the order of ψ such that $\sigma\psi\sigma^{-1} = \psi^h$

If ψ (or any nontrivial element of H) fixes pointwise a 2-dimensional plane P in $\mathbb{R}^4$, then σ fixes setwise the same 2-dimensional plane and $P \oplus P^\perp$ is a G-invariant decomposition of $\mathbb{R}^4$. In this case we can reduce the problem to the analysis of the finite subgroups of O(2) and we are done.

We know that ψ can be conjugate by a matrix in SO(4) in the form:

$$\left(\begin{array}{c|c} A & 0_2 \\ \hline 0_2 & B \end{array}\right).$$

If A and B have different orders, then we obtain in H a nontrivial element fixing pointwise a 2-dimensional plane, hence we get the thesis.

We can suppose that A and B have the same order.

We consider ψ as a complex matrix; if v is an eigenvector of ϕ corresponding to the eigenvalues λ, then $\sigma(v)$ is an eigenvector for ϕ corresponding to the eigenvalues λ^h.

Suppose first that $\lambda^h = \lambda$ for an eigenvalue λ; since the multiplicative order of λ equals the order of ψ, we obtain $\sigma\psi\sigma^{-1} = \psi$ and we get the thesis.

Then we can suppose that $\lambda^h \neq \lambda$ for each eigenvalue λ; σ induces a bijection on the set of eigenvalues that does not fix any of them. If one of the eigenvalues is -1, then the order of ψ is two and the matrix is diagonal (A and B have the same order); in this case ψ is central in G. Therefore ψ has two or four different eigenvalues, in any case σ^4 leaves invariant each eigenvalue, hence σ^4 commutes with ψ. Since the order of σH is odd, $\sigma^4 H$ generates G/H and we obtained that G is an abelian group.

□

3. The knot case

In this section we prove Theorem 1.1 in the knot case.

We recall that, since L is a knot, M is a $\mathbb{Z}_2$-homology sphere and, by Smith theory (see [12]), the fixed point set of an involution acting on M is either empty or a simple closed curve.

The method we use to investigate M/G is to pass through iterated quotients using a subnormal series of subgroups of G. This method can be applied thanks to the fact that, if G is a group acting on a manifold M and H is a normal subgroup of G, then the action of G induces an action of G/H on the quotient M/H.

In [10] it is proved that either there exists a hyperelliptic involution central in G or G is isomorphic to a subgroup of $\mathbb{Z}_2 \times \mathbb{S}_4$. We consider the two cases.

Case 1: G contains a central hyperelliptic involution.

Let t be a hyperelliptic involution contained in the centre of G and we suppose that L is the hyperbolic 3-bridge knot that is the projection of the fixed point set of t on $M/\langle t\rangle \cong S^3$.

The whole group G projects to $M/\langle t\rangle$ and the quotient M/G can be factorized through $(M/\langle t\rangle)/(G/\langle t\rangle)$. The Thurston orbifold geometrization theorem (see [4]) and the spherical space form conjecture for free actions on S^3 proved by Perelmann (see [13]) imply that every finite group of diffeomorphism of the 3-sphere is conjugate to a finite subgroup of SO(4). The finite subgroups of SO(4) were classified by Seifert and Threlfall ([19] and [20], see also [11] for a more geometric approach in terms of quotient orbifolds). Thus we can suppose that $G/\langle t\rangle$ is a group of isometries of S^3 leaving setwise invariant the knot L. We remark that, since L is not a trivial knot, by the positive solution of Smith conjecture (see [12]), $G/\langle t\rangle$ acts faithfully on L. In particular $G/\langle t\rangle$ is cyclic or dihedral (see Proposition 2.1).

If the fixed point set (that may be empty) of a symmetry of L is disjoint from the knot, we call it *a L-rotation*. If the fixed point set of a symmetry of L intersects the knot in two points, we call it *a L-reflection*.

Suppose first that $G/\langle t\rangle$ consists only of L-rotations. In this case $G/\langle t\rangle$ is cyclic and there are at most two simple closed curves that are fixed pointwise by some nontrivial element of $G/\langle t\rangle$ (see Lemma 2.2), thus the singular set of M/G is a link with at most three components. We recall that the quotient of a 3-sphere by an isometry with non-empty fixed point set is a 3-orbifold with the 3-sphere as underlying topological space and a trivial knot as singular set, while the quotient by an isometry acting freely is a lens space. Therefore if $G/\langle t\rangle$ is generated by elements with nonempty fixed point set the underlying topological space of the orbifold M/G isa 3-sphere, otherwise it is a lens-space.

If the group $G/\langle t\rangle$ contains a reflection of L, then $G/\langle t\rangle$ is either dihedral or isomorphic to $\mathbb{Z}_2$. In any case, by Lemma 2.2, the underlying topological space of the quotient is a 3-sphere.

If $G/\langle t\rangle \cong \mathbb{Z}_2$, then the orbifold M/G can be obtained as the quotient of $M/\langle t\rangle$ by a L-reflection; so the singular set of M/G is a theta-curve. For what concerns the dihedral case, first of all we recall that for $G/\langle t\rangle$ to be dihedral means that it is generated, up to conjugacy, by a L-reflection s and by a L-rotation r. In Lemma 2.2 we defined two tori T_1 and T_2 such that $T_1 \cup T_2 = S^3$ that are left invariant by G. Referring to the notation of the proof of Lemma 2.2, we define $C_1 = \{(x,y,z,w) \in S^3 | x^2+y^2=0\}$ and $C_2 = \{(x,y,z,w) \in S^3 | x^2+y^2=1\}$; these curves are the cores of the tori T_1 and T_2. We can suppose by conjugacy that the fixed-point sets of the L-rotations are contained in $C_1 \cup C_2$.

We consider then the singular set of $S^3/(G/\langle t\rangle)$ (where the knot L is not considered singular). The singular set of $S^3/(G/\langle t\rangle)$ is contained in the union of the projection of $C_1 \cup C_2$ with the projection of the fixed point sets of the L-reflections.

Let n be the order of the L-rotation r. We distinguish two cases: n odd or n even.

If n is odd, then all the L-reflections are conjugate. Therefore if we consider the fixed point sets of the L-reflections, the projections of these fixed point sets

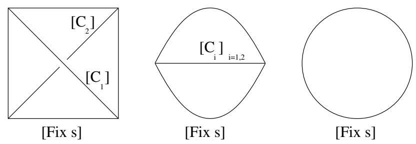


Figure 2: Possible singular sets of $S^3/(G/\langle t\rangle)$ if $G/\langle t\rangle$ is dihedral and n odd.

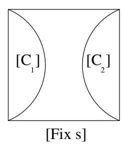


Figure 3: Another combinatorial setting build with a closed curve and two edges.

are all identified in a unique closed curve in the quotient $S^3/\langle r\rangle$. The involution s acts as a reflection also on the curves C_1 and C_2. If we consider the action of the projection of s to $S^3/\langle r\rangle$, we can describe the possible combinatorial settings for the singular set. These are represented in Figure 2. Notice that the singular set of $S^3/\langle r\rangle$ can also be empty or have only one component; the number of components of $S^3/\langle r\rangle$ depends on the number of the simple closed curves, that are fixed pointwise by any L-rotation. We denote by $[C_1]$, $[C_2]$ and $[\mathrm{Fix}\, s]$ the projections to $S^3/(G/\langle t\rangle)$ of C_1, C_2 and $\mathrm{Fix}\, s$, respectively.

However here there is something to remark. The first graph is only one of the two combinatorial settings that can be built with a closed curve and two edges with different endpoints. The second possibility is the graph in Figure 3. By Lemma 2.2 we can choose up to conjugacy $s : S^3 \subset \mathbb{R}^4 \to S^3$ as the map sending (x, y, z, w) to $(x, -y, z, -w)$, and it is easy to see that the fixed point set of r meets alternately C_1 and C_2, so this graph does not occur.

