Publicaciones Matemáticas del Uruguay

Volumen 12, Año 2011

Publicaciones Matemáticas del Uruguay

Published by:
IMERL-Facultad de Ingeniería
CMAT-Facultad de Ciencias
Universidad de la República
http://imerl.fing.edu.uy/pmu

ISSN: 0797-1443

Credits:
Cover design: J. Rodriguez Hertz
LaTeX editor: J. Rodriguez Hertz
using LaTeX's 'confproc' package, version 0.7 (by V. Verfaille)

Mario Wschebor, former editor of the Publicaciones Matemáticas del Uruguay, and founder of the IFUM, passed away while this volume was in print. We dedicate this volume to his memory.

Contents

Preface

A review of some recent results on Random Polynomials over $\mathbb{R}$ and over $\mathbb{C}$.
DIEGO ARMENTANO 1

Rice formulas and Gaussian waves II.
JEAN-MARC AZAÏS, JOSÉ R. LEÓN, and MARIO WSCHEBOR 15

On automorphism groups of fiber bundles
MICHEL BRION 39

On the focusing of Cramér - von Mises test
ALEJANDRA CABAÑA and ENRIQUE CABAÑA 67

Feuilletage de Hirsch, mesures harmoniques et g-mesures
BERTRAND DEROIN and CONSTANTIN VERNICOS 79

On existence of smooth critical subsolutions of the Hamilton-Jacobi Equation
ALBERT FATHI 87

Paths towards adaptive estimation for Instrumental Variable Regression
JEAN-MICHEL LOUBES and CLÉMENT MARTEAU 99

Semisimple Hopf algebras and their representations
SONIA NATALE 123

An example concerning the Theory of Levels for codimension-one foliations
ANDRÉS NAVAS 169

Accessibility and abundance of ergodicity in dimension three: a survey.
FEDERICO RODRIGUEZ HERTZ, JANA RODRIGUEZ HERTZ, and RAÚL URES 177

PREFACE

This volume contains the proceedings of the Colloquium celebrating the opening of the Franco-Uruguayan Institute of Mathematics (IFUM), which is an International Associate Laboratory (LIA) of the French National Center for Scientific Research (CNRS). This meeting took place in December 8-11, 2009, in Punta del Este, Uruguay, and was enriched with the participation of many specialists in the areas of Probability, Algebra and Dynamical Systems, from Argentina, France and Uruguay.

We are grateful to the Scientific Committee, specially to Viviane Baladi, for entrusting to us the edition of these proceedings. We are also indebted to CSIC, PEDECIBA-Matemática and IFUM for supporting the edition of this volume.

Last but not least, we counted on the generous collaboration of the authors and the referees, without whom this volume would have not been possible. We wish to express our gratitude to all of them.

Jana Rodriguez Hertz
Montevideo, November 2011.

A REVIEW OF SOME RECENT RESULTS ON RANDOM POLYNOMIALS OVER $\mathbb{R}$ AND OVER $\mathbb{C}$.

DIEGO ARMENTANO

Abstract. This article is divided in two parts. In the first part we review some recent results concerning the expected number of real roots of random system of polynomial equations. In the second part we deal with a different problem, namely, the distribution of the roots of certain complex random polynomials. We discuss a recent result in this direction, which shows that the associated points in the sphere (via the stereographic projection) are surprisingly well-suited with respect to the minimal logarithmic energy on the sphere.

1. Introduction

Let us consider a system of m polynomial equations in m unknowns over a field $\mathbb{K}$,

$$f_i(x) := \sum_{\|j\|\leq d_i} a_j^{(i)} x^j \quad (i = 1, \ldots, m). \tag{1}$$

The notation in (1) is the following: $x := (x_1, \ldots, x_m)$ denotes a point in $\mathbb{K}^m$, $j := (j_1, \ldots, j_m)$ a multi-index of non-negative integers, $\|j\| = \sum_{h=1}^m j_h$, $x^j = x^{j_1} \cdots x^{j_m}$, $a_j^{(i)} = a_{j_1,\ldots,j_m}^{(i)}$, and d_i is the degree of the polynomial f_i.

We are interested in the solutions of the system of equations

$$f_i(x) = 0 \quad (i = 1, \ldots, m), \tag{2}$$

lying in some subset V of $\mathbb{K}^m$. Throughout this review we are mainly concerned with the case $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$.

Key words and phrases. Random Polynomials; System of Random Equations; Bernstein Basis, Logarithmic Energy, Elliptic Fekete Points.

If we choose at random the coefficients $\{a_j^{(i)}\}$, then the solution of the system (2) becomes a random subset of $\mathbb{K}^m$. This is the main object of this review.

In the first part of this paper we focus on the real case. The main problem we consider is that of understanding $N^f(V)$: the number of solutions lying in the Borel subset V of $\mathbb{R}^m$.

In the second part we deal with a different problem: How are the roots of complex polynomials distributed?

This article is organized as follows:
In *Section 2* we start with some historical remarks on random polynomials. After that we move to the case of random systems of equations. We mention some recent results for centered Gaussian distributions. In *Section 2.1* we consider the non-centered case, which has also been called "smooth-analysis" in the last years. That is, we start with a fixed (non-random) polynomial system, then we perturb it with a polynomial noise, and we ask what can be said about the number of roots of the perturbed system. In *Section 2.2* we review a result which computes the expected number of roots of a random system of polynomial equations expressed in a different basis, namely, the Bernstein basis. Finally in *Section 3* we focus on the complex case. We discuss a recent result concerning the distribution of points in the sphere associated with roots of random complex polynomials.

This review follows the talk given by the author in the colloquium which was held the inauguration of the Franco-Uruguayan Institute of Mathematics, in Punta del Este, Uruguay, on December 2009.

2. The Number of Real Roots of Random Polynomials

The study of the expectation of the number of real roots of a random polynomial started in the thirties with the work of Block and Polya [7]. Further investigations were made by Littlewood and Offord [14]. However, the first sharp result is due to M. Kac (see Kac[11, 12]), who gives the asymptotic value

$$\mathbb{E}\left(N^f(\mathbb{R})\right) \approx \frac{2}{\pi}\log d, \quad \text{as} \quad d \to +\infty,$$

when the coefficients of the degree d univariate polynomial f are Gaussian centered independent random variables $N(0,1)$ (see the book by Bharucha–Reid and Sambandham [6]).

The first important result in the study of real roots of random system of polynomial equations is due to Shub and Smale [20] in 1992, where the authors computed the expectation of $N^f(\mathbb{R}^m)$ when the coefficients are Gaussian centered independent random variables having variances:

$$\mathbb{E}\left[(a_j^{(i)})^2\right] = \frac{d_i!}{j_1!\cdots j_m!\,(d_i - \|j\|)!}. \tag{3}$$

Their result was

$$\mathbb{E}\left(N^f(\mathbb{R}^m)\right) = \sqrt{d_1\cdots d_m}, \tag{4}$$

that is, the square root of the Bézout number associated to the system. The proof is based on a double fibration manipulation of the coarea formula. Some extensions of their work, including new results for one polynomial in one variable, can be found in Edelman–Kostlan[10]. There are also other extensions to multi-homogeneous systems in McLennan[16], and, partially, to sparse systems in Rojas[17] and Malajovich–Rojas[15]. A similar question for the number of critical points of real-valued polynomial random functions has been considered in Dedieu–Malajovich[9].

The probability law of the Shub–Smale model defined in (3) has the simplifying property of being invariant under the action of the orthogonal group in $\mathbb{R}^m$. In Kostlan[13] one can find the classification of all Gaussian probability distributions over the coefficients with this geometric invariant property.

In 2005, Azaïs and Wschebor gave a new and deep insight to this problem. The key point is using the Rice formula for random Gaussian fields (cf. Azaïs–Wschebor[5]). This formula allows one to extend the Shub–Smale result to other probability distributions over the coefficients. A general formula for $\mathbb{E}(N^f(V))$ when the random functions $f_i\,(i=1,\dots,m)$ are stochastically independent and their law is centered and invariant under the orthogonal group on $\mathbb{R}^m$ can be found in Azaïs–Wschebor[4]. This includes the Shub–Smale formula

(4) as a special case. Moreover, Rice formula appears to be the instrument to consider a major problem in the subject which is to find the asymptotic distribution of $N^f(V)$ (under some normalization). The only published results of which the author is aware concern asymptotic variances as $m \to +\infty$. (See Wschebor[25] for a detailed description in this direction and a simpler proof of Shub–Smale result).

2.1. **Non-centered Systems.** The aim of this section is to remove the hypothesis that the coefficients have zero expectation.

One way to look at this problem is to start with a non-random system of equations (the "signal")

$$P_i(x) = 0 \quad (i = 1, \dots, m), \tag{5}$$

perturb it with a polynomial noise $X_i(x)$ $(i = 1, \dots, m)$, that is, consider

$$P_i(x) + X_i(x) = 0 \quad (i = 1, \dots, m),$$

and ask what one can say about the number of roots of the new system, or, how much the noise modifies the number of roots of the deterministic part. (For short, we denote $N^f = N^f(\mathbb{R}^m)$).

Roughly speaking, we prove in *Theorem 1* that if the relation signal over noise is neither too big nor too small, in a sense that will be made precise later on, there exist positive constants C, θ, where $0 < \theta < 1$, such that

$$\mathbb{E}(N^{P+X}) \leq C\,\theta^m \mathbb{E}(N^X). \tag{6}$$

Inequality (6) becomes of interest if the starting non-random system (5) has a large number of roots, possibly infinite, and m is large. In this situation, the effect of adding polynomial noise is a reduction at a geometric rate of the expected number of roots, as compared to the centered case in which all the P_i's are identically zero.

For simplicity we assume that the polynomial noise X has the Shub-Smale distribution. However, one should keep in mind that the result can be extended to other orthogonally invariant distributions (cf. Armentano–Wschebor[2]).

Before the statement of *Theorem 1* below, we need to introduce some additional notations.

In this simplified situation, one only needs hypotheses concerning the relation between the signal P and the Shub-Smale noise X, which roughly speaking should neither be too small nor too big.

Since X has the Shub-Smale distribution, from (3) we get

$$\mathrm{Var}(X_i(x)) = (1 + \|x\|^2)^{d_i}, \quad \forall x \in \mathbb{R}^m, \qquad (i = 1, \dots, m).$$

Define

$$H(P_i) := \sup_{x \in \mathbb{R}^m} \left\{ (1 + \|x\|) \cdot \left\| \nabla \left(\frac{P_i}{(1 + \|x\|^2)^{d_i/2}} \right)(x) \right\| \right\},$$

$$K(P_i) := \sup_{x \in \mathbb{R}^m \setminus \{0\}} \left\{ (1 + \|x\|^2) \cdot \left| \frac{\partial}{\partial \rho} \left(\frac{P_i}{(1 + \|x\|^2)^{d_i/2}} \right)(x) \right| \right\},$$

for $i = 1, \dots, m$, where $\| \cdot \|$ is the Euclidean norm, and $\frac{\partial}{\partial \rho}$ denotes the derivative in the direction defined by $\frac{x}{\|x\|}$, at each point $x \neq 0$.

For $r > 0$, put:

$$L(P_i, r) := \inf_{\|x\| \geq r} \frac{P_i(x)^2}{(1 + \|x\|^2)^{d_i}} \quad (i = 1, \dots, m).$$

One can check by means of elementary computations that for each P as above, one has

$$H(P) < \infty, \ K(P) < \infty.$$

With these notations, we introduce the following hypotheses on the systems as m grows:

H_1)

$$A_m = \frac{1}{m} \cdot \sum_{i=1}^{m} \frac{H^2(P_i)}{i} = \mathrm{o}(1) \quad \text{as } m \to +\infty \tag{7a}$$

$$B_m = \frac{1}{m} \cdot \sum_{i=1}^{m} \frac{K^2(P_i)}{i} = \mathrm{o}(1) \quad \text{as } m \to +\infty. \tag{7b}$$

H_2) There exist positive constants r_0, ℓ such that if $r \geq r_0$:

$$L(P_i, r) \geq \ell \quad \text{for all } i = 1, \dots, m.$$

Theorem 1. *Under the hypotheses H_1) and H_2), one has*

$$\mathbb{E}(N^{P+X}) \leq C\,\theta^m \mathbb{E}(N^X), \tag{8}$$

where C, θ are positive constants, $0 < \theta < 1$.

2.1.1. *Remarks on the statement of Theorem 1.*

- It is obvious that our problem does not depend on the order in which the equations

$$P_i(x) + X_i(x) = 0 \quad (i = 1, \ldots, m)$$

  appear. However, conditions (7a) and (7b) in hypothesis H_3) do depend on the order. One can state them by saying that there exists an order $i = 1, \ldots, m$ on the equations, such that (7a) and (7b) hold true.
- Condition H_1) can be interpreted as a bound on the quotient signal over noise. In fact, it concerns the gradient of this quotient. In (7b) the radial derivative appears, which happens to decrease faster as $\|x\| \to \infty$ than the other components of the gradient.

  Clearly, if $H(P_i)$, $K(P_i)$ are bounded by fixed constants, (7a) and (7b) are verified. Also, some of them may grow as $m \to +\infty$ provided (7a) and (7b) remain satisfied.
- Hypothesis H_2) goes – in some sense – in the opposite direction: For large values of $\|x\|$ we need a lower bound of the relation signal over noise.
- A result of the type of *Theorem 1* can not be obtained without putting some restrictions on the relation signal over noise. In fact, consider the system

$$P_i(x) + \sigma\, X_i(x) = 0 \quad (i = 1, \ldots, m), \tag{9}$$

  where σ is a positive real parameter. If we let $\sigma \to +\infty$, the relation signal over noise tends to zero and the expected number of roots will tend to $\mathbb{E}(N^X)$. On the other hand, if $\sigma \downarrow 0$, $\mathbb{E}(N^X)$ can have different behaviours. For example, if P is a "regular" system, the expected value of the number of roots of (9) tends to the number of roots of $P_i(x) = 0$, $(i =$

$1, \dots, m$), which may be much bigger than $\mathbb{E}(N^X)$. In this case, the relation signal over noise tends to infinity.

- As it was mentioned before we can extend *Theorem 1* to other orthogonally invariant distributions. However, for the general version we need to add more hypotheses.

In the next paragraphs we are going to give two simple examples. For the proof of *Theorem 1* and more examples with different noises see Armentano–Wschebor[2].

2.1.2. *Some Examples.* We assume that the degrees d_i are uniformly bounded.

For the first example, let

$$P_i(x) = \|x\|^{d_i} - r^{d_i},$$

where d_i is even and r is positive and remains bounded as m varies. Then, one has:

$$\frac{\partial}{\partial \rho}\left(\frac{P_i}{(1+\|x\|^2)^{d_i/2}}\right)(x) = \frac{d_i\,\|x\|^{d_i-1} + d_i\, r^{d_i}\,\|x\|}{(1+\|x\|^2)^{\frac{d_i}{2}+1}} \leq \frac{d_i(1+r^{d_i})}{(1+\|x\|^2)^{3/2}}$$

$$\nabla\left(\frac{P_i}{(1+\|x\|^2)^{d_i/2}}\right)(x) = \frac{d_i\,\|x\|^{d_i-2} + d_i\, r^{d_i}}{(1+\|x\|^2)^{\frac{d_i}{2}+1}}\, x$$

which implies

$$\left\|\nabla\left(\frac{P_i}{(1+\|x\|^2)^{d_i/2}}\right)(x)\right\| \leq \frac{d_i(1+r^{d_i})}{(1+\|x\|^2)^{3/2}}.$$

Again, since the degrees $d_1, \dots, d_m$ are bounded by a constant that does not depend on m, H_1) follows. H_2) also holds under the same hypothesis.

Notice that an interest in this choice of the P_i's lies in the fact that obviously the system $P_i(x) = 0$ ($i = 1, \dots, m$) has an infinite number of roots (all points in the sphere of radius r centered at the origin are solutions), but the expected number of roots of the perturbed system is geometrically smaller than the Shub–Smale expectation, when m is large.

Our second example is the following: Let T be a polynomial of degree d in one variable that has d distinct real roots. Define:

$$P_i(x_1, \dots, x_m) = T(x_i) \quad (i = 1, \dots, m).$$

One can easily check that the system verifies our hypotheses, so that there exist C, θ positive constants, $0 < \theta < 1$ such that

$$\mathbb{E}(N^{P+X}) \leq C\, \theta^m d^{m/2},$$

where we have used the Shub–Smale formula when the degrees are all the same. On the other hand, it is clear that $N^P = d^m$ so that the diminishing effect of the noise on the number of roots can be observed. A number of variations of these examples for P can be constructed, but we will not pursue the subject here.

2.2. **Other Polynomial Basis.** Up to now all probability measures were introduced in a particular basis, namely, the monomial basis $\{x^j\}_{\|j\| \leq d}$. However, in many situations, polynomial systems are expressed in different basis, for example, orthogonal polynomials, harmonic polynomials,Bernstein polynomials, etc. So, it is a natural question to ask: *What can be said about $N^f(V)$ when the randomization is performed in a different basis?*

For the case of random orthogonal polynomials see Barucha-Reid and Sambandham[6], and Edelman–Kostlan[10] for random harmonic polynomials.

In this section following Armentano–Dedieu[3] we give an answer to the average number of real roots of a random system of equations expresed in the Bernstein basis. Let us be more precise:

The Bernstein basis is given by:

$$b_{d,k}(x) = \binom{d}{k} x^k (1-x)^{d-k},\ 0 \leq k \leq d,$$

in the case of univariate polynomials, and

$$b_{d,j}(x_1, \dots, x_m) = \binom{d}{j} x_1^{j_1} \dots x_m^{j_m} (1 - x_1 - \dots - x_m)^{d - \|j\|},\ \|j\| \leq d,$$

for polynomials in m variables, where $j = (j_1, \dots, j_m)$ is a multi-integer, and $\binom{d}{j}$ is the multinomial coefficient.