To obtain the singular set of M/G, to each graph in Figure 2 we add $[L]$, where $[L]$ is the projection of L to $S^3/(G/\langle t\rangle)$. Since s is a L-reflection, $[L]$ is an edge with endpoints contained in $[\mathrm{Fix}\, s]$. Figure 4 contains all the possibilities, up to knotting; all the edges, except $[C_1]$ and $[C_2]$, must have singularity index two.

On the other hand if n is even, then we do not have a unique conjugacy class for all the L-reflections of $G/\langle t\rangle$. Since the fixed point sets of all the elements in the same conjugacy class project to a single curve in the quotient $S^3/(G/\langle t\rangle)$, we take into consideration from now on only $\mathrm{Fix}\, s$ and $\mathrm{Fix}\, rs$, taking one representative element for each conjugacy class. In this case the

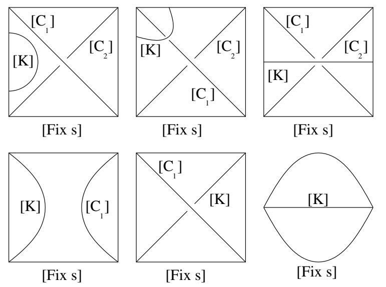


Figure 4: Possible singular sets of M/G when $G/\langle t\rangle$ is dihedral of order $2n$ with n odd.

fixed point sets of the L-reflections are not all identified in the quotient, but are collected into two different subsets of the singular set of $S^3/(G/\langle t\rangle)$, that we can denote simply by $[\mathrm{Fix}\, s]$ and $[\mathrm{Fix}\, rs]$.

Notice also that if n is even, $r^{n/2}$ is a central element in $G/\langle t\rangle$, hence we have a L-rotation fixing setwise $\mathrm{Fix}\, s$ and $\mathrm{Fix}\, rs$.

The type of action of $r^{n/2}$ on $\mathrm{Fix}\, s$ and $\mathrm{Fix}\, rs$ influences how these curves project to $S^3/\langle r\rangle$. In fact, according if this element has empty or non-empty fixed point set, different situations occur.

If $\mathrm{Fix}\, r^{n/2}$ is empty, then it acts on $\mathrm{Fix}\, s$ and $\mathrm{Fix}\, rs$ as a rotation, and the fixed point sets of s and rs project to two distinct closed curves in $S^3/\langle r\rangle$. We remark that $S^3/\langle r\rangle$ is not a 3-sphere. The projection of s to $S^3/\langle r\rangle$ is an involution which acts as a reflection on the projections of C_1 and C_2 and such that its fixed point set consists of the projections of $\mathrm{Fix}\, s$ and $\mathrm{Fix}\, rs$. The possible combinatorial structures of the singular set of $S^3/(G/\langle t\rangle)$ are presented in Figure 5.

We recall that L meets both $\mathrm{Fix}\, s$ and $\mathrm{Fix}\, sr$ and for the singular set of M/G we obtain one possibility for each of the graphs, as shown in Figure 6 (all the edges but $[C_1]$ and $[C_2]$ must have singularity index two).

If $\mathrm{Fix}\, r^{\frac{n}{2}}$ is non-empty, then clearly it coincides either with C_1 or with C_2. In this case, since $r^{\frac{n}{2}}$ commutes both with s and with rs and also with any other involution of $G/\langle t\rangle$, we obtain that $r^{n/2}$ acts as a strong inversion on both of the closed curves $\mathrm{Fix}\, s$ and $\mathrm{Fix}\, rs$ (see Lemma 2.2). Therefore the projections of $\mathrm{Fix}\, s$ and $\mathrm{Fix}\, rs$ are two arcs in $S^3/\langle r\rangle$ with both endpoints in common. Moreover the endpoints of $[\mathrm{Fix}\, s]$ and $[\mathrm{Fix}\, rs]$ in $S^3/(G/\langle t\rangle)$ coincide with the endpoints of the arc given by the projection of $\mathrm{Fix}\, r^{\frac{n}{2}}$. If C_1 is $\mathrm{Fix}\, r^{\frac{n}{2}}$, as before C_2 links $[\mathrm{Fix}\, s]$ and $[\mathrm{Fix}\, rs]$; the roles of C_1 and C_2 can be exchanged. So the possible settings for the singular set of $S^3/(G/\langle t\rangle)$ are the ones represented in Figure 7.

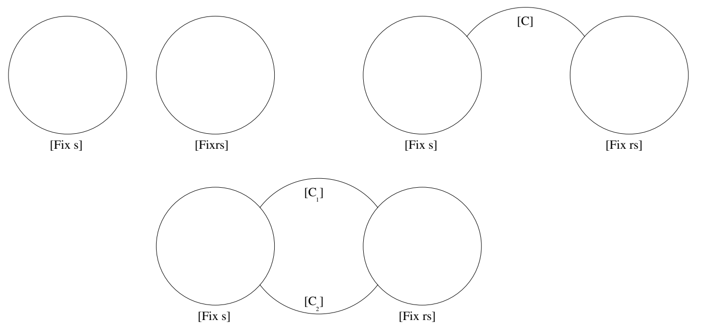


Figure 5: Possible singular sets of $S^3/(G/\langle t\rangle)$ when $G/\langle t\rangle$ is dihedral of order $2n$ with n even and the central involution acts freely.

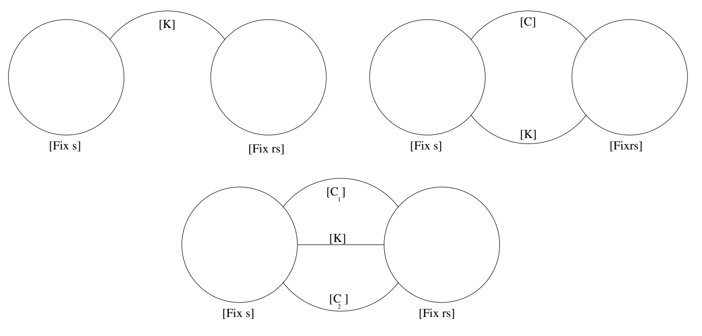


Figure 6: Possible singular sets of M/G when $G/\langle t\rangle$ is dihedral of order $2n$ with n even and the central involution acts freely.

From these we can build three different graphs that, up to knottings, are the possible singular sets of M/G, two from the first graph of Figure 7 and one from the second. The admissible results are shown in Figure 8 (again all the edges, except $[C_1]$ and $[C_2]$, have singularity index two).

Case 2: G is isomorphic to a subgroup of $\mathbb{Z}_2 \times \mathbb{S}_4$.

By [16] the number of hyperelliptic involution is at most three. We recall that by [1] and [18] two hyperelliptic involutions commute and their fixed point sets meet in two points. Here we consider groups containing a non-central hyperelliptic involution. In this case G contains a conjugacy class of hyperelliptic involutions with two or three elements (the property to be hyperelliptic is invariant under conjugation). These groups are described in the proof of [10, Theorem 1] (case c) and d) - pages 7 and 8.)

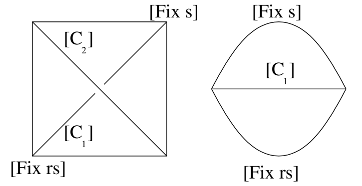


Figure 7: Possible singular sets of $S^3/(G/\langle t\rangle)$ when $G/\langle t\rangle$ is dihedral of order $2n$ with n even and the central involution does not act freely.

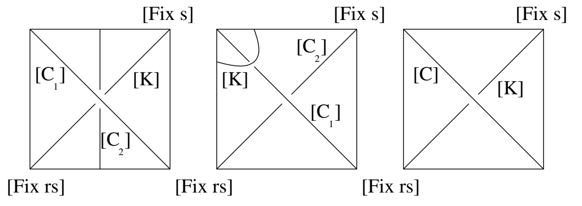


Figure 8: Possible singular sets of M/G when $G/\langle t\rangle$ is dihedral of order $2n$ with n even and the central involution does not act freely.

Case 2.1: $G \cong \mathbb{D}_8$.

This case occurs if we have a conjugacy class of hyperelliptic involutions with two elements which we denote by t_1 and t_2. By the properties of hyperelliptic involutions, t_1 and t_2 generate an elementary subgroup of rank 2 in G, and we have a subnormal series,

$$\langle t_1\rangle \lhd \langle t_1, t_2\rangle \lhd G.$$

The orbifold $M/\langle t_1\rangle$ has S^3 as underlying topological space and a knot as singular set. We consider now $M/\langle t_1, t_2\rangle$ which is diffeomorphic to the quotient of $(M/\langle t_1\rangle)$ by the projection of $\langle t_1, t_2\rangle$ to $M/\langle t_1\rangle$. Since t_2 has non empty fixed point set and is a $\mathrm{Fix}\, t_1$-reflection, we obtain that the underlying topological space of $M/\langle t_1, t_2\rangle$ is S^3 and its singular set is a knotted theta curve (a graph with two vertices and three edges; each of the three edges connects the two vertices).

Now we consider the action of $G/\langle t_1, t_2\rangle$ on $M/\langle t_1, t_2\rangle$. For a period two action on a theta-curve θ we have only three possibilities that are represented in Figure 9.

We can make some remarks on the actions represented. The first action fixes all the three edges and interchanges the vertices, therefore it acts as a rotation with period two around an axis that intersects all the three edges of the theta-curve and leaves the fixed point sets of t_1 and t_2 invariant; the second one acts

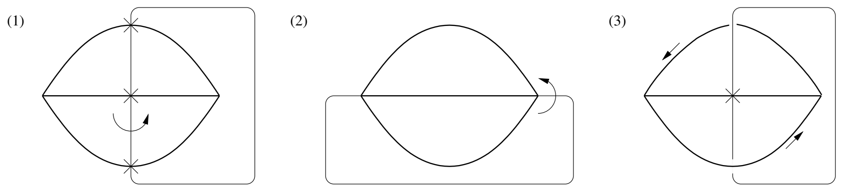


Figure 9: Possible actions with period two on a theta curve.

as a rotation of order two around an axis that contains one of the edges of the theta-curve, therefore it fixes one entire edge and the vertices and interchanges the remaining two edges; the third one acts as a rotation again of order two, but this time the axis intersects only one of the edges and in only one point, therefore it fixes only the intersection point of the theta-curve with the axis, leaves setwise invariant the edge that intersects the axis and interchanges the other two edges and the vertices. Since we already know that $\mathbb{D}_8$ interchanges the fixed point sets of t_1 and t_2, the first action is obviously not possible. Since the non-trivial element of $G/\langle t_1, t_2\rangle$ has non-empty fixed point set, the orbifold M/G has S^3 as underlying topological space; the singular set is, up to knottings, one of the two graphs represented in Figure 10 (a theta curve and a "pince-nez" graph). If we obtain a theta curve, then one of the edge has singularity index four (in this case the elements of order four in G have nonempty fixed point set).

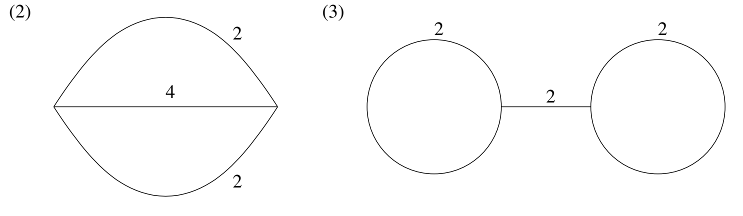


Figure 10: A theta-curve and a "pince-nez" graph.

Case 2.2: $G \cong \mathbb{Z}_2 \times \mathbb{D}_8$.

This is the second group occurring in the proof of [10, Theorem 1], when the existence of a conjugacy class of hyperelliptic involutions with two elements is assumed; we denote again the two hyperelliptic involutions by t_1 and t_2.

The first two quotients we consider are the same of the preceding case and we obtain that $M/\langle t_1, t_2\rangle$ is known.

Let A be the subgroup of G obtained by extending $\langle t_1, t_2\rangle$ by a non-trivial element of the centre of G (t_1 and t_2 cannot be in the centre of G). This means $A \cong \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$ and $A \lhd G$. The subnormal series we consider in this case is the following:

$$\langle t_1 \rangle \lhd \langle t_1, t_2 \rangle \lhd A \lhd G$$

We consider now the projection of the action of A on $M/\langle t_1, t_2 \rangle$. A acts leaving both hyperelliptic involutions t_1 and t_2 fixed. This means that it does not interchange the fixed point sets of t_1 and t_2. This time the only possible action of the three represented in Figure 9 is the first and the resulting singular set of M/A can be represented, up to knottings, as in Figure 11, by a tetrahedral graph. Since the action of $A/\langle t_1, t_2 \rangle$ is not free, the underlying topological space of M/A is S^3.

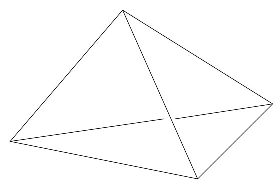


Figure 11: Singular set of M/A: a tetrahedral graph.

The last extension to take into consideration is $A \lhd G$, in particular we consider the action of G/A on M/A. We ask what actions of period two are combinatorially admissible on a tetrahedral graph. These are represented in Figure 12.

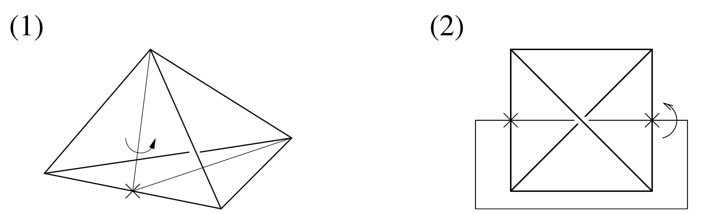


Figure 12: Possible actions of period two on a tetrahedral graph.

The actions represented are respectively a rotation around an axis containing one of the edges and meeting in a point the opposite one (1) and a rotation around an axis meeting a couple of non adjacent edges in one point (2). Therefore we obtain that M/G is an orbifold with underlying topological space a sphere S^3 and with two possible singular sets, that are the graphs represented in Figure 13 (always up to possible knottings).

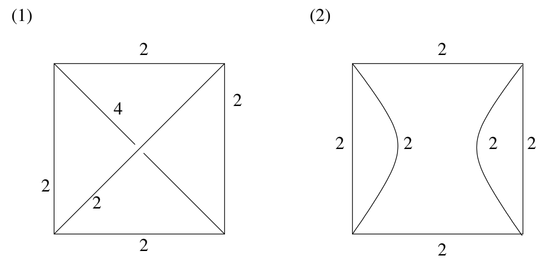


Figure 13: Graphs that can occur as singular sets of M/G, when $G \cong \mathbb{Z}_2 \times \mathbb{D}_8$.

Case 2.3: $G \cong \mathbb{A}_4$**.**

The case $G \cong \mathbb{A}_4$ is the first we encounter in which M admits a conjugacy class of three hyperelliptic involutions. This condition is satisfied in all the remaining cases (see proof of [10, Theorem 1]) and we will denote the three hyperelliptic involutions by t_1, t_2 and t_3. Just as before, we consider a subnormal series of subgroups of $\mathbb{A}_4$:

$$\langle t_1 \rangle \trianglelefteq \langle t_1, t_2 \rangle \trianglelefteq G.$$

The first two quotients we need to perform are the same encountered in the previous cases, therefore we begin analyzing the projection of the action of G on the last quotient $M/\langle t_1, t_2 \rangle$, that we recall is an orbifold with underlying topological space S^3 and singular set a theta-curve. Noticing that the index of $\langle t_1, t_2 \rangle$ in G is three, it follows that a non-trivial element $G/\langle t_1, t_2 \rangle$ acts faithfully as a rotation with period three on $M/\langle t_1, t_2 \rangle$. There is only one action of this type, that is a rotation around an axis that passes through the vertices of the theta-curve; the rotation permutes the three edges cyclically.

The action of $G/\langle t_1, t_2 \rangle$ on $M/\langle t_1, t_2 \rangle$ is clearly not free, so the underlying topological space of M/G is necessarily S^3. Moreover we can notice that the result of this action is again a theta-curve, but with different singularity indices of the edges as shown in Figure 14.