Let us consider the set of real polynomial systems in m variables,

$$f_i(x_1,\ldots,x_m) = \sum_{\|j\|\le d_i} a_j^{(i)} b_{d,j}(x_1,\ldots,x_m) \qquad (i=1,\ldots,m).$$

Take the coefficients $a_j^{(i)}$ to be independent Gaussian standard random variables.

Define

$$\tau : \mathbb{R}^m \to \mathbb{P}\left(\mathbb{R}^{m+1}\right)$$

by

$$\tau(x_1,\ldots,x_m) = [x_1,\ldots,x_m, 1-x_1-\ldots-x_m].$$

Here $\mathbb{P}(\mathbb{R}^{m+1})$ is the projective space associated with $\mathbb{R}^{m+1}$, $[y]$ is the class of the vector $y\in\mathbb{R}^{m+1}$, $y\neq 0$, for the equivalence relation defining this projective space. The (unique) orthogonally invariant probability measure in $\mathbb{P}(\mathbb{R}^{m+1})$ is denoted by λ_m.

With the above notation the following theorem holds:

Theorem 2. (1) *For any Borel set V in $\mathbb{R}^m$ we have*

$$\mathbb{E}\left(N^f(V)\right) = \lambda_m(\tau(V))\sqrt{d_1\ldots d_m}.$$

In particular

(2) $\mathbb{E}\left(N^f\right) = \sqrt{d_1\ldots d_m}$,

(3) $\mathbb{E}\left(N^f(\Delta^m)\right) = \sqrt{d_1\ldots d_m}/2^m$, *where*

$$\Delta^m = \{x\in\mathbb{R}^m \ :\ x_i\ge 0 \text{ and } x_1+\ldots+x_m\le 1\},$$

(4) *When $m=1$, for any interval $I=[\alpha,\beta]\subset\mathbb{R}$, one has*

$$\mathbb{E}\left(N^f(I)\right) = \frac{\sqrt{d}}{\pi}\left(\arctan(2\beta-1)-\arctan(2\alpha-1)\right).$$

The fourth assertion in *Theorem 2* is deduced from the first assertion but it also can be derived from Crofton's formula (see for example Edelman–Kostlan[10]).

For the proof of *Theorem 2* see Armentano–Dedieu[3]

3. Distribution of Complex Roots of Random Polynomials

In this part we will see that points in the sphere associated with roots of Shub–Smale complex analogue random polynomials via the stereographic projection, are surprisingly well-suited with respect to the minimal logarithmic energy on the sphere. That is, they provide a fairly good approximation to a classical minimization problem over the sphere, namely, the Elliptic Fekete points problem.

Next paragraphs follows closely Armentano–Beltrán–Shub[1], where one can find proofs and more detailed references.

Given $x_1, \ldots, x_N \in \mathbb{S}^2 = \{x \in \mathbb{R}^3 : \|x\| = 1\}$, let

$$V(x_1, \ldots, x_N) = \ln \prod_{1 \le i < j \le N} \frac{1}{\|x_i - x_j\|} = - \sum_{1 \le i < j \le N} \ln \|x_i - x_j\| \tag{10}$$

be the logarithmic energy of the N-tuple $x_1, \ldots, x_N$. Let

$$V_N = \min_{x_1, \ldots, x_N \in \mathbb{S}^2} V(x_1, \ldots, x_N)$$

denote the minimum of this function. N-tuples minimizing the quantity (10) are usually called Elliptic Fekete Points. The problem of finding (or even approximate) such optimal configurations is a classical problem (see White[23] for its origins).

During the last decades this problem has attracted much attention, and the number of papers concerning it has grown amazingly. The reader may see Kuijlaars-Saff[19] for a nice survey.

In the list of Smale's problems for the XXI Century [22], problem number 7 reads:

Can one find $x_1, \ldots, x_N \in \mathbb{S}^2$ such that

$$V(x_1, \ldots, x_N) - V_N \le c \ln N, \tag{11}$$

c a universal constant?

More precisely, Smale demands a real number algorithm in the sense of Blum–Cucker–Shub–Smale[8] that with input N returns a N-tuple $x_1, \ldots, x_N$ satisfying equation (11), and such that the running time is polynomial on N.

One of the main difficulties when dealing with this problem is that the value of V_N is not even known up to logarithmic precision. In Rakhmanov–Saff–Zhou[18] the authors proved that if one defines C_N by

$$V_N = -\frac{N^2}{4}\ln\left(\frac{4}{e}\right) - \frac{N\ln N}{4} + C_N N, \tag{12}$$

then,

$$-0.112768770... \le \liminf_{N\to\infty} C_N \le \limsup_{N\to\infty} C_N \le -0.0234973...$$

Let $X_1, \ldots, X_N$ be independent random variables with common uniform distribution over the sphere. One can easily show that the expected value of the function $V(X_1, \ldots, X_N)$ in this case is,

$$\mathbb{E}(V(X_1, \ldots, X_N)) = -\frac{N^2}{4}\ln\left(\frac{4}{e}\right) + \frac{N}{4}\ln\left(\frac{4}{e}\right). \tag{13}$$

Thus, this random choice of points in the sphere with independent uniform distribution already provides a reasonable approach to the minimal value V_N, accurate to the order of $O(N\ln N)$.

On one side, this probability distribution has an important property, namely, invariance under the action of the orthogonal group on the sphere. However, on the other hand this probability distribution lacks on correlation between points. More precisely, in order to obtain well-suited configurations one needs some kind of repelling property between points, and in this direction independence is not favorable. Hence, it is a natural question whether other handy orthogonally invariant probability distributions may yield better expected values. Here is where complex random polynomials comes into account.

Given $z \in \mathbb{C}$, let

$$\hat{z} := \frac{(z,1)}{1+|z|^2} \in \mathbb{C}\times\mathbb{R} \cong \mathbb{R}^3$$

be the associated points in the Riemann Sphere, i.e. the sphere of radius $1/2$ centered at $(0,0,1/2)$. Finally, let

$$X = 2\hat{z} - (0,0,1) \in \mathbb{S}^2$$

be the associated points in the unit sphere.

Given a polynomial f in one complex variable of degree N, we consider the mapping

$$f \mapsto V(X_1, \dots, X_N),$$

where X_i $(i = 1, \dots, N)$ are the associated roots of f in the unit sphere. Notice that this map is well defined in the sense that it does not depend on the way we choose to order the roots.

Theorem 3. *Let $f(z) = \sum_{k=0}^{N} a_k z^k$ be a complex random polynomial, such that the coefficients a_k are independent complex random variables, such that the real and imaginary parts of a_k are independent (real) Gaussian random variables centered at 0 with variance $\binom{N}{k}$. Then, with the notations above,*

$$\mathbb{E}\left(V(X_1, \dots, X_N)\right) = -\frac{N^2}{4} \ln\left(\frac{4}{e}\right) - \frac{N \ln N}{4} + \frac{N}{4} \ln \frac{4}{e}.$$

Comparing *Theorem 3* with equations (12) and (13), we see that the value of V is surpringsingly small at points coming from the solution set of this random polynomials. More precisely, necessarily many random realizations of the coefficients will produce values of V below the average and very close to V_N, possibly close enough to satisfy equation (11).

Notice that, taking the homogeneous counterpart of f, *Theorem 3* can be restated for random homogeneous polynomials and considering its complex projective solutions, under the identification of $\mathbb{P}(\mathbb{C}^2)$ with the Riemann sphere. In this fashion, the induced probability distribution over the space of homogeneous polynomials in two complex variables corresponds to the classical unitarily invariant Hermitian structure of the respective space (see Blum–Cucker–Shub–Smale[8]). Therefore, the probability distribution of the roots in $\mathbb{P}(\mathbb{C}^2)$ is invariant under the action of the unitary group.

It is not difficult to prove that the unitary group action over $\mathbb{P}(\mathbb{C}^2)$ correspond to the special orthogonal group of the unit sphere. Hence, the distribution of the associated random roots on the sphere is orthogonally invariant. Thus, *Theorem 3* is another geometric confirmation of the repelling property of the roots of this Gaussian random polynomials.

For a proof of *Theorem 3* and a more detailed discussion on this account see Armentano–Beltrán–Shub[1]. See also Shub–Smale[21].

References

[1] Armentano D., Beltrán C. and Shub M., *Minimizing the discrete logarithmic energy on the sphere: The role of random polynomials.* Transaction of the American Mathematical Society, 363 (2011), 2955-2965.

[2] Armentano D. and M. Wschebor, *Random systems of polynomial equations. The expected number of roots under smooth analysis.* Bernoulli 15, No. 1, (2009), 249-266.

[3] Armentano D. and Dedieu J.P., *On the Average Number of real roots of a Bernstein polynomial system.* Journal of Complexity, v.25 n.4, p.339-342, 2009.

[4] Azaïs J.-M and Wschebor M., On the Roots of a Random System of Equations. The Theorem of Shub and Smale and Some Extensions, *Found. Comput. Math.*, (2005), 125-144.

[5] Azaïs J.-M and Wschebor M., Level sets and extrema of random processes and fields. John Wiley and Sons 2009.

[6] Bharucha-Reid A. T. and Sambandham M., *Random Polynomials*, Probability and Mathematical Statistics, Academic Press, Orlando, FL, 1986.

[7] Bloch A. and Polya G.,*On the number of real roots of a random algebraic equation.* Proc. Cambridge Philos. Soc., 33:102114, 1932.

[8] Blum L., Cucker F., Shub M. and Smale S., *Complexity and real computation*, Springer-Verlag, New York, 1998.

[9] Dedieu J.-P. and Malajovich G., *On the number of minima of a random polynomial.* Journal of Complexity 24 (2008) 89-108.

[10] Edelman A. and Kostlan E, How many zeros of a random polynomial are real?,*Bull. Amer. Math. Soc.* (N.S.)**32**,(1), (1995), 1-37.

[11] Kac M., *On the average number of real roots of a random algebraic equation.* Bull. Am. Math. Soc. 49 (1943) 314-320 and 938.

[12] Kac M., *On the average number of real roots of a random algebraic equation (II).* Proc. London Math. Soc. 50 (1949) 390-408.

[13] Kostlan E., *On the expected number of real roots of a system of radom polynomial equations.* In: Foundations of Computational Mathematics, Hong Kong 2002, 149-188. World Sci. Pub., 2002.

[14] Littlewood, J. E. and Offord, A. C., *On the number of real roots of a random algebraic equation.* J. London Math. Soc., 13:288295, 1938.

[15] Malajovich G. and Rojas J.M, *High probability analysis of the condition number of sparse polynomial systems*, in *Theoret. Comput. Sci.*, 315(2-3), pp. 524-555, (2004).

[16] McLennan A., *The expected number of real roots of a multihomogeneous system of polynomial equations*, in *Amer. J. Math.*, 124(1), pp. 49-73, (2002).

[17] Rojas J.M., *On the average number of real roots of certain random sparse polynomial systems*, in *The mathematics of numerical analysis*, (Park City, UT, 1995) vol. 32, Lectures in Appl. Math., pp. 689-699, Amer. Math. Soc., Providence, RI, (1996).

[18] Rakhmanov E.A., Saff E.B. and Zhou Y.M. , *Minimal discrete energy on the sphere*, Math. Res. Letters **1** (1994), 647–662.

[19] Saff, E.B. and Kuijlaars, A.B.J., *Distributing many points on a sphere.* Math. Intelligencer 19 (1997), no. 1, 5–11.

[20] Shub M. and Smale S., *Complexity of Bezout's theorem. II. Volumes and probabilities*, Computational algebraic geometry (Nice, 1992), Progr. Math., vol. 109, Birkhäuser Boston, Boston, MA, 1993, pp. 267–285.

[21] ______, *Complexity of Bezout's theorem. III. Condition number and packing*, J. Complexity **9** (1993), no. 1, 4–14, Festschrift for Joseph F. Traub, Part I.

[22] Smale S., *Mathematical problems for the next century*, Mathematics: frontiers and perspectives, Amer. Math. Soc., Providence, RI, 2000, pp. 271–294.

[23] Whyte L.L., *Unique arrangements of points on a sphere*, Amer. Math. Monthly **59** (1952), 606–611.

[24] Wschebor M., On the Kostlan-Shub-Smale model for random polynomial systems. Variance of the number of roots, in *J. of Complexity*, 21, pp. 773-789, (2005).

[25] Wschebor M., *Systems of random equations. A review of some recent results.* In and out of equilibrium. 2, 559–574, Progr. Probab., 60, Birkhuser, Basel, 2008.

Centro de Matemática, Universidad de la República. Montevideo, Uruguay

RICE FORMULAS AND GAUSSIAN WAVES II.

JEAN-MARC AZAÏS, JOSÉ R. LEÓN, AND MARIO WSCHEBOR

Abstract. We prove a certain number of results on specular points and dislocations of random waves which we have announced without proof in [1] or for which only an outline of proof has been given in this reference. Along the paper, waves are Gaussian and the basic tools are Rice formulas.
The main results are on the first two moments and in some special case, also weak convergence is obtained.

1. Introduction

This paper is a continuation of [1] in which we study the zeroes of certain random waves that appear in oceanography and optics. Our aim here is to give full proofs of certain results that were stated without proof in that paper, or for which proofs have been only sketched.

Our interest lies in the geometry of the set of zeros of random fields with low-dimensional parameter set (smaller or equal than 3). In general, only a restricted number of geometrical characteristics of these sets can be described with the methods we use, namely the so-called Rice formulas.

When the set of zeros is 0-dimensional, Rice formulas permit to express the moments of the number of zeros by means of certain integrals depending upon a description of the probability law of the random fields. If it is 1-dimensional one can do something similar

2000 *Mathematics Subject Classification.* Primary 60G15; Secondary 60G60 78A10 78A97 86A05.

Key words and phrases. Rice formula, Gaussian waves, specular points, dislocations of wavefronts.

with length instead of number of zeros, if it is 2-dimensional with area-measure, and so on. One can also extend the same methods to weighted zeros, that is, compute the moments of total weight in the 0-dimensional case and the integral of a weight function on the 0-level set of the random field in the other cases.

We compute moments that are useful to make statistics on certain parameters appearing in the law of the random field. In some situations we can go further and obtain weak limit theorems for certain re-normalizations of natural functionals of the paths which are of interest.

These are special cases of the general problem of computing moments of the geometric measure of the level sets of random fields. For this purpose, Rice formulas have been developed since the pioneering work of Rice [10]. We refer to the book by Azaïs and Wschebor [2] for an extended presentation of the subject and for proofs of the general formulas we use.

In this paper we will consider two classes of 0-sets of random fields: specular points (Section 2) and dislocations of wave fronts (Section 3). For details not mentioned here and other geometrical properties of waves which can be studied with analogous methods, we refer to [1], [4], [9].

All random fields are assumed to have continuously differentiable paths and to be Gaussian, a hypothesis that is useful to be able to perform the computations associated with Rice formulas, but can fail to approximate physical reality in certain cases.

We use the following notations: $\sigma_d(B)$ the d-dimensional Hausdorff measure of a Borel set B. If f is a function of d variables we denote f_i the partial derivative with respect to the i-th variable. M^T denotes the transpose of a matrix M. $(const)$ is a positive constant whose value may change from one occurrence to another. $p_\xi(x)$ is the density of the random variable or vector ξ, whenever it exists. λ_k ($k = 0, 1, 2, ...$) denotes the k-th spectral moment of a stationary random process defined on the real line.

2. Specular points

2.1. **Specular points for one-parameter processes.** Specular points of a curve are defined as follows: We take cartesian coordinates Oxz in the plane and assume the curve is the graph of a C^1-function $z = W(x)$. A light source placed at $(0, h_1)$ emits a ray that is reflected at the point $(x, W(x))$ of the curve and the reflected ray is registered by an observer placed at $(0, h_2)$.

Using the equality between the angles of incidence and reflection with respect to the normal vector to the curve - i.e. $N(x) = (-W'(x), 1)$ - an elementary computation gives:

$$W'(x) = \frac{\alpha_2 r_1 - \alpha_1 r_2}{x(r_2 - r_1)}, \tag{1}$$

where $\alpha_i := h_i - W(x)$ and $r_i := \sqrt{x^2 + \alpha_i^2}$, i=1,2.

The points $(x, W(x))$ of the curve such that x is a solution of (1) are called "specular points". When the curve is random, one of our aims is to study the probability distribution of the number of specular points such the abscise $x \in A$, where A is a Borel subset of the line.

The following approximation is due to M.S. Longuet-Higgins (see [7], [8]): Suppose that h_1 and h_2 are big with respect to $W(x)$ and x, then $r_i = \alpha_i + x^2/(2\alpha_i) + O(h_i^{-3})$. Then, (1) can be approximated by

$$W'(x) \simeq \frac{x}{2}\frac{\alpha_1 + \alpha_2}{\alpha_1\alpha_2} \simeq \frac{x}{2}\frac{h_1 + h_2}{h_1 h_2} = kx, \tag{2}$$

where

$$k := \frac{1}{2}\Big(\frac{1}{h_1} + \frac{1}{h_2}\Big).$$

Set $Y(x) := W'(x) - kx$ and $SP(A)$ the number of roots of $Y(x)$ belonging to the set A. We will call $SP(A)$ the "Longuet-Higgins approximation" of the number of specular points, when the parameter

k tends to 0.