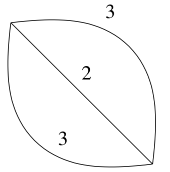


Figure 14: Singular set of M/G, when $G \cong \mathbb{A}_4$.

Case 2.4: $G \cong \mathbb{Z}_2 \times \mathbb{A}_4$**.**

The subnormal series of subgroups we use this time is the following:

$$\langle t_1 \rangle \lhd \langle t_1, t_2 \rangle \lhd A \lhd G,$$

where A is, as before, the normal subgroup of G isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$ obtained extending $\langle t_1, t_2 \rangle$ by an element of the centre of the group G. In light of what we saw in Case 2.2, we already know that M/A is an orbifold with underlying topological space the 3-sphere S^3 and singular set a tetrahedral graph. Therefore we can analyse directly the action of G/A on M/A. The group G/A has order three and there is only one admissible action of order three on a tetrahedral graph that is a rotation around an axis passing through one of the vertices of the graph that permutes cyclically the three edges containing the vertex fixed by the action, as well the three edges not containing the vertex. The singular set of M/G is shown, up to knottings, in Figure 15.

Notice that here too the singularity indices of the edges are different. Since the action of G/A is not free, the underlying topological space of M/G is S^3.

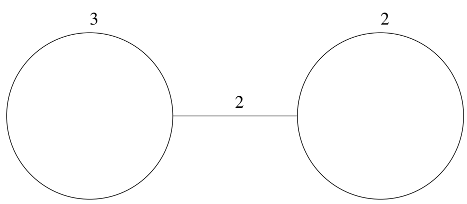


Figure 15: Singular set of M/G, when $G \cong \mathbb{Z}_2 \times \mathbb{A}_4$.

Case 2.5: $G \cong \mathcal{S}_4$**.**

Since t_1, t_2 and t_3 are conjugate and commute, the subgroup H of G isomorphic to $\mathbb{A}_4$ contains the three hyperelliptic involutions. To study M/G we consider the following subnormal series of subgroups:

$$\langle t_1 \rangle \lhd \langle t_1, t_2 \rangle \lhd H \lhd G$$

We already know that the underlying topological space of M/H is the 3-sphere and that the singular set is the theta-curve represented in Figure 14, up to knottings. Now the point is to understand how G/H acts on M/H. Since it is clear that two fixed point sets that have different singularity indices cannot be identified, the possible actions of G/H are the three represented in Figure 16.

Nevertheless in this case we can exclude some actions. For example the second action would produce as quotient a theta-curve that has one edge of singularity index four, and hence we would have an element $\alpha \in \mathcal{S}_4$ of order four

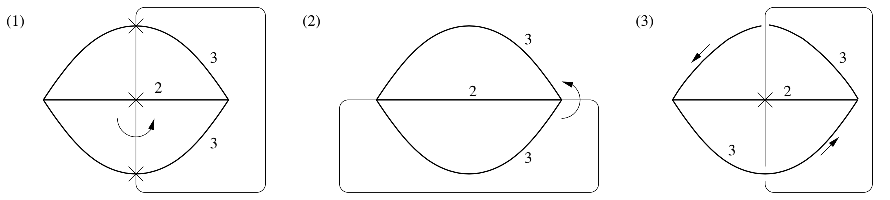


Figure 16: Possible actions of order two on the singular set of M/H.

and with non-empty fixed point set. This would mean that α^2 is a hyperelliptic involution of M, then α projects to a symmetry that fixes pointwise a 3-bridge and by the positive solution of Smith Conjecture, this is impossible.

In this case we have a reason to reject also the third action. Notice that on $M/\langle t_1, t_2\rangle$ acts also the dihedral group $\mathbb{D}_6 \cong G/\langle t_1, t_2\rangle$. The dihedral group with six elements is generated by a transformation of order two and by a transformation of order three. Since the transformation of order three acts on $M/\langle t_1, t_2\rangle$, it must act also on its singular set, which is a theta-curve. Therefore, as we have already seen, the fixed point set of this transformation must be non-empty. Since the relation between the transformation of order two and the rotation of order three is dihedral in $\mathbb{D}_6$, the fixed point set of the involution is non-empty too and the involution acts as a strong inversion on the fixed point set of the rotation of order three (see Proposition 2.1).

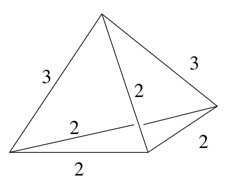


Figure 17: Singular set of M/G, when $G \cong \mathcal{S}_4$.

This means that the two fixed point sets intersect and this implies that the result of the action of $G/\langle t_1, t_2\rangle$ on S^3 produces as singular set of the quotient a theta curve with two edges of singularity index two and one edge of singularity index three. This theta curve must be contained in the singular graph of M/G, but this does not happen for the "pince-nez" graph that we would obtain as singular set of the third action (while the tetrahedral graph resulting from the first case contains such a graph). Therefore both the second and the third actions are not admissible.

Finally the only possible combinatorial setting of the singular set of the orbifold M/G is the tetrahedral graph shown in Figure 17.

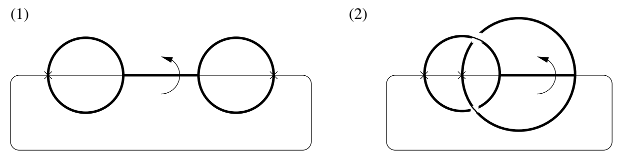


Figure 18: Reflections on the "pince-nez" graphs.

Case 2.6: $G \cong \mathbb{Z}_2 \times \mathcal{S}_4$**.**

We consider the following subnormal series of subgroups of G

$$\langle t_1 \rangle \lhd \langle t_1, t_2 \rangle \lhd A \lhd J \lhd G,$$

where A, as before, is the subgroup of G isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$ obtained extending $\langle t_1, t_2 \rangle$ by an element that belongs to the centre of the group G and A is the normal subgroup of G isomorphic to $\mathbb{Z}_2 \times \mathbb{A}_4$ and containing A as normal subgroup. We already know that the underlying topological space of M/J is S^3 and that its singular set is the "pince-nez" graph represented in Figure 15 (always up to knottings).

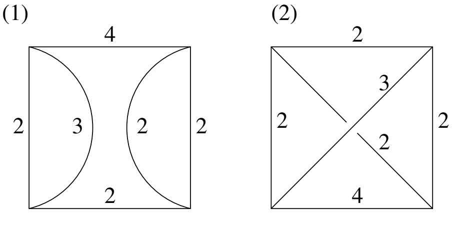


Figure 19: Admissible singular sets for M/G, when $G \cong \mathbb{Z}_2 \times \mathcal{S}_4$.

It is clear that the action of G/J on M/J is a transformation with period two, since $\mathbb{Z}_2 \times \mathbb{A}_4$ has index two in $\mathbb{Z}_2 \times \mathcal{S}_4$. The peculiarity of this case is that the singular set of the quotient is different according to the knotting of the "pince-nez" graph. In fact the action on the graph is combinatorially unique, but the result depends on the order of intersections of the two loops with the axis of the involution. The action with period two on these graphs is always a rotation that fixes pointwise the middle edge and leaves invariant the two loops, operating a reflection on each of them, but we have to distinguish between the two cases represented by the graphs of Figure 18. In the first case one of the two arcs, in which is divided the axis by the the intersection points between the axis and the first loop, does not contain any intersection point of the second loop and the axis; in the second case both the arcs contain an intersection point of the second loop and the axis.

We obtain two possible singular sets for the orbifold M/G, which are shown, up to knottings, in Figure 19. In any case the underlying topological space is S^3.

In the table in Figure 20 we summarize the situation.

4. The link case

In this section our aim is to generalize the work done in Chapter 3 on 2-fold branched coverings of 3-bridge knots extending our considerations to 2-fold branched coverings of 3-bridge links. In light of the definition of bridge number we can deduce that 3-bridge links can have a maximum of three components. Moreover, in contrast with the case of hyperbolic knots, the constituent knots of a hyperbolic link can also be all trivial.

We denote by t the hyperelliptic involution that is the covering transformation of L. In the last part of the proof of [16, Theorem 1] it is proved that, if L has more than one component, then t is central in G. Therefore we have that $M/G \cong (M/\langle t\rangle)/(G/\langle t\rangle)$ and each element of G projects to a symmetry of L.

of S^3 lifts to a finite group acting on M and containing t in its centre; by Thurston orbifold geometrization Theorem we can suppose that it is contained up to conjugacy in G.

In this case, since a long list of graphs would be produced (with respect to the one of the previous chapter), we don't consider the singular set of the quotients. Our only aim this time is to analyse what the underlying topological space of this quotient is.

If L is a link, the symmetry group of L is not as simple as when L is a knot: it is no more true that it is a subgroup of a dihedral group. Let G_0 be the normal subgroup of G which consists of the elements fixing setwise each component of $\operatorname{Fix} t$. We denote by $\bar{G}$ (resp. $\bar{G}_0$) the quotient $G/\langle t\rangle$ (resp. $G_0/\langle t\rangle$); the group $\bar{G}_0$ fixes setwise each component of L.

Clearly we have that the quotient group $\bar{G}/\bar{G}_0$ is a subgroup of the symmetry group $\mathcal{S}_n$, where n is the number of components of the link L, hence in our case $\bar{G}/\bar{G}_0$ is either a subgroup of $\mathcal{S}_3$ or a subgroup of $\mathcal{S}_2$. By Proposition 2.1, $\bar{G}_0 \leq \mathbb{Z}_2(\mathbb{Z}_m \times \mathbb{Z}_n)$ for some $m, n \in \mathbb{N}$, i.e. $\bar{G}_0$ is isomorphic to a subgroup of a generalized dihedral group. Therefore $\bar{G}_0$ has an abelian subgroup of rank at most two of index at most two. We separately analyse the different cases.

Case 1: $\bar{G}_0 \cong \mathbb{Z}_2(\mathbb{Z}_m \times \mathbb{Z}_n)$, i.e. $\bar{G}_0$ is generalized dihedral.
We recall that the underlying topological space of $M/\langle t\rangle$ is S^3. By Lemma 2.2, we obtain that the underlying topological space of the orbifold $(M/\langle t\rangle)/\bar{G}_0 \cong M/G_0$ is S^3. What remains to study now is the action of G/G_0 on M/G_0, but G/G_0 is either a subgroup of S_3 or a subgroup of S_2, and hence it is cyclic or

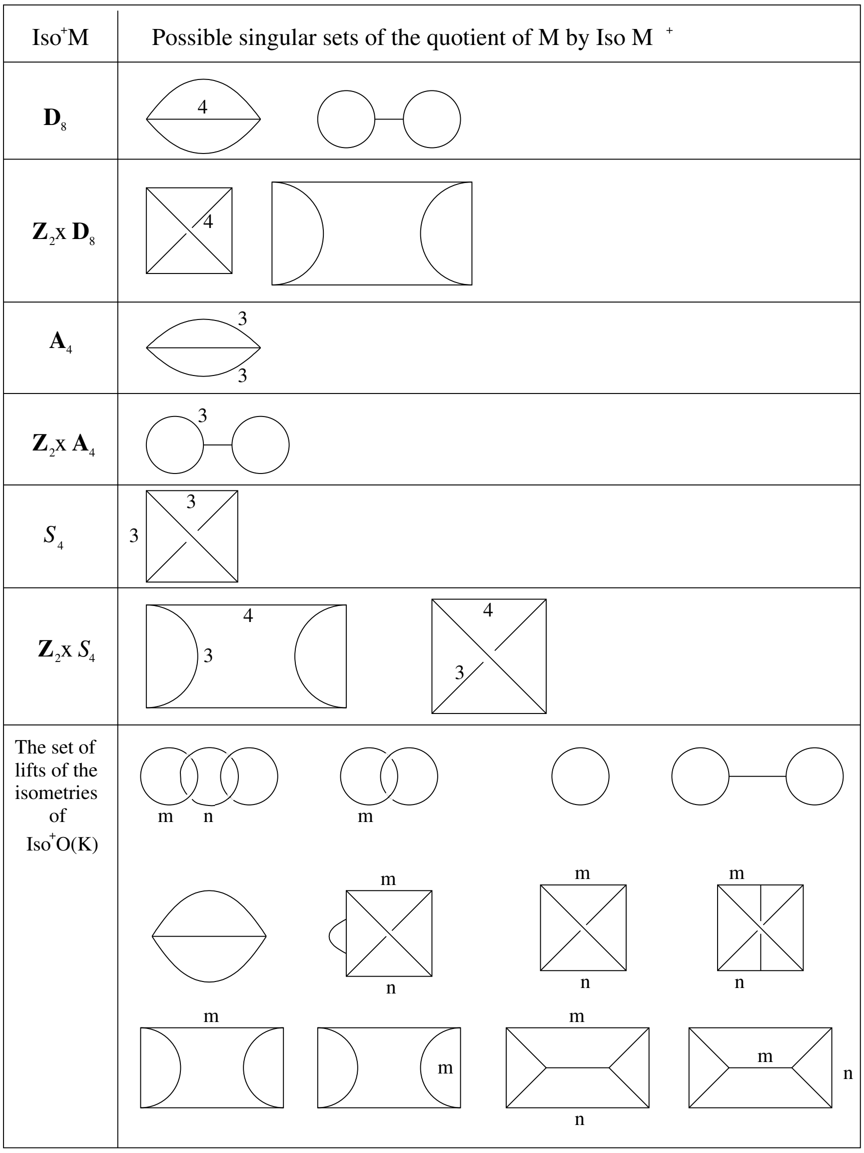


Figure 20: Admissible singular set for M/G. For the edges of singularity index 2 we omit the label.

dihedral. By Lemma 2.2, we obtain that the underlying topological space of $M/G \cong (M/G_0)/(G/G_0)$ is either S^3 or a lens space.

Case 2: $\bar{G}_0 \leq (\mathbb{Z}_m \times \mathbb{Z}_n)$, i.e. $\bar{G}_0$ is abelian.
In this case what is missing with respect to the previous one is the action of a strong inversion on a component of L. Again two cases can occur: rank $\bar{G}_0 = 1$ or rank $\bar{G}_0 = 2$.

Case 2.1: rank $\bar{G}_0 = 2$.
In this case $\bar{G}_0$ admits a subgroup either of type $\mathbb{Z}_p \times \mathbb{Z}_p$ with p an odd prime or of type $\mathbb{Z}_2 \times \mathbb{Z}_2$.

We begin showing that the first case cannot occur. Suppose that there exists a subgroup D of $\bar{G}_0$ such that $D \cong \mathbb{Z}_p \times \mathbb{Z}_p$ for some prime $p > 2$. Let L_i be a connected component of the link L and let X_i be the subgroup of D made of the isometries that fix pointwise the component L_i. The group D/X_i acts faithfully on the component L_i, hence the group D/X_i must be either cyclic or dihedral. Clearly it cannot be dihedral, being a quotient of an abelian group. This means that D/X_i is cyclic, in particular isomorphic to $\mathbb{Z}_p$. Therefore, by Lemma 2.2 the group X_i is one of the two subgroups of D isomorphic to $\mathbb{Z}_p$ that admit nonempty fixed-point set. This argument holds true for all the components of L: the components of L are two and, since D can be simultaneously conjugate to block-diagonal matrices (see proof of Lemma 2.2), L is the Hopf Link. Since the Hopf link is a well known 2-bridge link, this leads to a contradiction.

Suppose now that $\bar{G}_0$ contains a subgroup D isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$, in this case we prove that the underlying topological space of M/G_0 is S^3. Again we consider the quotient D/X_i, where X_i is the normal subgroup of D consisting of the isometries that fix pointwise the component L_i of L. In this case D/X_i can be either cyclic or the whole group D.