Assume now that $\{W(x) : x \in \mathbb{R}\}$ is a centered Gaussian stationary process with C^2-paths. In [1] an exact formula has been obtained for the expectation of the number of specular points belonging to the interval $[a, b]$. This is an integral formula, well-adapted to numerical computation and it turns out that $\mathbb{E}(SP([a, b]))$ is a very accurate approximation, for example, for ocean waves.

Also, the Longuet-Higgins approximation is tractable from a mathematical point of view, and one can go much farther than expectation in the description of the law of the number of specular points. More precisely in [1] it is proved that:

(1) Adding some hypotheses on the law of the process $\{W(x) : x \in \mathbb{R}\}$ (paths of class C^4 and some mixing condition, such as δ-dependence or a controlled decay of correlation), it follows that

$$\operatorname{Var}\big[SP(\mathbb{R})\big] = \theta\frac{1}{k} + O(1) \text{ as } k \to 0,$$

where θ is a constant that can be computed by means of an explicit formula from the covariance of the given Gaussian process, which is well-adapted to numerical computation.

This implies that the coefficient of variation of the random variable $SP(\mathbb{R})$ tends to zero in a controlled manner, namely:

$$\frac{\sqrt{\operatorname{Var}(SP(\mathbb{R}))}}{\mathbb{E}(SP(\mathbb{R}))} \sim \sqrt{\frac{\theta\pi k}{2\lambda_4}} \text{ as } k \to 0, \tag{3}$$

since

$$\mathbb{E}(SP(\mathbb{R})) \sim \sqrt{\frac{2\lambda_4}{\pi}}\frac{1}{k},$$

((3) corrects a small error in [1]).

(2) With some additional requirement on the smoothness of the paths of the process, under the same asymptotic, the natural renormalization of $SP(\mathbb{R})$ tends to the standard normal

distribution $\Phi(x)$, that is, for every $x \in \mathbb{R}$:

$$\mathbb{P}\Big(\frac{SP(\mathbb{R}) - (2\lambda_4/\pi)^{1/2}/k}{(\theta/k)^{1/2}} \leq x\Big) \to \Phi(x) \text{ as } k \to 0.$$

2.2. **Specular points for two-parameter processes.** Let us consider in $\mathbb{R}^3$ a coordinate system Oxyz, and a C^1-function $z = W(x, y)$. The following definition of specular points of the graph extends naturally the one we gave above for functions of one real variable.

The source of light is placed at the point $(0, 0, h_1)$ and the observer at $(0, 0, h_2)$. The point (x, y) is said to be a specular point if the normal vector $n(x, y) = (-W_x, -W_y, 1)$ to the graph at $(x, y, W(x, y))$ satisfies the following two conditions:

- the angles with the incident ray $I = (-x, -y, h_1 - W)$ and the reflected ray $R = (-x, -y, h_2 - W)$ are equal (for short the argument (x, y) has been removed),
- it belongs to the plane generated by I and R.

Setting $\alpha_i = h_i - W$ and $r_i = \sqrt{x^2 + y^2 + \alpha_i}$, $i = 1, 2$, as in the one-parameter case we have:

$$\begin{aligned} W_x &= \frac{x}{x^2+y^2}\frac{\alpha_2 r_1 - \alpha_1 r_2}{r_2 - r_1}, \\ W_y &= \frac{y}{x^2+y^2}\frac{\alpha_2 r_1 - \alpha_1 r_2}{r_2 - r_1}. \end{aligned} \tag{4}$$

When h_1 and h_2 are large, the system above can be approximated by

$$\begin{aligned} W_x &= kx \\ W_y &= ky, \end{aligned} \tag{5}$$

under the same conditions as in dimension 1. This is the Longuet-Higgins approximation for two-parameter functions.

For each subset Q of $\mathbb{R}^2$, we denote by $SP(Q)$, the number of approximate specular points in the sense of (5) such that $(x, y) \in Q$. In the remaining of this paragraph we limit our attention to this approximation and to the case in which $\{W(x, y) : (x, y) \in \mathbb{R}^2\}$ is a centered Gaussian stationary random field with C^3-paths.

We need some additional notation: μ denotes the spectral measure of the random field, which is a Borel measure on $\mathbb{R}^2$ and $\lambda_{ij}, i, j = 0, 1, 2, ...$ the spectral moments

$$\lambda_{ij} = \int_{\mathbb{R}^2} u^i v^j \mu(du, dv),$$

whenever they are well-defined.

In [1] one can find the statement of certain results on the behavior of expectation and variance of $SP(Q)$ under the asymptotic $k \to 0$. We give full proofs of these results below. For the time being, what is known for variance and coefficient of variation is weaker than in the one-dimensional parameter case.

Let us define:

$$\mathbf{Y}(x, y) := \begin{pmatrix} W_x(x, y) - kx \\ W_y(x, y) - ky \end{pmatrix}. \tag{6}$$

Under the non-degeneracy condition $\lambda_{20}\lambda_{02} - \lambda_{11}^2 \neq 0$, the random field $\{Y(x, y) : x, y \in \mathbb{R}\}$ satisfies the hypotheses of Theorem 6.2. in [2], and we can write the Rice formula:

$$\begin{aligned} \mathbb{E}\big(SP(Q)\big) &= \int_Q \mathbb{E}\big(|\det \mathbf{Y}'(x, y)| \big| \mathbf{Y}(x, y) = 0\big) p_{\mathbf{Y}(x,y)}(\mathbf{0})\ dxdy \\ &= \int_Q \mathbb{E}\big(|\det \mathbf{Y}'(x, y)|\big) p_{\mathbf{Y}(x,y)}(\mathbf{0})\ dxdy, \end{aligned} \tag{7}$$

since for fixed (x, y) the random matrix $\mathbf{Y}'(x, y)$ and the random vector $\mathbf{Y}(x, y)$ are independent, so that the condition in the conditional expectation can be removed.

The density in the right hand side of (7) has the expression

(8)
$$p_{\mathbf{Y}(x,y)}(\mathbf{0}) = p_{(W_x,W_y)}(kx,ky)$$
$$= \tfrac{1}{2\pi}\tfrac{1}{\sqrt{\lambda_{20}\lambda_{02}-\lambda_{11}^2}} \exp\Big[-\tfrac{k^2}{2(\lambda_{20}\lambda_{02}-\lambda_{11}^2)}\big(\lambda_{02}x^2 - 2\lambda_{11}xy + \lambda_{20}y^2\big)\Big].$$

To compute the expectation of the absolute value of the determinant in the right hand side of (7), which does not depend on x,y, we use the method of [3] (see also [6]). Set $\Delta := \det \mathbf{Y}'(x,y) = (W_{xx}-k)(W_{yy}-k) - W_{xy}^2$.

We have

$$\mathbb{E}(|\Delta|) = \mathbb{E}\left[\frac{2}{\pi}\int_0^{+\infty} \frac{1-\cos(\Delta t)}{t^2}dt\right]. \tag{9}$$

Define

$$h(t) := \mathbb{E}\left[\exp\big(it[(W_{xx}-k)(W_{yy}-k) - W_{xy}^2]\big)\right].$$

Then

$$\mathbb{E}(|\Delta|) = \frac{2}{\pi}\Big(\int_0^{+\infty} \frac{1-\mathfrak{Re}[h(t)]}{t^2}dt\Big). \tag{10}$$

To compute $h(t)$ we define

$$A = \begin{pmatrix} 0 & 1/2 & 0 \\ 1/2 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

and Σ the variance matrix of $W_{xx}, W_{yy}, W_{x,y}$

$$\Sigma := \begin{pmatrix} \lambda_{40} & \lambda_{22} & \lambda_{31} \\ \lambda_{22} & \lambda_{04} & \lambda_{13} \\ \lambda_{31} & \lambda_{13} & \lambda_{22} \end{pmatrix}.$$

Let$\Sigma^{1/2}A\Sigma^{1/2} = P\ diag(\Delta_1, \Delta_2, \Delta_3)P^T$ where P is orthogonal. Then by a diagonalization argument

$$
\begin{aligned}
(11)\quad h(t) &= e^{itk^2} \\
&\mathbb{E}\Big(\exp\big[it\big((\Delta_1 Z_1^2 - k(s_{11}+s_{21})Z_1) + (\Delta_2 Z_2^2 - k(s_{12}+s_{22})Z_2) \\
&\qquad + (\Delta_3 Z_3^2 - k(s_{13}+s_{23})Z_3)\big)\big]\Big),
\end{aligned}
$$

where (Z_1, Z_2, Z_3) is standard normal and s_{ij} are the entries of $\Sigma^{1/2}P^T$.

One can check that if ξ is a standard normal variable and τ, μ are real constants, $\tau > 0$:

$$
\begin{aligned}
\mathbb{E}\big(e^{i\tau(\xi+\mu)^2}\big) &= (1-2i\tau)^{-1/2} e^{\frac{i\tau\mu^2}{(1-2i\tau)}} \\
&= \frac{1}{(1+4\tau^2)^{1/4}} \exp\Big[\frac{-2\tau}{1+4\tau^2} + i\big(\varphi + \frac{\tau\mu^2}{1+4\tau^2}\big)\Big],
\end{aligned}
$$

where

$$\varphi = \frac{1}{2}\arctan(2\tau),\ 0 < \varphi < \pi/4.$$

Replacing in (11), we obtain for $\mathfrak{Re}[h(t)]$ the formula:

$$
(12)\quad \mathfrak{Re}[h(t)] = \Big[\prod_{j=1}^{3} \frac{d_j(t,k)}{\sqrt{1+4\Delta_j^2 t^2}}\Big] \cos\Big(\sum_{j=1}^{3}\big(\varphi_j(t) + k^2 t\psi_j(t)\big)\Big),
$$

where, for $j = 1, 2, 3$:

- $d_j(t,k) = \exp\Big[-\dfrac{k^2t^2}{2}\ \dfrac{(s_{1j}+s_{2j})^2}{1+4\Delta_j^2t^2}\Big]$,
- $\varphi_j(t) = \dfrac{1}{2}\arctan(2\Delta_j t),\ 0 < \varphi_j < \pi/4$,
- $\psi_j(t) = \dfrac{1}{3} - t^2\dfrac{(s_{1j}+s_{2j})^2\Delta_j}{1+4\Delta_j^2t^2}$.

Introducing these expressions in (10) and using (8) we obtain a new formula which has the form of a rather complicated integral. However, it is well adapted to numerical evaluation.

On the other hand, this formula allows us to compute the equivalent as $k \to 0$ of the expectation of the total number of specular points under the Longuet-Higgins approximation. In fact, a first order expansion of the terms in the integrand gives a somewhat more accurate result, that we state as a theorem:

Theorem 1.

$$\mathbb{E}\big(SP(\mathbb{R}^2)\big) = \frac{m_2}{k^2} + O(1), \tag{13}$$

where

$$m_2 = \int_0^{+\infty} \frac{1 - \big[\prod_{j=1}^3 (1 + 4\Delta_j^2 t^2)\big]^{-1/2} \cos\big(\sum_{j=1}^3 \varphi_j(t)\big)}{t^2} dt \tag{14}$$

$$= \int_0^{+\infty} \frac{1 - 2^{-3/2}\big[\prod_{j=1}^3 \big(A_j\sqrt{1 + A_j}\big)\big]\big(1 - B_1B_2 - B_2B_3 - B_3B_1\big)}{t^2} dt,$$

where

$$A_j = A_j(t) = \big(1 + 4\Delta_j^2 t^2\big)^{-1/2}, \ B_j = B_j(t) = \sqrt{(1 - A_j)/(1 + A_j)}.$$

Notice that m_2 only depends on the eigenvalues $\Delta_1, \Delta_2, \Delta_3$ and is easily computed numerically.

We now consider the variance of the total number of specular points in two dimensions, looking for analogous results to the one-dimensional case, in view of their interest for statistical applications. It turns out that the computations become much more involved. The statements on variance and speed of convergence to zero of the coefficient of variation that we give below include only the order of the asymptotic behavior in the Longuet-Higgins approximation, but not the constant. However, we still consider them to be useful. If one refines the computations one can give rough bounds on the generic constants in Theorem 2 and Corollary 1 on the basis of additional hypotheses on the random field.

Now we add the following hypothesis to the set already required to study the expectation of the specular points under the Longuet-Higgins asymptotic. We express $W''(\mathbf{0})$ in the reference system xOy of $\mathbb{R}^2$ as the 2×2 symmetric centered Gaussian random matrix:

$$W''(\mathbf{0}) = \begin{pmatrix} W_{xx}(\mathbf{0}) & W_{xy}(\mathbf{0}) \\ W_{xy}(\mathbf{0}) & W_{yy}(\mathbf{0}) \end{pmatrix}.$$

The function

$$\mathbf{z} \rightsquigarrow \Delta(\mathbf{z}) = \det\big[\operatorname{Var}\big(W''(0)\mathbf{z}\big)\big],$$

defined on $\mathbf{z} = (z_1, z_2)^T \in \mathbb{R}^2$, is a non-negative homogeneous polynomial of degree 4 in the pair z_1, z_2. We will assume the non-degeneracy condition:

$$\min\{\Delta(\mathbf{z}) : \|\mathbf{z}\| = 1\} = \underline{\Delta} > 0. \tag{15}$$

Theorem 2. *Let us assume that $\{W(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^2\}$ satisfies the above conditions and that it is also δ-dependent, $\delta > 0$, that is, $\mathbb{E}\big(W(\mathbf{x})W(\mathbf{y})\big) = 0$ whenever $\|\mathbf{x} - \mathbf{y}\| > \delta$.*

Then, for k small enough:

$$\operatorname{Var}\big(SP(\mathbb{R}^2)\big) \;\leq\; \frac{L}{k^2},$$

where L is a positive constant depending upon the law of the random field.

A direct consequence of Theorems 1 and 2 is the following:

Corollary 1. *Under the same hypotheses of Theorem 2, for k small enough, one has:*

$$\frac{\sqrt{\operatorname{Var}\big(SP(\mathbb{R}^2)\big)}}{\mathbb{E}\big(SP(\mathbb{R}^2)\big)} \;\leq\; L_1 k,$$

where L_1 is a new positive constant.

Proof of Theorem 2. Let us denote $T = SP(\mathbb{R}^2)$. We have:

$$\operatorname{Var}(T) = \mathbb{E}(T(T-1)) + \mathbb{E}(T) - [\mathbb{E}(T)]^2. \tag{16}$$

We have already computed the equivalents as $k \to 0$ of the second and third term in the right-hand side of (16). Our task in what follows is to consider the first term.

The proof is performed using Rice formula for the second factorial moment of the number of roots of the random field Y. We apply Theorem 6.3. of [2] for dimension $d = 2$ and $k = 2$. Then,

$$\begin{aligned}
&\mathbb{E}(T(T-1)) = \\
&= \int\int_{\mathbb{R}^2 \times \mathbb{R}^2} \mathbb{E}\Big(|\det \mathbf{Y}'(\mathbf{x})||\det \mathbf{Y}'(\mathbf{y})| \,\Big|\, \mathbf{Y}(\mathbf{x}) = \mathbf{0}, \mathbf{Y}(\mathbf{y}) = \mathbf{0}\Big) \\
&\hspace{20em} p_{\mathbf{Y}(\mathbf{x}),\mathbf{Y}(\mathbf{y})}(\mathbf{0},\mathbf{0})\; d\mathbf{x}d\mathbf{y} \\
&= \int\int_{\|\mathbf{x}-\mathbf{y}\|>\delta} \dots\, d\mathbf{x}d\mathbf{y} + \int\int_{\|\mathbf{x}-\mathbf{y}\|\le\delta} \dots\, d\mathbf{x}d\mathbf{y} = J_1 + J_2.
\end{aligned}$$

For J_1 we proceed as in the proof of Theorem 1 of [1], using the δ-dependence and the evaluations therein. We obtain:

$$J_1 = \frac{m_2^2}{k^4} + \frac{O(1)}{k^2}. \tag{17}$$

Let us show that for small k,

$$J_2 = \frac{O(1)}{k^2}. \tag{18}$$

In view of (16), (13) and (17) this suffices to prove the theorem.

We do not perform all detailed computations. The key point consists in evaluating the behavior of the integrand that appears in J_2 near the diagonal $\mathbf{x} = \mathbf{y}$, where the density $p_{\mathbf{Y}(\mathbf{x}),\mathbf{Y}(\mathbf{y})}(\mathbf{0},\mathbf{0})$ degenerates and the conditional expectation tends to zero.

For the density, using the invariance under translations of the law of $W'(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^2$, we have:

$$\begin{aligned}
p_{\mathbf{Y}(\mathbf{x}),\mathbf{Y}(\mathbf{y})}(\mathbf{0},\mathbf{0}) &= p_{W'(\mathbf{x}),W'(\mathbf{y})}(k\mathbf{x}, k\mathbf{y}) \\
&= p_{W'(\mathbf{0}),W'(\mathbf{y}-\mathbf{x})}(k\mathbf{x}, k\mathbf{y}) \\
&= p_{W'(\mathbf{0}),[W'(\mathbf{y}-\mathbf{x})-W'(\mathbf{0})]}(k\mathbf{x}, k(\mathbf{y}-\mathbf{x})).
\end{aligned}$$

Perform the Taylor expansion, for small $\mathbf{z} = \mathbf{y} - \mathbf{x} \in \mathbb{R}^2$:

$$W'(\mathbf{z}) = W'(\mathbf{0}) + W''(\mathbf{0})\mathbf{z} + O(\|\mathbf{z}\|^2).$$

Using the non-degeneracy assumption (15) and the fact that $W'(\mathbf{0})$ and $W''(\mathbf{0})$ are independent, we can show that for $\mathbf{x}, \mathbf{z} \in \mathbb{R}^2$, $\|\mathbf{z}\| \leq \delta$:

$$p_{\mathbf{Y}(\mathbf{x}),\mathbf{Y}(\mathbf{y})}(\mathbf{0}, \mathbf{0}) \leq \frac{C_1}{\|\mathbf{z}\|^2} \exp\big[- C_2 k^2 (\|\mathbf{x}\| - C_3)^2 \big],$$

where C_1, C_2, C_3 are positive constants.