We suppose that X_i is trivial for some component L_i, thus D contains an L_i-reflection which we denote by α. Suppose that exists an element $\beta \in \bar{G}_0$ of order different from two. This element must act as a rotation on the i-th component of L, but then, since α is a L_i-reflection, we have that $\alpha\beta\alpha^{-1} = \beta^{-1}$. This implies that $\bar{G}_0$ admits a dihedral subgroup, that leads to a contradiction, being $\bar{G}_0$ abelian. Therefore we obtain that $\bar{G}_0$ is isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$. In this case we are done. In fact, since $\mathbb{Z}_2 \times \mathbb{Z}_2$ is always generated by a couple of involutions with non empty fixed point set, then the underlying topological space of $M/G_0 \cong (M/\langle t\rangle)/(G_0/\langle t\rangle)$ is S^3.

On the other hand, if X_i is non-trivial for each component L_i of the link L, then again we obtain a contradiction by Lemma 2.2. In fact, if D contains two involutions with non empty fixed point set, then L should be the Hopf link. If D contains three involutions with non empty fixed point set, the components of L should intersect in two points and this is impossible.

Summarizing we obtained that if $\bar{G}_0$ has rank two, then M/G_0 has always underlying topological space S^3. As in the previous case, the group G/G_0 is either cyclic or isomorphic to $\mathcal{S}_3$, and this implies that the underlying topological space of M/G is either S^3 or a lens space.

Case 2.2: rank $\bar{G}_0 = 1$, i.e. $\bar{G}_0$ is cyclic.
The quotient of S^3 by a cyclic group of isometries is an orbifold with underlying topological space either S^3 or a lens space.

We distinguish two cases: $\bar{G}_0$ admits at least one element acting with non empty fixed point set or $\bar{G}_0$ acts freely.

If $\bar{G}_0$ does not act freely, then each element in the normalizer of G_0 fixes setwise each curve fixed by a nontrivial element of G_0. In fact the different curves are fixed pointwise by elements of different order. This means that G fixes setwise at least a closed curve, therefore, thanks to Proposition 2.1, we can say that G must be a subgroup of a generalized dihedral group and we are done by Lemma 2.2.

On the other hand if $\bar{G}_0$ acts freely, then the analysis of the quotient M/G is more complicated. If $\bar{G} = \bar{G}_0$, then the underlying topological space is a lens space.

Otherwise the quotient G/G_0 is isomorphic to $\mathbb{Z}_2$, $\mathbb{Z}_3$ or $\mathcal{S}_3$.

If $\bar{G}/\bar{G}_0 \cong \mathbb{Z}_3$, then by Lemma 2.2 the group $\bar{G}$ is abelian and the underlying topological space of M/G is either S^3 or a lens space. If the components are three and $\bar{G}/\bar{G}_0 \cong \mathbb{Z}_2$, then one of the component of L is fixed setwise by $\bar{G}$ and we are done.

In the remaining cases we can suppose that $\bar{G}$ has an abelian subgroup of index two. Up to now we were able to prove that the underlying topological space of M/G is either S^3 or a lens space, unfortunately in the remaining cases some groups can admit as underlying topological space of the quotient a prism manifold. By analyzing the remaining groups case by case, we can deduce more information about the situations in which a prism manifold can occur, but at this point we prefer to give a shorter argument that simply exclude tetrahedral, octahedral and icosahedral manifolds as underlying topological space of the quotient. Since $\bar{G}$ has an abelian subgroup of index two, the group leaves invariant a fibration of S^3 (see [11]). The quotient orbifold $S^3/\bar{G}$ admits a Seifert fibration induced by the fibration of S^3 left invariant by G. By [11, Lemma 2], the base 2-orbifold B of $S^3/\bar{G}$ is the quotient of S^2 by the action of $\bar{G}$ (which is possibly non-faithfully.) Since G has an abelian subgroup of index 2, either B has a disk as underlying topological space or it is a 2-sphere with at most one singular point of index strictly greater than 2 (i.e. the cases with base 2-orbifold $S^2(2,3,3)$ $S^2(2,3,4)$ and $S^2(2,3,5)$ in [11, Table 4] are excluded).

If B has underlying topological space the 2-disk, then, by [6, Proposition 2.11], the underlying topological space of $S^3/\bar{G}$ (and hence of M/G) is either S^3 or a lens space.

On the other hand, if the base 2-orbifold has no boundary component, then, by forgetting the orbifold singularity of the fibers, we obtain, from the Seifert fibration of $S^3/\bar{G}$, a Seifert fibration of the underlying topological space of $S^3/\bar{G}$. The base 2-orbifold of the underlying topological space of $S^3/\bar{G}$ can be obtained from B by dividing the index of the singular points by the singularity index of the corresponding fibers. The Euler number of the fibration is not affected by the singularity forgetting process. Since $S^2(2,3,3)$ $S^2(2,3,4)$ and $S^2(2,3,5)$ are excluded as base 2-orbifolds, it turns out that the underlying topological space of $S^3/\bar{G}$ is S^3, a lens space or a prism manifold (see [14] and [11, Table 2,3,4])

5. An example

In this section we describe an infinite family of hyperbolic 3-manifolds such that each of them is the 2-fold branched covering of three inequivalent knots, two of them with bridge number equal to three and the third one with bridge number strictly greater than three.

For any triple of nonzero integers (i, j, k) we can define the 3-bridge knot K_{ijk} presented in Figure 21.

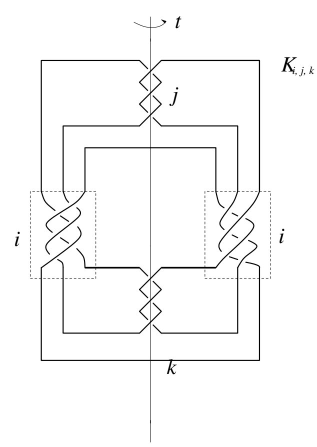


Figure 21: The knot K_{ijk}.

In Figure 21 we have also drawn the axis of a strong inversion t of K_{ijk}. Let $\mathcal{O}(K_{ijk})$ be the orbifold with underlying topological space S^3 and K_{ijk} as singular set of index 2. We consider the quotient orbifold $\mathcal{O}(\theta_{ijk}) := \mathcal{O}(K_{ijk})/\langle t\rangle$ which has S^3 as underlying topological space. The singular set is a theta-curve θ_{ijk} with edges e_1, e_2 and e_3 and constituent knots $A_1 = e_2 \cup e_3$, $A_2 = e_1 \cup e_3$

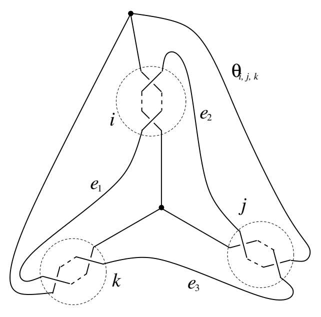


Figure 22: The theta curve θ_{ijk}.

and $A_3 = e_1 \cup e_2$. The theta curve θ_{ijk} is represented in Figure 22 (how to obtain this planar diagram of θ_{ijk} is explained in [21]).

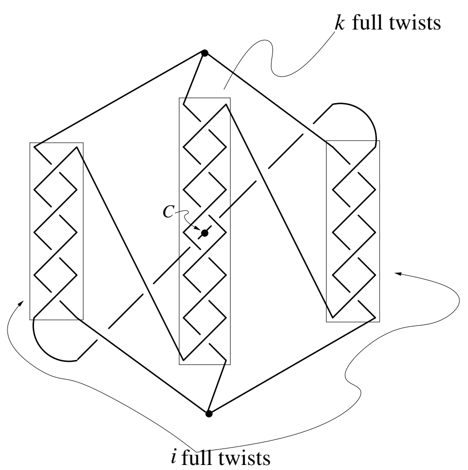


Figure 23: Symmetry of θ_{iik}.