Let us consider the conditional expectation. For each pair $\mathbf{x}, \mathbf{y}$ of different points in $\mathbb{R}^2$, denote by τ the unit vector $(\mathbf{y} - \mathbf{x})/\|\mathbf{y} - \mathbf{x}\|$ and $\mathbf{n}$ a unit vector orthogonal to τ. We denote respectively by $\partial_\tau \mathbf{Y}, \partial_{\tau\tau}\mathbf{Y}, \partial_{\mathbf{n}}\mathbf{Y}$ the first and second partial derivatives of the random field in the directions given by τ and $\mathbf{n}$.

Under the condition

$$\mathbf{Y}(\mathbf{x}) = \mathbf{0}, \mathbf{Y}(\mathbf{y}) = \mathbf{0},$$

we have the following simple bound on the determinant, based upon its definition and Rolle's Theorem applied to the segment $[\mathbf{x}, \mathbf{y}] = \{\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}\}$:

$$\begin{aligned} \big|\det \mathbf{Y}'(\mathbf{x})\big| &\leq \|\partial_\tau \mathbf{Y}(\mathbf{x})\| \|\partial_{\mathbf{n}} \mathbf{Y}(\mathbf{x})\| \\ &\leq \|\mathbf{y} - \mathbf{x}\| \sup_{\mathbf{s} \in [\mathbf{x}, \mathbf{y}]} \|\partial_{\tau\tau} \mathbf{Y}(\mathbf{s})\| \|\partial_{\mathbf{n}} \mathbf{Y}(\mathbf{x})\| \end{aligned} \tag{19}$$

So,

$$\mathbb{E}\Big(|\det \mathbf{Y}'(\mathbf{x})||\det \mathbf{Y}'(\mathbf{y})|\Big|\ \mathbf{Y}(\mathbf{x})=\mathbf{0},\mathbf{Y}(\mathbf{y})=\mathbf{0}\Big)$$
$$\leq \|\mathbf{y}-\mathbf{x}\|^2\mathbb{E}\Big[\sup_{\mathbf{s}\in[\mathbf{x},\mathbf{y}]}\|\partial_{\tau\tau}\mathbf{Y}(\mathbf{s})\|^2\|\partial_{\mathbf{n}}\mathbf{Y}(\mathbf{x})\|.\|\partial_{\mathbf{n}}\mathbf{Y}(\mathbf{y})\| \Big| W'(\mathbf{x})=k\mathbf{x},W'(\mathbf{y})=k\mathbf{y}\Big]$$
$$=\|\mathbf{z}\|^2\mathbb{E}\Big[\sup_{\mathbf{s}\in[\mathbf{0},\mathbf{z}]}\|\partial_{\tau\tau}\mathbf{Y}(\mathbf{s})\|^2\|\partial_{\mathbf{n}}\mathbf{Y}(\mathbf{0})\|.\|\partial_{\mathbf{n}}\mathbf{Y}(\mathbf{z})\| \Big| W'(\mathbf{0})=k\mathbf{x},\frac{W'(\mathbf{z})-W'(\mathbf{0})}{\|\mathbf{z}\|}=k\tau\Big],$$

where the last equality is again a consequence of the stationarity of the random field $\{W(\mathbf{x}):\mathbf{x}\in\mathbb{R}^2\}$.

At this point, we perform a Gaussian regression on the condition. For the condition, use again Taylor expansion, the non-degeneracy hypothesis and the independence of $W'(\mathbf{0})$ and $W''(\mathbf{0})$. Then, use the finiteness of the moments of the supremum of bounded Gaussian processes (see for example [2], Ch. 2), take into account that $\|z\|\leq\delta$ to get the inequality:
(20)

$$\mathbb{E}\Big(|\det \mathbf{Y}'(\mathbf{x})||\det \mathbf{Y}'(\mathbf{y})|\Big|\ \mathbf{Y}(\mathbf{x})=\mathbf{0},\mathbf{Y}(\mathbf{y})=\mathbf{0}\Big)\leq C_4\,\|\mathbf{z}\|^2\big(1+k\|\mathbf{x}\|\big)^4,$$

where C_4 is a positive constant. Summing up, we have the following bound for J_2:

(21)
$$\begin{aligned} J_2 &\leq C_1C_4\ \pi\delta^2\int_{\mathbb{R}^2}\big(1+k\|\mathbf{x}\|\big)^4\exp\big[-C_2k^2(\|\mathbf{x}\|-C_3)^2\big]d\mathbf{x}\\ &= C_1C_4\ 2\pi^2\delta^2\int_0^{+\infty}\big(1+k\rho\big)^4\exp\big[-C_2k^2(\rho-C_3)^2\big]\rho d\rho. \end{aligned}$$

Performing the change of variables $w=k\rho$, (18) follows. ∎

3. Dislocation of wave fronts

Dislocations are phase singularities of wavefronts. They correspond to lines of darkness in light propagation, or threads of silence in sound (see Berry and Dennis [3]). In a mathematical framework they can be defined as the loci of points where the amplitude of waves vanishes.

We represent the wave as

$$W(\mathbf{x},t) = \xi(\mathbf{x},t) + i\eta(\mathbf{x},t), \quad \text{where } \mathbf{x} \in \mathbb{R}^d.$$

The dislocations are the intersection of the two level sets $\xi(\mathbf{x},t) = 0$, $\eta(\mathbf{x},t) = 0$ of the two random surfaces ξ and η. We consider a fixed time, for instance $t = 0$.

For random waves, when $d = \mathbf{2}$ we will study the expectation of the random variable

$$\#\{\mathbf{x} \in S : \xi(\mathbf{x},0) = \eta(\mathbf{x},0) = 0\}.$$

When $d = \mathbf{3}$ one important quantity is the length of the level curve

$$\mathcal{L}\{\mathbf{x} \in S : \xi(\mathbf{x},0) = \eta(\mathbf{x},0) = 0\}.$$

In what follows, we will re-formulate some results in optics in the standard form of probability theory and give complete proofs for them. A short presentation can be found in [1].

3.1. Mean number of dislocation points.

Let us consider a space variable $\mathbf{x}$ in $\mathbb{R}^2$ and a random wave with real part $\xi(\mathbf{x})$ and imaginary part $\eta(\mathbf{x})$. We define $\{\mathbf{Z}(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^2\}$ as the random field taking values in $\mathbb{R}^2$, with coordinates $\xi(\mathbf{x}), \eta(\mathbf{x})$. We assume that these two coordinates are independent centered Gaussian stationary isotropic random fields with C^2-paths and the same distribution. With no loss of generality, we also assume that $\mathrm{Var}(\xi(\mathbf{x})) = 1$.

First, we are interested in the expectation of the number of dislocation points

$$d_2 := \mathbb{E}[\#\{\mathbf{x} \in S : \xi(\mathbf{x}) = \eta(\mathbf{x}) = 0\}],$$

where S is a subset of the parameter space having area equal to 1, for simplicity.

Then, using the Rice formula for Gaussian fields ([2] Theorem 6.2) we get:

$$d_2 = \int_S \mathbb{E}[|\det(\mathbf{Z}'(\mathbf{x}))| \big| \mathbf{Z}(\mathbf{x}) = 0] p_{\mathbf{Z}(\mathbf{x})}(0) dx, \tag{22}$$

where $p_{\mathbf{Z}(\mathbf{x})}(.)$ is the density of $\mathbf{Z}(\mathbf{x})$. One can easily check that this density is non-degenerate. Moreover, one has (use Proposition 6.5. of [2]) $\mathbb{P}(\exists \mathbf{x}, \mathbf{Z}(\mathbf{x}) = 0, \det[\mathbf{Z}'(\mathbf{x}) = 0]) = 0$. These two conditions imply the validity of (22).

Set $\lambda_2 = \mathrm{Var}(\xi_i(\mathbf{x})) = \mathrm{Var}(\eta_i(\mathbf{x}))$, $i = 1, 2$. The stationarity implies, first, that the integrand in (22) is constant and, second, that $\mathbf{Z}(\mathbf{x})$ and $\mathbf{Z}'(\mathbf{x})$ are independent, so that the conditional expectation is in fact an ordinary expectation.

The entries of $\mathbf{Z}'(\mathbf{x})$ are four independent centered Gaussian variables with variance λ_2, so that, up to the factor λ_2, $|\det(\mathbf{Z}'(\mathbf{x}))|$ is the area of the parallelogram generated by two independent standard Gaussian variables in $\mathbb{R}^2$. Using invariance of the distribution, the distribution of this volume is the product of independent square roots of a $\chi^2(2)$ and a $\chi^2(1)$ distributed random variables. An elementary calculation gives then: $\mathbb{E}[|\det(\mathbf{Z}'(\mathbf{x}))|] = \lambda_2$. Finally, we get

$$d_2 = \frac{1}{2\pi}\lambda_2.$$

This quantity is equal to $\frac{K_2}{4\pi}$ in Berry and Dennis [3] notation, giving their formula (4.6).

3.2. **Mean length of dislocation curve.** Now suppose that the space variable is of dimension **3** and the random field $\{\mathbf{Z}(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^3\}$ satisfies the same hypotheses as in the 2-dimensional parameter case. Generically the dislocation points form a curve $\mathcal{C}$:

$$\mathcal{C} = \{\mathbf{x} : \mathbf{Z}(\mathbf{x}) = 0\}.$$

Our aim is to compute for each measurable subset S of $\mathbb{R}^3$:

$$d_3 = \mathbb{E}[\mathcal{L}(\mathcal{C} \cap S)],$$

where $\mathcal{L}$ is the length of the curve which is always defined, at least, as the Hausdorff measure of dimension 1. The Rice formula to be applied is now [2] Th 6.8 that reads

$$d_3 = \int_S \mathbb{E}[(\det \mathbf{Z}'(\mathbf{x})\ \mathbf{Z}'(\mathbf{x})^T)^{1/2} \big| \mathbf{Z}(\mathbf{x}) = 0] p_{\mathbf{Z}(\mathbf{x})}(0) d\mathbf{x},$$

and the verification of the validity is performed in a similar way to the 2-dimensional case above. For simplicity, we may assume again that S has Lebesgue measure equal to 1. The expression can be simplified using the stationarity and the normalization of the variance, to get

$$d_3 = \frac{1}{2\pi} \mathbb{E}[(\det \mathbf{Z}'(\mathbf{x}) \mathbf{Z}'(\mathbf{x})^T)^{1/2}],$$

with

$$\mathbb{E}[(\det(\mathbf{Z}'(\mathbf{x}) \mathbf{Z}'(\mathbf{x})^T)^{1/2}] = \lambda_2 \mathbb{E}(V),$$

where V is the surface area of the parallelogram generated by two standard Gaussian variables in $\mathbb{R}^3$. The projection method gives

$$\mathbb{E}(V) = \mathbb{E}(XY) = \frac{4}{\sqrt{2\pi}} \sqrt{\frac{\pi}{2}} = 2,$$

Here X and Y are independent and X (resp. Y) is the square root of a $\chi^2(3)$-distributed (resp. $\chi^2(2)$-distributed) random variable.

So,

$$d_3 = \frac{\lambda_2}{\pi}.$$

In Berry and Dennis' notations [3] the last quantity is denoted by $\frac{k_2}{3\pi}$ giving their formula (4.5).

3.3. **Variance.** In this section, we limit ourselves to dimension **2** and the random field satisfies the hypotheses we introduced to compute the expectation of the number of dislocation points. We further assume that for $\mathbf{s}_1, \mathbf{s}_2 \in \mathbb{R}^2, \mathbf{s}_1 \neq \mathbf{s}_2$ the joint distribution of $\xi(\mathbf{s}_1), \xi(\mathbf{s}_2)$

does not degenerate. Let S be again a measurable subset of $\mathbb{R}^2$ having Lebesgue measure equal to 1.

The variance of the number of dislocations points is an important issue that can be obtained via the second factorial moment of the number of zeroes. More precisely:

$$\operatorname{Var}\big(N_S^{\mathbf{Z}}(\mathbf{0})\big) = \mathbb{E}\big(N_S^{\mathbf{Z}}(\mathbf{0})\big(N_S^{\mathbf{Z}}(\mathbf{0}) - 1\big)\big) + d_2 - d_2^2,$$

and using Theorem 6.3 of [2], we can write the formula:

$$\mathbb{E}\big(N_S^{\mathbf{Z}}(\mathbf{0})\big(N_S^{\mathbf{Z}}(\mathbf{0}) - 1\big)\big) = \int_{S\times S} A(\mathbf{s}_1, \mathbf{s}_2) d\mathbf{s}_1 d\mathbf{s}_2,$$

where

$$A(\mathbf{s}_1, \mathbf{s}_2) = \mathbb{E}\Big(|\det \mathbf{Z}'(\mathbf{s}_1)\det \mathbf{Z}'(\mathbf{s}_2)|\big|\mathbf{Z}(\mathbf{s}_1) = \mathbf{Z}(\mathbf{s}_2) = 0\Big) p_{\mathbf{Z}(\mathbf{s}_1,\mathbf{s}_2)}(0,0).$$

Taking into account that the law of the random field is invariant under translations and orthogonal transformations of $\mathbb{R}^2$, we have

$$A(\mathbf{s}_1, \mathbf{s}_2) = A\big((0,0),(r,0)\big) = A(r) \quad \text{whith } r = \|\mathbf{s}_1 - \mathbf{s}_2\|.$$

The function $A(r)$ has two intuitive interpretations. First it can be viewed as

$$A(r) = \lim_{\epsilon\to 0} \frac{1}{\pi^2\epsilon^4} \mathbb{E}\big[N\big(B((0,0),\epsilon)\big) \times N\big(B((r,0),\epsilon)\big)\big].$$

Second it is the density of the Palm distribution (a generalization of horizontal window conditioning of [5]) of the number of zeroes of $\mathbf{Z}$ per unit of surface, locally around the point $(r,0)$ given that there is a zero at $(0,0)$. $A(r)/d_2^2$ is called the "correlation density function" in [3].

To compute $A(r)$, we recall that $\xi_1, \xi_2, \eta_1, \eta_2$ denote the partial derivatives of ξ, η with respect to the first and second coordinate.

So,

$$\begin{aligned} A(r) &= \mathbb{E}\Big[|\det \mathbf{Z}'(0,0)\det \mathbf{Z}'(r,0)|\Big|\mathbf{Z}(0,0)=\mathbf{Z}(r,0)=\mathbf{0}_2\Big]p_{\mathbf{Z}(0,0),\mathbf{Z}(r,0)}(\mathbf{0}_4) \\ &= \mathbb{E}\Big[\big|\big(\xi_1\eta_2-\xi_2\eta_1\big)(0,0)\big(\xi_1\eta_2-\xi_2\eta_1\big)(r,0)\big)\big|\Big|\mathbf{Z}(0,0)=\mathbf{Z}(r,0)=\mathbf{0}_2\Big] \\ &\qquad p_{\mathbf{Z}(0,0),\mathbf{Z}(r,0)}(\mathbf{0}_4), \end{aligned} \tag{23}$$

where $\mathbf{0}_p$ denotes the null vector in dimension p.

The density is easy to compute

$$p_{\mathbf{Z}(0,0),\mathbf{Z}(r,0)}(\mathbf{0}_4)=\frac{1}{(2\pi)^2(1-\rho^2(r))},\ \text{where } \rho(r)=\int_0^\infty J_0(kr)\Pi(dk).$$

Here, J_0 is the Bessel function of the first kind of order 0. The spectral measure μ is invariant under the isometries of $\mathbb{R}^2$, so that the measure Π on $\mathbb{R}^+$ is defined to be such that for every $w\geq 0$, $\mu(\tau : \tau\in\mathbb{R}^2, \|\tau\|\leq w)=2\pi\Pi([0,w])$.

To compute the conditional expectation of the product of the absolute value of the determinants, we use again the same device as in [3], as well as the same notations. We have:

$$|w|=\frac{1}{\pi}\int_{-\infty}^{+\infty}(1-\cos(wt))t^{-2}dt. \tag{24}$$

$$\begin{cases} C := \rho(r) \\ E = \rho'(r) \\ H = -E/r \\ F = -\rho''(r) \\ F_0 = -\rho''(0) \end{cases}$$

The regression formulas imply that the conditional variance matrix of the vector

$$\mathbf{W}=\Big(\xi_1(\mathbf{0}),\xi_1(r,0),\xi_2(\mathbf{0}),\xi_2(r,0),\eta_1(\mathbf{0}),\eta_1(r,0),\eta_2(\mathbf{0}),\eta_2(r,0)\Big),$$

is given by

$$\Sigma = Diag\Big[\mathcal{A},\mathcal{B},\mathcal{A},\mathcal{B}\Big]$$

with

$$\mathcal{A} = \begin{pmatrix} F_0 - \frac{E^2}{1-C^2} & F - \frac{E^2C}{1-C^2} \\ F - \frac{E^2C}{1-C^2} & F_0 - \frac{E^2}{1-C^2} \end{pmatrix}$$

$$\mathcal{B} = \begin{pmatrix} F_0 & H \\ H & F_0 \end{pmatrix}.$$

Using formula (24) the expectation we have to compute is equal to

(25)

$$\frac{1}{\pi^2}\int_{-\infty}^{+\infty} dt_1 \int_{-\infty}^{+\infty} dt_2 t_1^{-2} t_2^{-2}\Big[1 - \frac{1}{2}T(t_1,0) - \frac{1}{2}T(-t_1,0) - \frac{1}{2}T(0,t_2) - \frac{1}{2}T(0,-t_2) + \frac{1}{4}T(t_1,t_2) + \frac{1}{4}T(-t_1,t_2) + \frac{1}{4}T(t_1,-t_2) + \frac{1}{4}T(-t_1,-t_2)\Big],$$

where

$$T(t_1,t_2) = \mathbb{E}\Big[\exp\big(i(w_1t_1 + w_2t_2)\big)\Big],$$

with

$$w_1 = \xi_1(\mathbf{0})\eta_2(\mathbf{0}) - \eta_1(\mathbf{0})\xi_2(\mathbf{0}) = \mathbf{W}_1\mathbf{W}_7 - \mathbf{W}_3\mathbf{W}_5$$

$$w_2 = \xi_1(r,0)\eta_2(r,0) - \eta_1(r,0)\xi_2(r,0) = \mathbf{W}_2\mathbf{W}_8 - \mathbf{W}_4\mathbf{W}_6.$$

$T(t_1,t_2) = \mathbb{E}\big(\exp(i\mathbf{W}^T\mathcal{H}\mathbf{W})\big)$ where $\mathbf{W}$ has the distribution $N(0,\Sigma)$ and

$$\mathcal{H} = \begin{bmatrix} 0 & 0 & 0 & \mathcal{D} \\ 0 & 0 & -\mathcal{D} & 0 \\ 0 & -\mathcal{D} & 0 & 0 \\ \mathcal{D} & 0 & 0 & 0 \end{bmatrix},$$

$$\mathcal{D} = \frac{1}{2}\begin{bmatrix} t_1 & 0 \\ 0 & t_2 \end{bmatrix}.$$

A standard diagonalization argument shows that

$$T(t_1,t_2) = \mathbb{E}\big(\exp(i\mathbf{W}^T\mathcal{H}\mathbf{W})\big) = \mathbb{E}\big(\exp(i\sum_{j=1}^{8}\lambda_j\zeta_j^2)\big),$$

where the ζ_j's are independent with standard normal distribution and the λ_j are the eigenvalues of $\Sigma^{1/2}\mathcal{H}\Sigma^{1/2}$. Using the characteristic function of the $\chi^2(1)$ distribution:

$$\mathbb{E}\big(\exp(i\mathbf{W}^T\mathcal{H}\mathbf{W})\big) = \prod_{j=1}^{8}(1-2i\lambda_j)^{-1/2}. \tag{26}$$

Clearly

$$\Sigma^{1/2} = Diag\Big[\mathcal{A}^{1/2}, \mathcal{B}^{1/2}, \mathcal{A}^{1/2}, \mathcal{B}^{1/2}\Big],$$

and

$$\Sigma^{1/2}\mathcal{H}\Sigma^{1/2} = \begin{bmatrix} 0 & 0 & 0 & \mathcal{M} \\ 0 & 0 & -\mathcal{M}^T & 0 \\ 0 & -\mathcal{M} & 0 & 0 \\ \mathcal{M}^T & 0 & 0 & 0 \end{bmatrix},$$

with $\mathcal{M} = \mathcal{A}^{1/2}\mathcal{D}\mathcal{B}^{1/2}$.

Let λ be an eigenvalue of $\Sigma^{1/2}\mathcal{H}\Sigma^{1/2}$. It is easy to check that λ^2 is an eigenvalue of $\mathcal{M}\mathcal{M}^T$. Respectively if λ_1^2 and λ_2^2 are the eigenvalues of $\mathcal{M}\mathcal{M}^T$, those of $\Sigma^{1/2}\mathcal{H}\Sigma^{1/2}$ are $\pm\lambda_1$(twice) and $\pm\lambda_2$ (twice).

Note that λ_1^2 and λ_2^2 are the eigenvalues of $\mathcal{M}\mathcal{M}^T = \mathcal{A}^{1/2}\mathcal{D}\mathcal{B}\mathcal{D}\mathcal{A}^{1/2}$ or equivalently, of $\mathcal{D}\mathcal{B}\mathcal{D}\mathcal{A}$. Using (26)

$$\begin{aligned}\mathbb{E}\big(\exp(i\mathbf{W}^T\mathcal{H}\mathbf{W})\big) &= \big(1+4(\lambda_1^2+\lambda_2^2)+16\lambda_1^2\lambda_2^2\big)^{-1} \\ &= \big(1+4\ tr(\mathcal{D}\mathcal{B}\mathcal{D}\mathcal{A})+16\det(\mathcal{D}\mathcal{B}\mathcal{D}\mathcal{A})\big)^{-1}\end{aligned}$$

where

$$\mathcal{D}\mathcal{B}\mathcal{D}\mathcal{A} = \frac{1}{4}\begin{bmatrix} t_1^2F_0(F_0-\frac{E^2}{1-C^2})+t_1t_2H(F-\frac{E^2C}{1-C^2}) & t_1^2F_0(F-\frac{E^2C}{1-C^2})+t_1t_2H(F_0-\frac{E^2}{1-C^2}) \\ t_1t_2H(F_0-\frac{E^2}{1-C^2})+t_2^2F_0(F-\frac{E^2C}{1-C^2}) & t_1t_2H(F-\frac{E^2C}{1-C^2})+t_2^2F_0(F_0-\frac{E^2}{1-C^2}) \end{bmatrix}$$

So,

$$4\ tr(\mathcal{D}\mathcal{B}\mathcal{D}\mathcal{A}) = (t_1^2+t_2^2)F_0(F_0-\frac{E^2}{1-C^2})+2t_1t_2H(F-\frac{E^2C}{1-C^2}) \tag{27}$$

$$16\det(\mathcal{D}\mathcal{B}\mathcal{D}\mathcal{A}) = t_1^2t_2^2\big[F_0^2-H^2\big]\big[(F_0-\frac{E^2}{1-C^2})^2-(F-\frac{E^2C}{1-C^2})^2\big], \tag{28}$$

giving

$$(29)\quad T(t_1,t_2) = \mathbb{E}\Big(\exp(i\mathbf{W}^T\mathcal{H}\mathbf{W})\Big)$$
$$= \Big(1 + (t_1^2+t_2^2)F_0(F_0 - \frac{E^2}{1-C^2}) + 2t_1t_2H(F - \frac{E^2C}{1-C^2})$$
$$+ t_1^2t_2^2\big[F_0^2 - H^2\big]\big[(F_0 - \frac{E^2}{1-C^2})^2 - (F - \frac{E^2C}{1-C^2})^2\big]\Big)^{-1}.$$

Performing the change of variable $t' = \sqrt{A_1}t$ with $A_1 = F_0(F_0 - \frac{E^2}{1-C^2})$ the integral (25) becomes

$$(30)\quad \frac{A_1}{\pi^2}\int_{-\infty}^{+\infty} dt_1 \int_{-\infty}^{+\infty} dt_2 t_1^{-2}t_2^{-2}$$
$$\Big[1 - \frac{1}{1+t_1^2}\frac{1}{1+t_2^2} + -\frac{1}{2}\Big\{\frac{1}{1+(t_1^2+t_2^2) - 2A_2t_1t_2 + t_1^2t_2^2Z} +$$
$$\frac{1}{1+(t_1^2+t_2^2) + 2A_2t_1t_2 + t_1^2t_2^2Z}\Big\}\Big]$$
$$= \frac{A_1}{\pi^2}\int_{-\infty}^{+\infty} dt_1 \int_{-\infty}^{+\infty} dt_2 t_1^{-2}t_2^{-2}$$
$$\Big[1 - \frac{1}{1+t_1^2} - \frac{1}{1+t_2^2} + \frac{1+(t_1^2+t_2^2)+t_1^2t_2^2Z}{\Big(1+(t_1^2+t_2^2)+t_1^2t_2^2Z\Big)^2 - 4A_2^2t_1^2t_2^2}\Big],$$

where

$$\begin{cases} A_2 = \frac{H}{F_0}\frac{F(1-C^2)-E^2C}{F_0(1-C^2)-E^2} \\ Z = \frac{F_0^2-H^2}{F_0^2}\Big[1 - (F - \frac{E^2C}{1-C^2})^2.(F_0 - \frac{E^2}{1-C^2})^{-2}\Big]. \end{cases}$$

In this form, and up to a sign change, this result is equivalent to Formula (4.43) of [3] (note that $A_2^2 = Y$ in [3]).

In order to compute the integral (30), first we obtain

$$\int_{-\infty}^{\infty} \frac{1}{t_2^2}\big[1 - \frac{1}{1+t_2^2}\big]dt_2 = \pi.$$

We split the other term into two integrals, thus we have for the first one

$$
\begin{aligned}
&\frac{1}{2}\int_{-\infty}^{\infty}\frac{1}{t_2^2}\Big[\frac{1}{1+(t_1^2+t_2^2)-2A_2t_1t_2+t_1^2t_2^2Z}-\frac{1}{1+t_1^2}\Big]dt_2\\
&\qquad=-\frac{1}{2(1+t_1^2)}\int_{-\infty}^{\infty}\frac{1}{t_2^2}\frac{(1+t_1^2Z)t_2^2-2A_2t_1t_2}{1+t_1^2-2A_2t_1t_2+(1+t_1^2Z)t_2^2}dt_2\\
&\qquad=-\frac{1}{2(1+t_1^2)}\int_{-\infty}^{\infty}\frac{1}{t_2^2}\frac{t_2^2-2Z_1t_1t_2}{t_2^2-2Z_1t_1t_2+Z_2}dt_2=I_1,
\end{aligned}
$$

where $Z_2=\frac{1+t_1^2}{1+Zt_1^2}$ and $Z_1=\frac{A_2}{1+Zt_1^2}$.

Similarly, for the second integral we get

$$
\begin{aligned}
&\frac{1}{2}\int_{-\infty}^{\infty}\frac{1}{t_2^2}\Big[\frac{1}{1+(t_1^2+t_2^2)+2A_2t_1t_2+t_1^2t_2^2Z}-\frac{1}{1+t_1^2}\Big]dt_2\\
&\qquad\qquad=-\frac{1}{2(1+t_1^2)}\int_{-\infty}^{\infty}\frac{1}{t_2^2}\frac{t_2^2+2Z_1t_1t_2}{t_2^2+2Z_1t_1t_2+Z_2}\,dt_2=I_2
\end{aligned}
$$

$$
\begin{aligned}
I_1+I_2&=-\frac{1}{2(1+t_1^2)}\int_{-\infty}^{\infty}\frac{1}{t_2^2}\Big[\frac{t_2^2-2Z_1t_1t_2}{t_2^2-2Z_1t_1t_2+Z_2}+\frac{t_2^2+2Z_1t_1t_2}{t_2^2+2Z_1t_1t_2+Z_2}\Big]dt_2\\
&=-\frac{1}{(1+t_1^2)}\int_{-\infty}^{\infty}\frac{t_2^2+(Z_2-4Z_1^2t_1^2)}{t_2^4+2(Z_2-2Z_1^2t_1^2)t_2^2+Z_2^2}\,dt_2\\
&=-\frac{1}{(1+t_1^2)}\frac{\pi(Z_2-2Z_1^2t_1^2)}{Z_2\sqrt{(Z_2-Z_1^2t_1^2)}}.
\end{aligned}
$$

In the third line we have used the formula provided by the method of residues. In fact, if the polynomial X^2-SX+P with $P>0$ has not root in $[0,\infty)$, then

$$
\int_{-\infty}^{\infty}\frac{t^2-\gamma}{t^4-St^2+P}\,dt=\frac{\pi}{\sqrt{P(-S+2\sqrt{P})}}(\sqrt{P}-\gamma).
$$

In our case $\gamma=-(Z_2-4Z_1^2t_1^2)$, $S=-2(Z_2-2Z_1^2t_1^2)$ and $P=Z_2^2$.

Therefore we get

$$A(r) = \frac{A_1}{4\pi^3(1-C^2)} \int_{-\infty}^{\infty} \frac{1}{t_1^2}\big[1 - \frac{1}{(1+t_1^2)} \frac{(Z_2 - 2Z_1^2 t_1^2)}{Z_2\sqrt{(Z_2 - Z_1^2 t_1^2)}}\big]\, dt_1.$$

Acknowledgement. This work has received financial support from European Marie Curie Network SEAMOCS.

References

[1] J-M. Azaïs, J. León and M. Wschebor, Rice formulae and Gaussian waves (2010) *Bernoulli*, vol. 17, No. 1, 170-193 (2011).

[2] J-M. Azaïs, and M. Wschebor, *Level sets and extrema of random processes and fields*, Wiley (2009).

[3] M.V. Berry, and M.R. Dennis, Phase singularities in isotropic random waves, *Proc. R. Soc. Lond*, A, 456, 2059-2079 (2000).

[4] L. Callenbach, P. Hänggi and S.J. Linz, Oscillatory systems driven by noise: Frequency and phase syncronisation, *Physical Review E* vol. 65, 051110 (2002).

[5] H. Cramér and M.R. Leadbetter, *Stationary and Related Stochastic Processes*, Wiley (1967).

[6] Li and Wei, Gaussian integrals involving absolute value functions. IMS Lecture Notes Monograph Series. Proceedings of the Conference in Luminy (IMS) (2009).

[7] M. S. Longuet-Higgins, Reflection and refraction at a random surface. I, II, III, *Journal of the Optical Society of America*, vol. 50, No. 9, 838-856 (1960).

[8] M. S. Longuet-Higgins, The statistical geometry of random surfaces. *Proc. Symp. Appl. Math.*, Vol. XIII, AMS Providence R.I., 105-143 (1962).

[9] A.O. Petters, B. Rider and A.M. Teguia, A Theory of Stochastic Microlensing I. Random Images, Shear and the Kac-Rice Formula. *J. Math. Physics*, vol. 50, pp. 122501, (2009).

[10] S.O. Rice,(1944-1945). Mathematical Analysis of Random Noise, *Bell System Tech. J.*, 23, 282-332; 24, 45-156 (1944-1945).

Université de Toulouse, IMT, ESP, F31062 Toulouse Cedex 9, France.
E-mail address: jean-marc.azais@math.univ-toulouse.fr

Escuela de Matemática. Facultad de Ciencias. Universidad Central de Venezuela. A.P. 47197, Los Chaguaramos, Caracas 1041-A, Venezuela.
E-mail address: jose.leon@ciens.ucv.ve

Centro de Matemática. Facultad de Ciencias. Universidad de la República. Calle Iguá 4225. 11400. Montevideo. Uruguay.
E-mail address: wschebor@cmat.edu.uy

ON AUTOMORPHISM GROUPS OF FIBER BUNDLES

MICHEL BRION

Abstract. We obtain analogues of classical results on automorphism groups of holomorphic fiber bundles, in the setting of group schemes. Also, we establish a lifting property of the connected automorphism group, for torsors under abelian varieties. These results will be applied to the study of homogeneous bundles over abelian varieties.

1. Introduction

This work arose from a study of homogeneous bundles over an abelian variety A, that is, of those principal bundles with base A and fiber an algebraic group G, that are isomorphic to all of their pull-backs by the translations of A (see [Br2]). In the process of that study, it became necessary to obtain algebro-geometric analogues of two classical results about automorphisms of fiber bundles in complex geometry. The first one, due to Morimoto (see [Mo]), asserts that the equivariant automorphism group of a principal bundle over a compact complex manifold, with fiber a complex Lie group, is a complex Lie group as well. The second one, a result of Blanchard (see [Bl]), states that a holomorphic action of a complex connected Lie group on the total space of a locally trivial fiber bundle of complex manifolds descends to a holomorphic action on the base, provided that the fiber is compact and connected.

Also, we needed to show the existence in the category of schemes of certain fiber bundles associated to a G-torsor (or principal bundles)

2010 *Mathematics Subject Classification.* 14L10, 14L15, 14L30.

$\pi : X \to Y$, where G is a connected group scheme and X, Y are algebraic schemes; namely, those fiber bundles $X \times^G Z \to Y$ associated to G-homogeneous varieties Z. Note that the fiber bundle associated to an arbitrary G-scheme Z exists in the category of algebraic spaces, but may fail to be a scheme (see [Bi, KM]).

Finally, we were led to a lifting result which reduces the study of homogeneous bundles to the case that the structure group is linear, and does not seem to have its holomorphic counterpart. It asserts that given a G-torsor $\pi : X \to Y$ where G is an abelian variety and X, Y are smooth complete algebraic varieties, the connected automorphism group of X maps onto that of Y under the homomorphism provided by the analogue of Blanchard's theorem.

In this paper, we present these preliminary results which may have independent interest, with (hopefully) modest prerequisites. Section 2 is devoted to a scheme-theoretic version of Blanchard's theorem: a proper morphism of schemes $\pi : X \to Y$ such that $\pi_*(\mathcal{O}_X) = \mathcal{O}_Y$ induces a homomorphism $\pi_* : \mathrm{Aut}^o(X) \to \mathrm{Aut}^o(Y)$ between the neutral components of the automorphism group schemes (Corollary 2.2). Our proof is an adaptation of that given in [Ak] in the setting of complex spaces.

In Section 3, we consider a torsor $\pi : X \to Y$ under a connected group scheme G, and show the existence of the associated fiber bundle $X \times^G G/H = X/H$ for any subgroup scheme $H \subset G$ (Theorem 3.3). As a consequence, $X \times^G Z$ exists when Z is the total space of a G-torsor, or a group scheme where G acts via a homomorphism (Corollary 3.4). Another application of Theorem 3.3 concerns the quasi-projectivity of torsors (Corollary 3.5); it builds on work of Raynaud, who showed e.g. the local quasi-projectivity of homogeneous spaces over a normal scheme (see [Ra]).