The three constituent knots are trivial and the preimage of e_1, e_2 respectively e_3 in the 2-fold cyclic branched covering of S^3 along A_1, A_2 respectively A_3 is $K_1 = K_{ijk}$, $K_2 = K_{jki}$ respectively $K_3 = K_{kij}$. Finally, if we take the two fold branched covering of K_1, K_2 or K_3, we get the same manifold M: the manifold M is the $\mathbb{D}_4$ covering of $\mathcal{O}(\theta_{ijk})$. In [21] it is proved that M is hyperbolic for $|i|, |j|, |k|$ sufficiently large. The isometry group of M was studied in [10]. Here we consider the case where two of the indices are equal, while the third one is different. If $\{i, j, k\}$ is not of the form $\{lm, (l-1)m\}$, with m

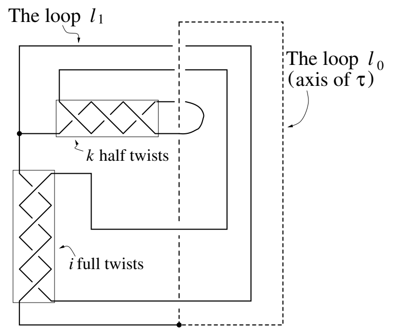


Figure 24: The singular set of M/G.

and l integers and l even, then the isometry group of M is isomorphic to $\mathbb{D}_8$ and it does not contain any orientation-reversing isometry (see [10, page 8]). We denote by G the isometry group of M and we suppose $i = j$. In G there are three conjugacy classes of involutions, two of them consist of hyperelliptic involutions and correspond to the 3-bridge knots. One of the hyperelliptic involutions is central in G. The involutions in the third conjugacy class are not hyperelliptic. In this section we prove that the quotient orbifold of M by one of these involutions has underlying topological space S^3; the singular set of this orbifold is a knot that has bridge number different from three. If $i = j$, the theta curve θ_{iik} has a symmetry of order 2 exchanging the two vertices and leaving setwise invariant only one of the edges. In the diagram of θ_{iik} presented in Figure 23 the symmetry is evident, it consists of a π-rotation around the point C. We denote this involution by τ.

The quotient of $\mathcal{O}(\theta_{iik})$ by τ is an orbifold with S^3 as underlying topological space and with the knotted pince-nez graph represented in Figure 24 as singular set. This orbifold is M/G (this situation corresponds to Case 2.1. in Section 3). We denote by l_0 and l_1 the loops of the pince-nez graph. In particular let l_1 be the projection of the axis of τ (the dotted line in Figure 24).

Now we consider the orbifold obtained by taking the 2-fold covering of M/G branched over the loop l_1. We remark that the loop l_1 is a trivial knot. This gives an orbifold $\mathcal{O}(\Gamma_{ik})$ with S^3 as underlying topological space and the theta curve Γ_{ik} represented in Figure 25 as singular set (of singularity index 2).

To draw explicitly Γ_{ik} we use a planar diagram of the singular set of M/G where l_1 has a trivial projection. In Figure 26 how to obtain such a diagram is explained. From this representation of the graph it is easy to reconstruct a diagram of Γ_{ik}. To help the reader, in Figure 25 we represent explicitly the axis of the involution acting on $\mathcal{O}(\Gamma_{ik})$ that gives M/G as quotient.

The orbifold $\mathcal{O}(\Gamma_{ik})$ is the quotient of M by the group generated by the

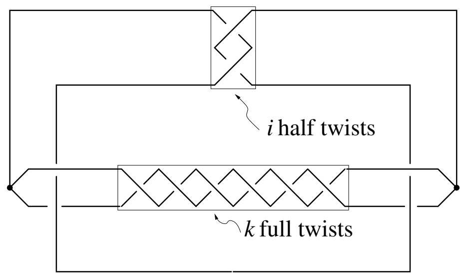


Figure 25: Γ_{ik}

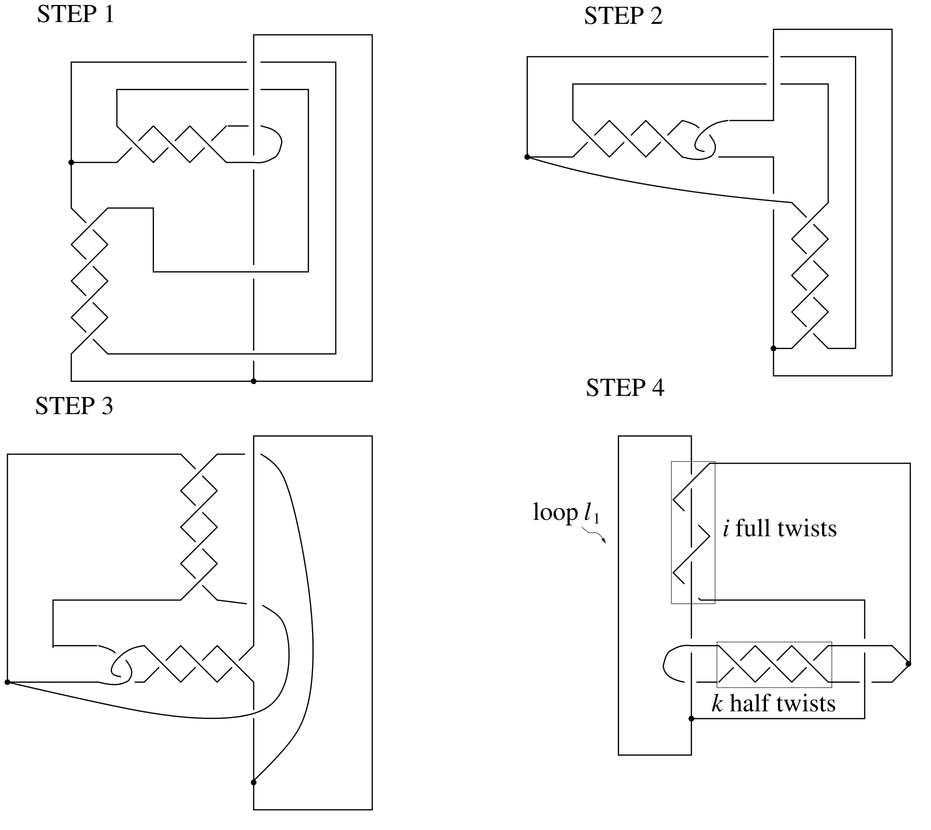


Figure 26: Equivalent diagrams of the singular set of M/G

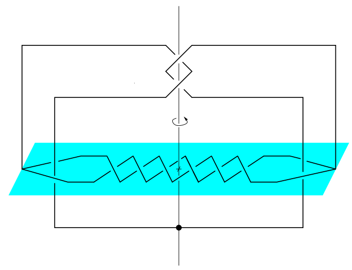


Figure 27: Γ_{ik}

central hyperelliptic involution and by one involution that is not hyperelliptic (we call u such an involution). The orbifold $M/\langle u\rangle$ is the two fold covering of $\mathcal{O}(\Gamma_{ik})$ branched over the appropriate constituent knot. We remark that all the three constituent knots of Γ_{ik} are trivial. This implies that the underlying topological space of $M/\langle u\rangle$ is S^3 and that M is the 2-fold covering of S^3 branched over the knot that is the singular set of $M/\langle u\rangle$. We denote this knot by L_{ik}. Since u is not hyperelliptic, the bridge number of L_{ik} is different from 3. Since the only knot with bridge number one is the unknot andthe 2-fold branched coverings of the 2-bridge knots are the lens spaces, these knots have bridge number strictly greater than three. We give a diagram of L_{ik} in Figure 28 and Figure 29 shows a procedure to get it. Indeed an explicit diagram of L_{ik} is not necessary to get the properties we need. We remark that the Heegaard splitting induced by a minimal bridge presentation of L_{ik} is not of minimal genus.

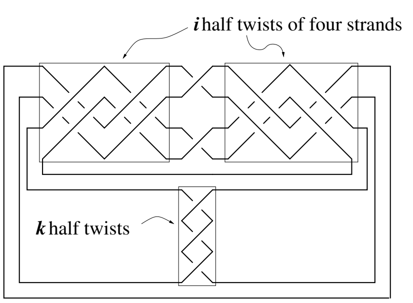


Figure 28: the knot L_{ik}

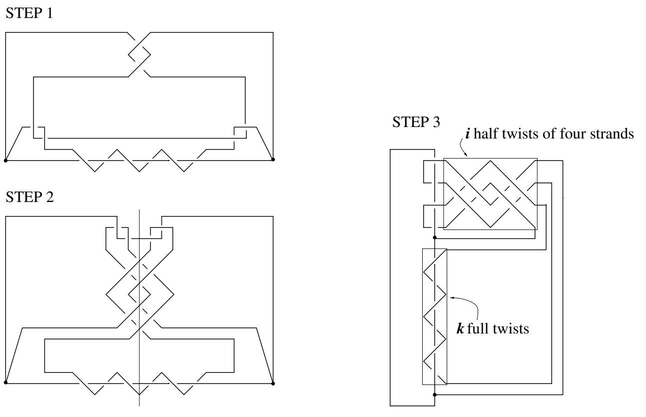


Figure 29: Equivalent diagrams of Γ_{ik}

Acknowledgements

The second author was partially supported by the FRA 2013 grant "Geometria e topologia delle varietà", Università di Trieste, and by the PRIN 2010-2011 grant "Varietà reali e complesse: geometria, topologia e analisi armonica".

References

[1] J. Berge and M. Scharlemann, *Multiple genus 2 Heegaard splittings: a missed case*, Algebr. Geom. Topol. **11** (2011), no. 3, 1781–1792.

[2] M. Boileau, S. Boyer, R. Cebanu, and G. Walsh, *Knot commensurability and the Berge conjecture*, Geom. Topol. **16** (2012), no. 2, 625–664.

[3] M. Boileau, S. Boyer, and G. Walsh, *On commensurability of knot complements*, preprint.

[4] M. Boileau, B. Leeb, and J. Porti, *Geometrization of 3-dimensional orbifolds*, Ann. of Math. **162** (2005), no. 1, 195–290.

[5] G. E. Bredon, *Introduction to compact transformation groups*, Academic Press, New York-London, 1972, Pure and Applied Mathematics, Vol. 46.

[6] W.D. Dunbar, *Fibered orbifolds and crystallographic groups*, Ph.D. thesis, Princeton University, 1981.

[7] J.E. Greene, *Lattices, graphs, and Conway mutation*, Invent. Math. **192** (2013), no. 3, 717–750.

[8] J. Johnson, *Bridge positions and branched covers*, blog post available at http://ldtopology.wordpress.com/?s=Bridge+positions+and+branched+covers.

[9] J. Kim, *Structures of geometric quotient orbifolds of three dimensional G-manifolds of genus two*, J. Korean Math. Soc. **46** (2009), no. 4, 859–893.

[10] M. Mecchia and M. Reni, *Hyperbolic 2-fold branched covering of 3-bridge knots*, Kobe J. Math. **17** (2000), no. 1, 1–19.

[11] M. Mecchia and A. Seppi, *Fibered spherical 3-orbifolds*, To appear in Rev. Mat. Iberoam. - arXiv:1307.0641v2 [math.GT].

[12] J. Morgan and H. Bass (editors), *The Smith conjecture*, Pure Appl. Math., vol. 112, Academic Press, Orlando, FL, 1984.

[13] J. Morgan and G. Tian, *Ricci flow and the Poincaré conjecture*, Clay Mathematics Monographs, vol. 3, American Mathematical Society, Providence, RI; Clay Mathematics Institute, Cambridge, MA, 2007.

[14] P. Orlik, *Seifert manifolds*, Lecture Notes in Mathematics, Vol. 291, Springer-Verlag, Berlin-New York, 1972.

[15] L. Paoluzzi, *Hyperbolic knots and cyclic branched covers*, Publ. Mat. **49** (2005), 257–284.

[16] M. Reni, *The isometry groups of genus 2 hyperbolic 3-manifolds*, Kobe J. Math. **15** (1998), no. 1, 77–84.

[17] D. Rolfsen, *Knots ans links*, Mathematics Lecture Series, vol. 7, Publish or Perish, Inc., Houston, TX, 1990.

[18] H. Rubinstein and M. Scharlemann, *Genus two Heegaard splittings of orientable 3-manifolds*, Proceedings of the Kirbyfest (Berkeley, CA, 1998), Geom. Topol. Monogr., vol. 2, Geom. Topol. Publ., Coventry, 1999, pp. 489–553.

[19] W. Threlfall and H. Seifert, *Topologische Untersuchung der Diskontinuitätsbereiche endlicher Bewegungsgruppen des dreidimensionalen sphärischen Raumes*, Math. Ann. **104** (1930), 1–70.

[20] W. Threlfall and H. Seifert, *Topologische Untersuchung der Diskontinuitätsbereiche endlicher Bewegungsgruppen des dreidimensionalen sphärischen Raumes (Schluß)*, Math. Ann. **107** (1932), 543–586.

[21] B. Zimmermann, *On hyperbolic knots with the same m-fold and n-fold cyclic branched coverings*, Topology Appl. **79** (1997), no. 2, 143–157.

Authors' addresses:

Annalisa Bruno
Università di Trieste
Dipartimento di Matematica e Geoscienze
via Valerio 12/1
34127, Trieste, Italy
E-mail: `annalisa.bruno18@gmail.com`

Mattia Mecchia
Università di Trieste
Dipartimento di Matematica e Geoscienze
via Valerio 12/1
34127, Trieste, Italy
E-mail: `mmecchia@units.it`

Received November 25, 2014
Revised December 4, 2014

Editorial Note

The Managing Editors would like to remind the authors of the ethical rules set out in the document of IMU “Best current practices for Journals” (see www.mathunion.org). In particular, we stress the following:

“Authors must abide by high standards of research integrity and good scholarship. It is the responsibility of authors to submit a well written, mathematically correct article, if necessary seeking advice if it is not written in their native language, to clearly describe any novel and non-trivial content, and to suitably acknowledge the contributions of others, including referees. Submission of a paper to a journal implies that it is not currently under consideration by any other journal, and that any substantial overlap with other published or submitted papers is duly acknowledged. In addition authors should be responsive to correspondence with the journal. Multiple authors should communicate fully, speak with one voice, and accept mutual responsibility in their communications with the journal. All authors are expected to have materially contributed to the paper, and to be familiar with its content.”

Alessandro Fonda
Emilia Mezzetti
Pierpaolo Omari
Maura Ughi

Contents

Editorial Policy

The journal *Rendiconti dell'Istituto di Matematica dell'Università di Trieste* publishes original articles in all areas of mathematics. Special regard is given to research papers, but attractive expository papers may also be considered for publication.

The journal usually appears in one issue per year. Additional issues may however be published. In particular, the Managing Editors may consider the publication of supplementary volumes related to some special events, like conferences, workshops, and advanced schools.

All submitted papers will be refereed. Manuscripts are accepted for review with the understanding that the work has not been published before and is not under consideration for publication elsewhere.

The journal can be obtained by exchange agreements with other similar journals.

Instructions for Authors

Authors are invited to submit their papers by e-mail directly to one of the Managing Editors in PDF format.

All the correspondence regarding the submission and the editorial process of the paper are done by e-mail.

Papers have to be written in one of the following languages: English, French, German, or Italian. Abstracts should not exceed ten printed lines, except for papers written in French, German, or Italian, for which an extended English summary is required.

After acceptance, manuscripts have to be prepared in LaTeX using the style *rendiconti.cls* which can be downloaded from the web page.

Any figure should be recorded in a single PDF, PS (PostScript), or EPS (Encapsulated PostScript) file.

INDIRIZZO DEL COMITATO DI REDAZIONE

Rendiconti dell'Istituto di Matematica dell'Università di Trieste
Dipartimento di Matematica e Geoscienze - Università - Via Valerio 12/1 - 34127 Trieste (Italy)
Tel. +39-040-5582635 — Fax +39-040-5582636 — Pagina web: http://rendiconti.dmi.units.it

Direttore Responsabile: DANIELE DEL SANTO

Periodico registrato il 25 Settembre 1968 al n. 358 del registro dei periodici del Tribunale di Trieste

Annual subscription rate: €50,00 (one issue, inclusive of any supplements).
Subscription inquiries should be sent to
EUT, Edizioni Università di Trieste, Piazzale Europa 1, 34127 Trieste, Italia.

ISSN 0049-4704

Contents

ISSN 0049-4704