The automorphism groups of torsors are studied in Section 4. In particular, we obtain a version of Morimoto's theorem: the equivariant automorphisms of a torsor over a proper scheme form a group scheme, locally of finite type (Theorem 4.2). Here our proof, based on an equivariant completion of the structure group, is quite different from the original one. We also analyze the relative equivariant automorphism group of such a torsor; this yields a version of Chevalley's structure theorem for algebraic groups in that setting (Proposition 4.3).

The final Section 5 contains a full description of relative equivariant automorphisms for torsors under abelian varieties (Proposition 5.1) and our lifting result for automorphisms of the base (Theorem 5.4).

Acknowledgements. Many thanks to Gaël Rémond for several clarifying discussions, and special thanks to the referee for very helpful comments and corrections. In fact, the final step of the proof of Theorem 3.3 is taken from the referee's report; the end of the proof of Corollary 2.2, and the proof of Corollary 3.4 (ii), closely follow his/her suggestions.

Notation and conventions. Throughout this article, we consider algebraic varieties, schemes, and morphisms over an algebraically closed field k. Unless explicitly mentioned, we will assume that the considered schemes are of finite type over k (such schemes are also called algebraic schemes). By a point of a scheme X, we will mean a closed point unless explicitly mentioned. A *variety* is an integral separated scheme.

We will use [DG] as a general reference for group schemes. Given such a group scheme G, we denote by $\mu_G : G \times G \to G$ the multiplication and by $e_G \in G(k)$ the neutral element. The neutral component of G is denoted by G^o, and the Lie algebra by $\mathrm{Lie}(G)$.

We recall that an *action* of G on a scheme X is a morphism

$$\alpha : G \times X \longrightarrow X, \quad (g, x) \longmapsto g \cdot x$$

such that the composite map

$$X \xrightarrow{e_G \times \mathrm{id}_X} G \times X \xrightarrow{\alpha} X$$

is the identity, and the square

$$\begin{array}{ccc} G \times G \times X & \xrightarrow{\mathrm{id}_G \times \alpha} & G \times X \\ {\scriptstyle \mu_G \times \mathrm{id}_X} \big\downarrow & & {\scriptstyle \alpha} \big\downarrow \\ G \times X & \xrightarrow{\alpha} & X \end{array}$$

commutes. We then say that X is a *G-scheme*. A morphism $f : X \to Y$ between two G-schemes is called *equivariant* if the square

$$\begin{array}{ccc} G \times X & \xrightarrow{\alpha} & X \\ {\scriptstyle \mathrm{id}_G \times f} \big\downarrow & & {\scriptstyle f} \big\downarrow \\ G \times Y & \xrightarrow{\beta} & Y \end{array}$$

commutes (with the obvious notation). We then say that f is a *G-morphism*.

A smooth group scheme will be called an algebraic group. By Chevalley's structure theorem (see [Ro, Theorem 16], or [Co] for a modern proof), every connected algebraic group G has a largest closed connected normal affine subgroup G_{aff}; moreover, the quotient $G/G_{\mathrm{aff}} =: A(G)$ is an abelian variety. This yields an exact sequence of connected algebraic groups

$$1 \longrightarrow G_{\mathrm{aff}} \longrightarrow G \longrightarrow A(G) \longrightarrow 1.$$

2. Descending automorphisms for fiber spaces

We begin with the following scheme-theoretic version of a result of Blanchard (see [Bl, Section I.1] and also [Ak, Lemma 2.4.2]).

Proposition 2.1. *Let G be a connected group scheme, X a G-scheme, Y a scheme, and $\pi : X \to Y$ a proper morphism such that $\pi_*(\mathcal{O}_X) = \mathcal{O}_Y$. Then there is a unique G-action on Y such that π is equivariant.*

Proof. We will consider a scheme Z as the ringed space $(Z(k), \mathcal{O}_Z)$ where the set $Z(k)$ is equipped with the Zariski topology; this makes sense as Z is of finite type.

We first claim that the abstract group $G(k)$ permutes the fibers of $\pi : X(k) \to Y(k)$ (note that these fibers are non-empty and connected, since $\pi_*(\mathcal{O}_X) = \mathcal{O}_Y$). Let $y \in Y(k)$ and denote by F_y the set-theoretic fiber of π at y, viewed as a closed reduced subscheme of X. Then the map

$$\varphi : G_{\mathrm{red}} \times F_y \longrightarrow Y, \quad (g, x) \longmapsto \pi(g \cdot x)$$

maps $\{e_G\} \times F_y$ to the point y. Moreover, G_{red} is a variety, and F_y is connected and proper. By the rigidity lemma (see [Mu, p. 43]), it follows that φ maps $\{g\} \times F_y$ to a point for any $g \in G(k)$, i.e., $g \cdot F_y \subset F_{g \cdot y}$. Thus, $g^{-1} \cdot F_{g \cdot y} \subset F_y$ and hence $g \cdot F_y = F_{g \cdot y}$. This implies our claim.

That claim yields a commutative square

$$\begin{array}{ccc} G(k) \times X(k) & \xrightarrow{\alpha} & X(k) \\ {\scriptstyle \mathrm{id}_G \times \pi} \big\downarrow & & {\scriptstyle \pi} \big\downarrow \\ G(k) \times Y(k) & \xrightarrow{\beta} & Y(k), \end{array}$$

where β is an action of the (abstract) group $G(k)$.

Next, we show that β is continuous. It suffices to show that $\beta^{-1}(Z)$ is closed for any closed subset $Z \subset Y(k)$. But $(\mathrm{id}_G, \pi)^{-1}\beta^{-1}(Z) = \alpha^{-1}\pi^{-1}(Z)$ is closed, and (id_G, π) is proper and surjective; this yields our assertion.

Finally, we define a morphism of sheaves of k-algebras

$$\beta^\# : \mathcal{O}_Y \longrightarrow \beta_*(\mathcal{O}_{G \times Y}).$$

For this, to any open subset $V \subset Y$, we associate a homomorphism of algebras

$$\beta^\#(V) : \mathcal{O}_Y(V) \longrightarrow \mathcal{O}_{G \times Y}\big(\beta^{-1}(V)\big).$$

By assumption, the left-hand side is isomorphic to $\mathcal{O}_X\big(\pi^{-1}(V)\big)$, and the right-hand side to

$$\mathcal{O}_{G \times X}\big((\mathrm{id}_G, \pi)^{-1}\beta^{-1}(V)\big) = \mathcal{O}_{G \times X}\big(\alpha^{-1}\pi^{-1}(V)\big).$$

We define $\beta^\#(V) := \alpha^\#\big(\pi^{-1}(V)\big)$. Now it is straightforward to verify that $(\beta, \beta^\#)$ is a morphism of locally ringed spaces; this yields a morphism of schemes $\beta : G \times Y \to Y$. By construction, β is the unique morphism such that the square

$$\begin{array}{ccc} G \times X & \xrightarrow{\alpha} & X \\ {\scriptstyle \mathrm{id}_G \times \pi}\big\downarrow & & {\scriptstyle \pi}\big\downarrow \\ G \times Y & \xrightarrow{\beta} & Y \end{array}$$

commutes.

It remains to show that β is an action of the group scheme G. Note that e_G acts on $X(k)$ via the identity; moreover, the composite morphism of sheaves

$$\mathcal{O}_Y \xrightarrow{\beta^{\#}} \beta_*(\mathcal{O}_{G \times Y}) \xrightarrow{(e_G \times \mathrm{id}_Y)^{\#}} \beta_*(\mathcal{O}_{\{e_G\} \times Y}) \cong \mathcal{O}_Y$$

is the identity, since so is the analogous morphism

$$\mathcal{O}_X \xrightarrow{\alpha^{\#}} \alpha_*(\mathcal{O}_{G \times X}) \xrightarrow{(e_G \times \mathrm{id}_X)^{\#}} \alpha_*(\mathcal{O}_{\{e_G\} \times X}) \cong \mathcal{O}_X$$

and $\pi_*(\mathcal{O}_X) = \mathcal{O}_Y$. Likewise, the square

$$\begin{array}{ccc} G \times G \times Y & \xrightarrow{\mathrm{id}_G \times \beta} & G \times Y \\ {\scriptstyle \mu_G \times \mathrm{id}_Y}\big\downarrow & & {\scriptstyle \beta}\big\downarrow \\ G \times Y & \xrightarrow{\beta} & Y \end{array}$$

commutes on closed points, and the corresponding square of morphisms of sheaves commutes as well, since the analogous square with Y replaced by X commutes. □

This proposition will imply a result of descent for group scheme actions, analogous to [Bl, Proposition I.1] (see also [Ak, Proposition 2.4.1]). To state that result, we need some recollections on automorphism functors.

Given a scheme S, we denote by $\mathrm{Aut}_S(X \times S)$ the group of automorphisms of $X \times S$ viewed as a scheme over S. The assignement $S \mapsto \mathrm{Aut}_S(X \times S)$ yields a group functor $Aut(X)$, i.e., a contravariant functor from the category of schemes to that of groups. If X is proper, then $Aut(X)$ is represented by a group scheme $\mathrm{Aut}(X)$, locally of finite type (see [MO, Theorem 3.7]). In particular, the neutral component $\mathrm{Aut}^o(X)$ is a group scheme of finite type. Also, recall that

$$\mathrm{Lie}\,\mathrm{Aut}(X) \cong \Gamma(X, T_X) \tag{1}$$

where the right-hand side denotes the Lie algebra of global vector fields on X, that is, of derivations of $\mathcal{O}_X$.

We now are in a position to state:

Corollary 2.2. *Let $\pi : X \to Y$ be a morphism of proper schemes such that $\pi_*(\mathcal{O}_X) = \mathcal{O}_Y$. Then π induces a homomorphism of group schemes*

$$\pi_* : \mathrm{Aut}^o(X) \longrightarrow \mathrm{Aut}^o(Y).$$

Proof. This is a formal consequence of Proposition 2.1. Specifically, let $G := \mathrm{Aut}^o(X)$ and consider the G-action on Y obtained in that proposition. This yields a automorphism of $Y \times G$ as a scheme over G,

$$(y, g) \longmapsto (g \cdot y, g),$$

and in turn a morphism (of schemes)

$$\pi_* : G \longrightarrow \mathrm{Aut}(Y).$$

Moreover, $\pi_*(e_G) = e_{\mathrm{Aut}(Y)}$ since e_G acts via the identity. As G is connected, it follows that the image of π_* is contained in $\mathrm{Aut}^o(Y) =: H$. In other words, we have a morphism of schemes $\pi_* : G \to H$ such that $\pi_*(e_G) = e_H$. It remains to check that π_* is a homomorphism; but this follows from the fact that π_* corresponds to the G-action on Y, and hence yields a morphism of group functors. □

Given two complete varieties X and Y, the preceding corollary applies to the projections

$$p : X \times Y \to X, \quad q : X \times Y \to Y$$

and yields homomorphisms

$$p_* : \mathrm{Aut}^o(X)\times\mathrm{Aut}^o(Y) \to \mathrm{Aut}^o(X), \quad q_* : \mathrm{Aut}^o(X)\times\mathrm{Aut}^o(Y) \to \mathrm{Aut}^o(Y).$$

This implies readily the following analogue of [Bl, Corollaire, p. 161]:

Corollary 2.3. *Let X and Y be complete varieties. Then the homomorphism*

$$(p_*, q_*) : \mathrm{Aut}^o(X \times Y) \longrightarrow \mathrm{Aut}^o(X) \times \mathrm{Aut}^o(Y)$$

is an isomorphism, with inverse the natural homomorphism

$$\mathrm{Aut}^o(X)\times\mathrm{Aut}^o(Y) \longrightarrow \mathrm{Aut}^o(X\times Y), \quad (g,h) \longmapsto \big((x,y) \mapsto (g(x), h(y)\big).$$

More generally, the isomorphism

$$\mathrm{Aut}^o(X \times Y) \cong \mathrm{Aut}^o(X) \times \mathrm{Aut}^o(Y)$$

holds for those proper schemes X and Y such that $\mathcal{O}(X) = \mathcal{O}(Y) = k$, but may fail for arbitrary proper schemes. Indeed, let X be a complete variety having non-zero global vector fields, and let $Y := \mathrm{Spec}\, k[\varepsilon]$ where $\epsilon^2 = 0$; denote by y the closed point of Y. Then we have an exact sequence

$$1 \longrightarrow \Gamma(X, T_X) \longrightarrow \mathrm{Aut}_Y(X \times Y) \longrightarrow \mathrm{Aut}(X) \longrightarrow 1,$$

where the map on the right is obtained by restricting to $X \times \{y\}$. This identifies the vector group $\Gamma(X, T_X)$ to a closed subgroup of $\mathrm{Aut}^o(X\times Y)$, which is not in the image of the natural homomorphism.

Likewise, $\mathrm{Aut}(X \times Y)$ is generally strictly larger than $\mathrm{Aut}(X) \times \mathrm{Aut}(Y)$ (e.g. take $Y = X$ and consider the automorphism $(x, y) \mapsto (y, x)$).

3. Torsors and asssociated fiber bundles

Consider a group scheme G, a G-scheme X, and a G-invariant morphism

$$\pi : X \longrightarrow Y, \tag{2}$$

where Y is a scheme. We say that X is a *G-torsor over Y*, if π is faithfully flat and the morphism

$$\alpha \times p_2 : G \times X \longrightarrow X \times_Y X, \quad (g, x) \longmapsto (g \cdot x, x) \tag{3}$$

is an isomorphism. The latter condition is equivalent to the existence of a faithfully flat morphism $f : Y' \to Y$ such that the pull-back torsor $\pi' : X\times_Y Y' \to Y'$ is trivial. (Since our schemes are assumed to be of finite type, π is quasi-compact and finitely presented; thus, there is no need to distinguish between the fppf and the fpqc topology).

For a G-torsor (2), the morphism π is surjective, and its geometric fiber $X_{\bar{y}}$ is isomorphic to $G_{\bar{y}}$ for any (possibly non-closed) point $y \in Y$. In particular, π is smooth if and only if G is an algebraic group; under that assumption, X is smooth (resp. normal) if and only if so is Y.

Also, note that π is a universal geometric quotient in the sense of [MFK, Section 0], and hence a universal categorical quotient (see [loc. cit., Proposition 0.1]). In particular, $Y(k) = X(k)/G(k)$ and $\mathcal{O}_Y = \pi_*(\mathcal{O}_X)^G$ (the subsheaf of G-invariants in $\pi_*(\mathcal{O}_X)$). Thus, we will also denote Y by X/G.

Remark 3.1. If G is an affine algebraic group, then every G-torsor (2) is *locally isotrivial*, i.e., for any point $y \in Y$ there exist an open subscheme $V \subset Y$ containing y and a finite étale surjective morphism $f : V' \to V$ such that the pull-back torsor $X \times_V V'$ is trivial (this result is due to Grothendieck, see [Ra, Lemme XIV 1.4] for a detailed proof). The local isotriviality of π also holds if G is an algebraic group and Y_{red} is normal, as a consequence of [loc. cit., Théorème XIV 1.2]. In particular, π is locally trivial for the étale topology in both cases.

Yet there exist torsors under algebraic groups that are not locally isotrivial, see [loc. cit., XIII 3.1] (reformulated in more concrete terms in [Br1, Example 6.2]) for an example where Y is a rational nodal curve, and G is an abelian variety having a point of infinite order.

Given a G-torsor (2) and a G-scheme Z, we may view $X \times Z$ as a G-scheme for the diagonal action, and ask if there exist a G-torsor $\varpi : X \times Z \to W$ where W is a scheme, and a morphism $q : W \to Y$ such that the square

$$\begin{array}{ccc} X \times Z & \xrightarrow{p_1} & X \\ \varpi \downarrow & & \pi \downarrow \\ W & \xrightarrow{q} & Y \end{array}$$

is cartesian; here p_1 denotes the first projection. Then q is called the *associated fiber bundle with fiber Z*. The quotient scheme W will be denoted by $X \times^G Z$.

The answer to this question is positive if Z admits an ample G-linearized invertible sheaf (as follows from descent theory; see [SGA1, Proposition 7.8] and also [MFK, Proposition 7.1]). In particular, the answer is positive if Z is affine. Yet the answer is generally negative, even if Z is a smooth variety; see [Bi]. However, associated fiber bundles do exist in the category of algebraic spaces, see [KM, Corollary 1.2].

Of special interest is the case that the fiber is a group scheme G' where G acts through a homomorphism $f : G \to G'$. Then $X' := X \times^G G'$ is a G'-torsor over Y, obtained from X by *extension of the structure group*. If f identifies G with a closed subgroup scheme of G', then X' comes with a G'-morphism to G'/G arising from the projection $X \times G' \to G'$. Conversely, the existence of such a morphism yields a reduction of structure group, in view of the following standard result:

Lemma 3.2. *Let G be a group scheme, H a subgroup scheme, and X a G-scheme equipped with a G-morphism $f : X \to G/H$. Denote by Z the fiber of f at the base point of G/H, so that Z is an H-scheme.*

Then f is faithfully flat, and the natural map $G \times Z \to X$ factors through a G-isomorphism $G \times^H Z \cong X$.

If $\pi : X \to Y$ is a G-torsor, then the restriction $\pi_{|Z} : Z \to Y$ is an H-torsor.

Proof. Form and label the cartesian square

$$\begin{array}{ccc} X' & \xrightarrow{f'} & G \\ {\scriptstyle q'}\downarrow & & \downarrow{\scriptstyle q} \\ X & \xrightarrow{f} & G/H \end{array}$$

where q denotes the quotient map. Then X' is a G-scheme and f' is a G-morphism with fiber Z at e_G. It follows readily that the morphism

$$G \times Z \longrightarrow X', \quad (g, z) \longmapsto g \cdot z$$

is an isomorphism, with inverse

$$X' \longrightarrow G \times Z, \quad x' \longmapsto \big(f'(x'), f'(x')^{-1} \cdot x'\big).$$

This identifies f' with the projection $G \times X \to G$; in particular, f' is faithfully flat. Since q is an H-torsor, f is faithfully flat as well; moreover, q' is an H-torsor. This yields the first assertion.

Next, the G-torsor $\pi : X \to Y$ yields a $G \times H$-torsor

$$F : G \times Z \longrightarrow Y, \quad (g, z) \longmapsto \pi(g \cdot z).$$

Moreover, F is the composite of the projection $G \times Z \to Z$ followed by $\pi_{|Z}$. Thus, $\pi_{|Z}$ is faithfully flat. It remains to show that the natural morphism $H \times Z \to Z \times_Y Z$ is an isomorphism. But this follows by considering the isomorphism (3) and taking the fiber of the morphism $f \times f : X \times_Y X \to G/H \times G/H$ at the base point of $G/H \times G/H$. $\square$

Returning to a G-torsor (2) and a G-scheme Z, we now show that the associated fiber bundle $X \times^G Z$ is a scheme in the case that G is connected and acts transitively on Z. Then $Z \cong G/H$ for some subgroup scheme $H \subset G$, and hence $X \times^G Z \cong X/H$ as algebraic spaces.

THEOREM 3.3. *Let G be a connected group scheme, $\pi : X \to Y$ a G-torsor, and $H \subset G$ a subgroup scheme. Then:*

(i) *π factors uniquely as the composite*

$$X \xrightarrow{p} Z \xrightarrow{q} Y, \tag{4}$$

where Z is a scheme, and p is an H-torsor.

(ii) *If H is a normal subgroup scheme of G, then q is a G/H-torsor.*

PROOF. (i) The uniqueness of the factorization (4) follows from the fact that p is a universal geometric quotient.

Also, the factorization (4) exists after base change under $\pi : X \to Y$: it is just the composite

$$G \times X \xrightarrow{r \times \mathrm{id}_X} G/H \times X \xrightarrow{p_2} X$$

where $r : G \to G/H$ is the quotient map, and p_2 the projection.

Thus, it suffices to show that the algebraic space X/H is representable by a scheme.

We first prove this assertion under the assumption that G *is a (connected) algebraic group.* We begin by reducing to the case that

X and Y are normal, quasi-projective varieties. For this, we adapt the argument of [Ra, pp. 206–207]. We may assume that $X = G \cdot U$, where $U \subset X$ is an open affine subscheme (since X is covered by open G-stable subschemes of that form). Then let $\nu : \tilde{Y} \to Y$ denote the normalization map of Y_{red}. Consider the cartesian square

$$\begin{array}{ccc} \tilde{X} & \longrightarrow & X \\ {\scriptstyle \tilde{\pi}}\big\downarrow & & {\scriptstyle \pi}\big\downarrow \\ \tilde{Y} & \xrightarrow{\nu} & Y \end{array}$$

and let $\tilde{U} := U \times_Y \tilde{Y}$. Then $\tilde{\pi}$ is a G-torsor, and hence $\tilde{X}$ is normal. Moreover, $\tilde{X} = G \cdot \tilde{U}$ contains $\tilde{U}$ as an open affine subset. Hence $\tilde{\pi}$ is quasi-projective by [Ra, Théorème VI 2.3]. Therefore, to show that X/H is a scheme, it suffices to check that $\tilde{X}/H$ is a scheme in view of [loc. cit., Lemme XI.3.2]. Thus, we may assume that X is normal and π is quasi-projective. Then we may further assume that Y is quasi-projective, and hence so is X. Now X is the disjoint union of its irreducible components, and each of them is G-stable; thus, we may assume that X is irreducible. This yields the desired reduction.

Thus, we may assume that there exists an ample invertible sheaf L on X; since X is normal, we may assume that L is G_{aff}-linearized. In view of [Br1, Lemma 3.2], it follows that there exists a G-morphism $X \to G/G_1$, where $G_1 \subset G$ is a subgroup scheme containing G_{aff} and such that G_1/G_{aff} is finite. By Lemma 3.2, this yields a G-isomorphism

$$X \cong G \times^{G_1} X_1$$

where $X_1 \subset X$ is a closed subscheme, stable under G_1. Moreover, the restriction $\pi_1 : X_1 \to Y$ is a G_1-torsor. Since G_1 is affine, so is the morphism π_1 and hence X_1 is quasi-projective.

We now show that π_1 factors as a G_{aff}-torsor $p_1 : X_1 \to X_1/G_{\mathrm{aff}}$, where X_1/G_{aff} is a quasi-projective scheme, followed by a G_1/G_{aff}-torsor $q_1 : X_1/G_{\mathrm{aff}} \to Y$. Indeed, the associated fiber bundle $X_1 \times^{G_1} G_1/G_{\mathrm{aff}}$ is a quasi-projective scheme, since G_1/G_{aff} is affine; we then take for p_1 the composite of the morphism $\mathrm{id}_{X_1} \times e_{G_1} : X_1 \to X_1 \times G_1$ with the natural morphism $X_1 \times G_1 \to X_1 \times^{G_1} G_1/G_{\mathrm{aff}}$. Then p_1 is

G_{aff}-invariant and fits into a commutative diagram

$$\begin{array}{ccccc} X_1 \times G_1 & \longrightarrow & X_1 \times G_1/G_{\mathrm{aff}} & \longrightarrow & X_1 \\ \downarrow & & \downarrow & & \downarrow \\ X_1 & \xrightarrow{p_1} & X_1 \times^{G_1} G_1/G_{\mathrm{aff}} & \xrightarrow{q_1} & X_1/G_1 = Y \end{array}$$

where the top horizontal arrows are the natural projections, and the vertical arrows are G_1-torsors; thus, p_1 is a G_{aff}-torsor.

Next, note that the smooth, quasi-projective G_{aff}-variety G admits a G_{aff}-linearized ample invertible sheaf. By the preceding step and [MFK, Proposition 7.1], it follows that $G \times^{G_{\mathrm{aff}}} X_1$ is a quasi-projective scheme; it is the total space of a G_1/G_{aff}-torsor over $X = G \times^{G_1} X_1$. Likewise, $G/H \times^{G_{\mathrm{aff}}} X_1$ is a quasi-projective scheme, the total space of a G_1/G_{aff}-torsor over

$$(G/H \times^{G_{\mathrm{aff}}} X_1)/(G_1/G_{\mathrm{aff}}) =: Z$$

It follows that $Z = G/H \times^{G_1} X_1$ fits into a cartesian square

$$\begin{array}{ccc} G \times X_1 & \xrightarrow{r \times \mathrm{id}_{X_1}} & G/H \times X_1 \\ \downarrow & & \downarrow \\ X & \xrightarrow{p} & Z \end{array}$$

where the vertical arrows are G_1-torsors; therefore, p is an H-torsor.

Finally, in the *general case*, we may assume that k has characteristic $p > 0$. For any positive integer n, we then have the n-th Frobenius morphism

$$F^n_G : G \longrightarrow G^{(n)}.$$

Its kernel G_n is a finite local subgroup scheme of G. Likewise, we have the n-th Frobenius morphism

$$F^n_X : X \longrightarrow X^{(n)}$$

and $G^{(n)}$ acts on $X^{(n)}$ compatibly with the G-action on X. In particular, F^n_X is invariant under G_n. Since the morphism F^n_X is finite, the sheaf of $\mathcal{O}_{X^{(n)}}$-algebras $\big((F^n_X)_*\mathcal{O}_X\big)^{G_n}$ is of finite type. Thus, the scheme

$$X/G_n := \mathrm{Spec}_{X^{(n)}} \big((F^n_X)_*\mathcal{O}_X\big)^{G_n}$$

is of finite type, and F^n_X is the composite of the natural morphisms $X \to X/G_n \to X^{(n)}$. Clearly, the formation of X/G_n commutes with faithfully flat base change; thus, the morphism

$$\pi_n : X \longrightarrow X/G_n$$

is a G_n-torsor, since this holds for the trivial G-torsor $G \times Y \to Y$. As a consequence, π factors through π_n, the G-action on X descends to an action of $G/G_n \cong G^{(n)}$ on X/G_n, and the map $X/G_n \to Y$ is a $G^{(n)}$-torsor. Note that $G^{(n)}$ is reduced, and hence a connected algebraic group, for $n \gg 0$.

Now consider the restriction

$$F^n_H : H \longrightarrow H^{(n)}$$

with kernel $H_n = H \cap G_n$. Then H acts on X/G_n via its quotient $H/H_n \cong H^{(n)} \subset G^{n)}$. By the preceding step, there exists an $H^{(n)}$-torsor $X/G_n \to (X/G_n)/H^{(n)} = X/G_nH$, and hence a G_nH-torsor

$$p_n : X \longrightarrow X/G_nH$$

where X/G_nH is a scheme (of finite type). We now set

$$Z := \mathrm{Spec}_{X/G_nH}\big((p_n)_*\mathcal{O}_X\big)^H$$

so that p_n factors through a morphism $p : X \to Z$. Then p is an H-torsor, since the formations of X/G_nH and Z commute with faithfully flat base change, and p is just the natural map $G \times Y \to G/H \times Y$ when π is the trivial torsor over Y. Likewise, the morphism $Z \to X/G_nH$ is finite, and hence the scheme Z is of finite type.

(ii) The composite map

$$G \times X \xrightarrow{\ \alpha\ } X \xrightarrow{\ p\ } X/H$$

is invariant under the action of $H \times H$ on $G \times X$ via $(h_1, h_2) \cdot (g, x) = (gh_1^{-1}, h_2 \cdot x)$. This yields a morphism $\beta : G/H \times X/H \to X/H$ which is readily seen to be an action. $\square$

Corollary 3.4. *Let again G be a connected group scheme.*

(i) Given two G-torsors $\pi_1 : X_1 \to Y_1$ and $\pi_2 : X_2 \to Y_2$, the associated torsor $X_1 \times X_2 \to X_1 \times^G X_2$ exists.

(ii) Given a homomorphism of group schemes $f : G \to G'$ and a G-torsor $\pi : X \to Y$, the G'-torsor $\pi' : G' \times^G X \to Y$ (obtained by extension of structure groups) exists.

Proof. (i) Apply Theorem 3.3 to the $G \times G$-torsor $X_1 \times X_2 \to Y_1 \times Y_2$ and to the diagonal embedding of G into $G \times G$.

(ii) Denote by $\bar{G}$ the (scheme-theoretic) image of f and by $p : G' \to G'/\bar{G}$ the quotient morphism. Then $p \times \pi : G' \times X \to G'/\bar{G} \times Y$ is a $\bar{G} \times G$-torsor. Moreover, $\bar{G} \times G$ is a connected group scheme, and contains G viewed as the image of the homomorphism $f \times \mathrm{id}$. Applying Theorem 3.3 again yields a G-torsor $G' \times X \to G' \times^G X$. Moreover, the trivial G'-torsor $G' \times X \to X$ descends to a G'-torsor $G' \times^G X \to X/G = Y$. $\square$

Corollary 3.5. *Let G be a connected algebraic group. Then every G-torsor (2) factors uniquely as the composite*

$$X \xrightarrow{p} Z \xrightarrow{q} Y, \tag{5}$$

where Z is a scheme, p is a G_{aff}-torsor, and q is an $A(G)$-torsor. Here p is affine and q is proper.

Moreover, the following conditions are equivalent:

1. *π is quasi-projective.*
2. *q is projective.*
3. *q admits a reduction of structure group to a finite subgroup scheme $F \subset A(G)$.*
4. *π admits a reduction of structure group to an affine subgroup scheme $H \subset G$.*

These conditions hold if X is smooth. In characteristic 0, they imply that q is isotrivial and π is locally isotrivial.

Proof. The existence and uniqueness of the factorization are direct consequences of Theorem 3.3. The assertions on p and q follow by descent theory (see [SGA1, Exposé VIII, Corollaires 4.8, 5.6]).

(1)⇒(2) is a consequence of [Ra, Lemme XIV 1.5 (ii)].

(2)⇒(1) holds since p is affine.

(2)⇒(3) follows from [Br1, Lemma 3.2].

(3)⇒(4) Let $H \subset G$ be the preimage of F. Then $G/H \cong A(G)/F$. By assumption, X/G_{aff} admits an $A(G)$-morphism to $A(G)/F$; this yields a G-morphism $X \to G/H$.

(4)⇒(3) Since $G_{\mathrm{aff}}H$ is affine (as a quotient of the affine group scheme $G_{\mathrm{aff}} \times H$), we may replace H with $G_{\mathrm{aff}}H$. Thus, we may assume that H is the preimage of a finite subgroup scheme $F \subset A(G)$. Then q admits a reduction of structure group to $A(G)/F$.

If X is smooth, then so are Y and Z; in that case, (3) follows from [Ro, Theorem 14] or alternatively from [Ra, Proposition XIII 2.6].

Also, (3) means that $Y \cong A(G) \times^F Z'$ as $A(G)$-torsors over $Z \cong Z'/F$, where Z' is a closed F-stable subscheme of Y. This yields a cartesian square

$$\begin{array}{ccc} A(G) \times Z' & \xrightarrow{p_2} & Z' \\ \downarrow & & \downarrow \\ Y & \xrightarrow{q} & Z \end{array}$$

where the vertical arrows are F-torsors, and hence étale in characteristic 0. This shows the isotriviality of q. Since p is locally isotrivial, so is π. □

Remarks 3.6. (i) The equivalent conditions in the preceding result do not generally hold in the setting of normal varieties. Specifically, given an elliptic curve G, there exists a G-torsor $\pi : X \to Y$ where Y is a normal affine surface and X is not quasi-projective; then of course π is not projective (see [Br1, Example 6.4], adapted from [Ra, XIII 3.2]).

(ii) If (4) holds, one may ask whether π admits a reduction of structure group to some affine algebraic subgroup $H \subset G$. The answer is trivially positive in characteristic 0, but negative in characteristic $p > 0$, as shown by the following example.

Choose an integer $n \geq 2$ not divisible by p, and let C denote the curve of equation $y^p = x^n - 1$ in the affine plane $\mathbb{A}^2$, minus all points $(x, 0)$ where x is a n-th root of unity. The group scheme μ_p of p-th roots of unity acts on $\mathbb{A}^2$ via $t \cdot (x, y) = (x, ty)$, and this action leaves C stable. The morphism $\mathbb{A}^2 \to \mathbb{A}^2$, $(x, y) \mapsto (x, y^p)$ restricts to a

μ_p-torsor

$$q : C \longrightarrow Y$$

where $Y \subset \mathbb{A}^2$ denotes the curve of equation $y = x^n - 1$ minus all points $(x, 0)$ with $x^n = 1$. Note that Y is smooth, whereas C is singular; both curves are rational, since the equation of C may be rewritten as $x^n = (y+1)^p$.

Next, let G be an ordinary elliptic curve, so that G contains μ_p, and denote by

$$\pi : X = G \times^{\mu_p} C \longrightarrow Y$$

the G-torsor obtained by extension of structure group (which exists since C is affine). Then X is a smooth surface.

We show that there exists no G-morphism $f : X \to G/H$, where H is an affine algebraic (or equivalently, finite) subgroup of G. Indeed, f would map the rational curve $C \subset X$ and all its translates by G to points of the elliptic curve G/H, and hence f would factor through a G-morphism $G/\mu_p \to G/H$. As a consequence, $\mu_p \subset H$, a contradiction.

(iii) Given a torsor (2) under a group scheme (of finite type) G, there exists a unique factorization

$$X \xrightarrow{\ p\ } Z = X/G^o_{\mathrm{red}} \xrightarrow{\ q\ } Y \tag{6}$$

where Z is a scheme, p is a torsor under the connected algebraic group G^o_{red}, and q is finite. (Indeed, $Z = X \times^G G/G^o_{\mathrm{red}}$ as in the proof of Theorem 3.3).

4. Automorphism groups of torsors

To any G-torsor $\pi : X \to Y$ as in Section 3, one associates several groups of automorphisms:

- the automorphism group of X as a scheme over Y, denoted by $\mathrm{Aut}_Y(X)$ and called the relative automorphism group,
- the automorphism group of the pair (X, Y), denoted by $\mathrm{Aut}(X, Y)$: it consists of those pairs $(\varphi, \psi) \in \mathrm{Aut}(X) \times \mathrm{Aut}(Y)$ such that

the square

$$\begin{array}{ccc} X & \xrightarrow{\varphi} & X \\ {\scriptstyle\pi}\Big\downarrow & & {\scriptstyle\pi}\Big\downarrow \\ Y & \xrightarrow{\psi} & Y \end{array}$$

commutes,

- the automorphism group of X viewed as a G-scheme, denoted by $\mathrm{Aut}^G(X)$ and called the equivariant automorphism group.

Clearly, the projection $p_2 : \mathrm{Aut}(X) \times \mathrm{Aut}(Y) \to \mathrm{Aut}(Y)$ yields an exact sequence of (abstract) groups

$$1 \longrightarrow \mathrm{Aut}_Y(X) \longrightarrow \mathrm{Aut}(X,Y) \xrightarrow{p_2} \mathrm{Aut}(Y).$$

Also, note that each G-morphism $\varphi : X \to X$ descends to an morphism $\psi : Y \to Y$, since $\pi \circ \varphi : X \to Y$ is G-invariant and π is a categorical quotient. The assignement $\varphi \in \mathrm{Aut}^G(X) \mapsto \psi =: \pi_*(\varphi) \in \mathrm{Aut}(Y)$ yields an identification of $\mathrm{Aut}^G(X)$ with a subgroup of $\mathrm{Aut}(X,Y)$, and an exact sequence of groups

$$1 \longrightarrow \mathrm{Aut}^G_Y(X) \longrightarrow \mathrm{Aut}^G(X) \xrightarrow{\pi_*} \mathrm{Aut}(Y). \tag{7}$$

Moreover, we may view the equivariant automorphisms as those pairs (ϕ, ψ) where $\psi \in \mathrm{Aut}(Y)$, and $\phi : X \to X_\psi$ is a G-morphism. Here X_ψ denotes the G-torsor over Y obtained by pull-back under ψ; note that ϕ is an isomorphism, as a morphism of G-torsors over the same base.

The relative automorphism group is described by the following result, which is certainly well-known but for which we could not locate any appropriate reference:

Lemma 4.1. *Let $\pi : X \to Y$ be a G-torsor. Then the map*

$$\mathrm{Hom}(X,G) \longrightarrow \mathrm{Aut}_Y(X), \quad (f : X \to G) \longmapsto \big(F : X \to X, \quad x \mapsto f(x)\cdot x\big)$$

is an isomorphism of groups, which restricts to an isomorphism

$$\mathrm{Hom}^G(X,G) \cong \mathrm{Aut}^G_Y(X). \tag{8}$$

Here $\mathrm{Hom}^G(X,G) \subset \mathrm{Hom}(X,G)$ *denotes the subset of morphisms that are equivariant for the given* G*-action on* X*, and the* $G-$*action on itself by conjugation.*

If G *is commutative, then* $\mathrm{Aut}^G_Y(X) \cong \mathrm{Hom}(Y,G)$.

Proof. Let $u \in \mathrm{Aut}_Y(X)$. Then $u \times \mathrm{id}_X$ is an automorphism of $X \times_Y X$ over X. In view of the isomorphism (3), $u \times \mathrm{id}_X$ yields an automorphism of $G \times X$ over X, thus of the form $(g,x) \mapsto (F(g,x),x)$ for a unique $F \in \mathrm{Hom}(G \times X, G)$. In other words, $u(g \cdot x) = F(g,x) \cdot x$. Thus, $u(x) = f(x) \cdot x$, where $f := F(e_G, -) \in \mathrm{Hom}(X,G)$. This yields the claimed isomorphism $\mathrm{Hom}(X,G) \cong \mathrm{Aut}_Y(X)$, equivariant for the action of G on $\mathrm{Hom}(X,G)$ via $(g \cdot f)(x) = gf(g^{-1} \cdot x)g^{-1}$ and on $\mathrm{Aut}_Y(X)$ by conjugation. Taking invariants, we obtain the isomorphism (8).

If G is commutative, then $\mathrm{Hom}^G(X,G)$ consists of the G-invariant morphisms $X \to G$; these are identified with the morphisms $Y = X/G \to G$. $\square$

The preceding considerations adapt to group functors of automorphisms, that associate to any scheme S the groups $\mathrm{Aut}_{Y \times S}(X \times S)$, $\mathrm{Aut}_S(X \times S, Y \times S)$ and their equivariant analogues. We will denote these functors by $Aut_Y(X)$, $Aut(X,Y)$, $Aut^G(X)$ and $Aut^G_Y(X)$. The exact sequence (7) readily yields an exact sequence of group functors

$$1 \longrightarrow Aut^G_Y(X) \longrightarrow Aut^G(X) \xrightarrow{\pi_*} Aut(Y). \tag{9}$$

Also, by Lemma 4.1, we have a functorial isomorphism

$$\mathrm{Aut}_{Y \times S}(X \times S) \cong \mathrm{Hom}(X \times S, G).$$

In other words, $Aut_Y(X)$ is isomorphic to the group functor

$$Hom(X,G) : S \longmapsto \mathrm{Hom}(X \times S, G).$$

As a consequence, $Aut^G_Y(X)$ is isomorphic to $Hom^G(X,G) : S \mapsto \mathrm{Hom}^G(X \times S, G)$. This readily yields isomorphisms

$$\mathrm{Lie}\,\mathrm{Aut}_Y(X) \cong \mathrm{Hom}\big(X, \mathrm{Lie}(G)\big) \cong \mathcal{O}(X) \otimes \mathrm{Lie}(G),$$

$$\mathrm{Lie}\,\mathrm{Aut}^G_Y(X) \cong \mathrm{Hom}^G\big(X, \mathrm{Lie}(G)\big) \cong \big(\mathcal{O}(X) \otimes \mathrm{Lie}(G)\big)^G.$$

We now obtain a finiteness result for $Aut^G(X)$, analogous to a theorem of Morimoto (see [Mo, Théorème, p. 158]):

THEOREM 4.2. *Consider a G-torsor $\pi : X \to Y$ where G is a group scheme, X a scheme, and Y a proper scheme. Then the functor $Aut^G(X)$ is represented by a group scheme, locally of finite type, with Lie algebra $\Gamma(X, T_X)^G$.*

PROOF. The assertion on the Lie algebra follows from the G-isomorphism (1).

To show the representability assertion, we first reduce to the case that G is a connected affine algebraic group. Let G_{aff} denote the largest closed normal affine subgroup of G, or equivalently of G^o_{red}. Then $Aut^G(X)$ is a closed subfunctor of $Aut^{G_{\mathrm{aff}}}(X)$. Moreover, the factorizations (5) and (6) yield a factorization of π as

$$X \xrightarrow{p} X/G_{\mathrm{aff}} \xrightarrow{q} X/G^o_{\mathrm{red}} \xrightarrow{r} Y$$

where p is a torsor under G_{aff}, q a torsor under $G^o_{\mathrm{red}}/G_{\mathrm{aff}}$, and r is a finite morphism. Since q and r are proper, X/G_{aff} is proper as well. This yields the desired reduction.

Next, we may embed G as a closed subgroup of $\mathrm{GL}(V)$ for some finite-dimensional vector space V. Let Z denote the closure of G in the projective completion of $\mathrm{End}(V)$. Then Z is a projective variety equipped with an action of $G \times G$ (arising from the $G \times G$-action on $\mathrm{End}(V)$ via left and right multiplication) and with an ample $G \times G$-linearized invertible sheaf. By construction, G (viewed as a $G \times G$-variety via left and right multiplication) is the open dense $G \times G$-orbit in Z.

As seen in Section 3, the associated fiber bundle $X \times^G Z$ (for the left G-action on Z) exists; it is equipped with a G-action arising from the right G-action on Z. Moreover, $X \times^G Z$ contains $X \times^G G \cong X$ as a dense open G-stable subscheme. Also, recall the cartesian square

$$\begin{array}{ccc} X \times Z & \xrightarrow{p} & X \\ {\scriptstyle\varpi}\downarrow & & {\scriptstyle\pi}\downarrow \\ X \times^G Z & \xrightarrow{q} & Y. \end{array}$$

Since Z is complete and π is faithfully flat, it follows that q is proper, and hence so is $X \times^G Z$.

Now let S be a scheme, and $\varphi \in \mathrm{Aut}_S^G(X \times S)$. Then φ yields an S-automorphism

$$\phi : X \times Z \times S \longrightarrow X \times Z \times S, \quad (x, z, s) \longmapsto \big(\varphi(x, s), z, s\big).$$

Consider the action of $G \times G$ on $X \times Z \times S$ given by

$$(g_1, g_2) \cdot (x, z, s) = \big(g_1 \cdot x, (g_1, g_2) \cdot z, s\big).$$

Then ϕ is $G \times G$-equivariant, and hence yields an automorphism $\Phi \in \mathrm{Aut}_S^G(X \times^G Z \times S)$ which stabilizes $X \times^G (Z \setminus G) \times S$. Moreover, the assignement $\varphi \mapsto \Phi$ identifies $\mathrm{Aut}_S^G(X \times S)$ with the stabilizer of $X \times^G (Z \setminus G) \times S$ in $\mathrm{Aut}_S^G(X \times^G Z \times S)$. Thereby, $Aut^G(X)$ is identified with a closed subfunctor of $Aut(X \times^G Z)$; the latter is represented by a group scheme of finite type, since $X \times^G Z$ is proper. $\square$

For simplicity, we denote by $\mathrm{Aut}^G(X)$ the group scheme defined in the preceding theorem. Since $Aut_Y^G(X)$ is a closed subfunctor of $Aut^G(X)$, it is also represented by a group scheme (locally of finite type) that we denote likewise by $\mathrm{Aut}_Y^G(X)$. Further properties of this relative automorphism group scheme are gathered in the following:

Proposition 4.3. *Let $\pi : X \to Y$ be a torsor under a connected algebraic group G, where Y is a proper scheme. Then the factorization $X \xrightarrow{p} Z = X/G_{\mathrm{aff}} \xrightarrow{q} Y$ (obtained in Corollary 3.5) yields an exact sequence of group schemes*

$$(10) \qquad 1 \longrightarrow \mathrm{Aut}_Z^{G_{\mathrm{aff}}}(X) \longrightarrow \mathrm{Aut}_Y^G(X) \xrightarrow{p_*} \mathrm{Aut}_Y^{A(G)}(Z).$$

Moreover, $\mathrm{Aut}_Z^{G_{\mathrm{aff}}}(X)$ is affine of finite type,

If Y is a (complete) variety, then the neutral component of $\mathrm{Aut}_Y^{A(G)}(Z)$ is just $A(G)$; it is contained in the image of p_*.

Proof. We first show that $\mathrm{Aut}_Z^{G_{\mathrm{aff}}}(X)$ is affine of finite type. By Lemma 4.1, we have

$$Aut_Z^{G_{\mathrm{aff}}}(X) \cong Hom^{G_{\mathrm{aff}}}(X, G_{\mathrm{aff}}).$$

Moreover, there exists a closed G_{aff}-equivariant immersion of G_{aff} into an affine space V where G_{aff} acts via a representation. Thus,

$Aut_Z^{G_{\mathrm{aff}}}(X)$ is a closed subfunctor of $Hom^{G_{\mathrm{aff}}}(X,V)$. But the latter is represented by an affine space (of finite dimension), namely, the space of global sections of the associated vector bundle $X \times^{G_{\mathrm{aff}}} V$ over the proper scheme $X/G_{\mathrm{aff}} = Z$. This completes the proof.

Next, we obtain (10). We start with the exact sequence (9) for the G_{aff}-torsor p, which translates into an exact sequence of group schemes

$$1 \longrightarrow \mathrm{Aut}_Z^{G_{\mathrm{aff}}}(X) \longrightarrow \mathrm{Aut}^{G_{\mathrm{aff}}}(X) \xrightarrow{p_*} \mathrm{Aut}(Z).$$

Taking G-invariants yields the exact sequence of group schemes

$$1 \longrightarrow \mathrm{Aut}_Z^{G}(X) \longrightarrow \mathrm{Aut}_Y^{G}(X) \xrightarrow{p_*} \mathrm{Aut}(Z).$$

But G acts on the affine scheme $\mathrm{Aut}_Z^{G_{\mathrm{aff}}}(X)$ through its quotient $G/G_{\mathrm{aff}} = A(G)$, an abelian variety. So this G-action must be trivial, that is, $\mathrm{Aut}_Z^G(X) = \mathrm{Aut}_Z^{G_{\mathrm{aff}}}(X)$.

We now show that $A(G) = \mathrm{Aut}_Y^{A(G),o}(Z)$ if Y (or equivalently Z) is a variety. Since $A(G)$ is commutative, we have a homorphism $f : A(G) \to \mathrm{Aut}_Y^{A(G)}(Z)$. The induced homomorphism of Lie algebras is the natural map

$$\mathrm{Lie}\, A(G) \longrightarrow \mathrm{Lie}\, \mathrm{Aut}_Y^{A(G)}(Z) = \big(\mathcal{O}(Z) \otimes \mathrm{Lie}\, A(G)\big)^{A(G)}$$

which is an isomorphism since $\mathcal{O}(Z) = k$. This yields our assertion.

Finally, we show that $A(G)$ is contained in the image of p_*. Indeed, the neutral component of the center of G is identified with a subgroup of $\mathrm{Aut}_Y^G(X)$, and is mapped onto $A(G)$ under the quotient homomorphism $G \to G/G_{\mathrm{aff}}$ (as follows from [Ro, Corollary 5, p. 440]). □

Observe that the exact sequence (10) yields an analogue for torsors of Chevalley's structure theorem; it gives back that theorem when applied to the trivial torsor G.

5. Lifting automorphisms for abelian torsors

We begin by determining the relative equivariant automorphism groups of torsors under abelian varieties:

PROPOSITION 5.1. *Let G be an abelian variety and $\pi : X \to Y$ a G-torsor, where X and Y are complete varieties. Then the group scheme $Aut_Y^G(X)$ is isomorphic to* $\mathrm{Hom}_{\mathrm{gp}}\big(A(Y), G\big) \times G$. *Here $A(Y)$ denotes the Albanese variety of Y, and* $\mathrm{Hom}_{\mathrm{gp}}\big(A(Y), G\big)$ *denotes the space of homomorphisms of algebraic groups $A(Y) \to G$; this is a free abelian group of finite rank, viewed as a constant group scheme.*

PROOF. By Lemma 4.1, we have a functorial isomorphism

$$\mathrm{Aut}_{Y\times S}^G(X \times S) \cong \mathrm{Hom}(Y \times S, G).$$

Choose a point $y_0 \in Y$. For any $f \in \mathrm{Hom}(Y \times S, G)$, consider the morphism

$$\varphi : Y \times S \longrightarrow G, \quad (y, s) \longmapsto f(y, s) - f(y_0, s)$$

where the group law of the abelian variety G is denoted additively. We claim that φ factors through the projection $Y \times S \to Y$. For this, we may replace k with a larger field, and assume that S has a k-rational point s_0; we may also assume that S is connected. Then the morphism

$$\psi : Y \times S \longrightarrow G, \quad (y, s) \longmapsto f(y, s) - f(y, s_0)$$

maps $Y \times \{s_0\}$ to a point. By a scheme-theoretic version of the rigidity lemma (see [SS, Theorem 1.7]), it follows that ψ factors through the projection $Y \times S \to S$. Thus, $f(y, s) - f(y, s_0) = f(y_0, s) - f(y_0, s_0)$ which shows the claim.

By that claim, we may write

$$f(y, s) = \varphi(y) + \psi(s)$$

where $\varphi : Y \to G$ and $\psi : S \to G$ are morphisms such that $\varphi(y_0) = 0$. Now let $a : Y \to A(Y)$ be the Albanese morphism, normalized so that $a(y_0) = 0$. Then φ factors through a unique homorphism $\Phi : A(Y) \to A$, and $f = (\Phi \circ a) + \psi$ where Φ is an S-point of $\mathrm{Hom}_{\mathrm{gp}}\big(A(Y), G\big)$, and ψ an S-point of G. $\square$

Next, we obtain a preliminary result which again is certainly well-known, but for which we could not locate any reference:

Lemma 5.2. *Assume that k has characteristic 0. Let $\pi : Z \to Y$ be a finite étale morphism, where Y and Z are complete varieties.*

Then the natural homomorphism $\pi_* : \mathrm{Aut}(Z, Y) \to \mathrm{Aut}(Y)$ restricts to an isogeny $\mathrm{Aut}^o(Z, Y) \to \mathrm{Aut}^o(Y)$ on neutral components.

If π is a Galois cover with group F (that is, an F-torsor), then $\mathrm{Aut}^o(Z, Y)$ is the neutral component of $\mathrm{Aut}^F(Y)$.

Proof. We set for simplicity $H := \mathrm{Aut}^o(Y)$; this is a connected algebraic group in view of the characteristic-0 assumption. For any $h \in H(k)$, denote by Z_h the étale cover of Y obtained from Z by pull-back under h. Then the covers Z_h, $h \in H(k)$, are all isomorphic by [SGA1, Exposé X, Corollaire 1.9]. Thus, every $h \in H(k)$ lifts to some $\tilde{h} \in \mathrm{Aut}(Z)(k)$. In other words, the image of the projection $\pi_* : \mathrm{Aut}(Z, Y) \to \mathrm{Aut}(Y)$ contains H. It follows that π_* restricts to a surjective homomorphism $\mathrm{Aut}^o(Z, Y) \to \mathrm{Aut}^o(Y)$; its kernel is finite by Galois theory.

If π is an F-torsor, then π_* has kernel F, by Galois theory again. In particular, $\mathrm{Aut}(Z, Y)$ normalizes F. The action of the neutral component $\mathrm{Aut}^o(Z, Y)$ by conjugation on the finite group F must be trivial; this yields the second assertion. □

Remark 5.3. In particular, with the notation and assumptions of the preceding lemma, all elements of $\mathrm{Aut}^o(Y)$ lift to automorphisms of Z. But this does not generally hold for elements of $\mathrm{Aut}(Y)$. For a very simple example, take $k = \mathbb{C}$, Z the elliptic curve $\mathbb{C}/2\mathbb{Z} + i\mathbb{Z}$, Y the elliptic curve $\mathbb{C}/\mathbb{Z} + i\mathbb{Z}$, and π the natural morphism. Then the multiplication by i defines an automorphism of Y which admits no lift under the double cover π.

We now come to the main result of this section:

Theorem 5.4. *Let G be an abelian variety and $\pi : X \to Y$ a G-torsor, where X and Y are complete varieties. Then G centralizes $\mathrm{Aut}^o(X)$; equivalently, $\mathrm{Aut}^o(X) = \mathrm{Aut}^{G,o}(X)$. Moreover, there exists a closed subgroup $H \subset \mathrm{Aut}^o(X)$ such that $\mathrm{Aut}^o(X) = GH$ and $G \cap H$ is finite.*

If k has characteristic 0 and X (or equivalently Y) is smooth, then the homomorphism $\pi_* : \mathrm{Aut}^G(X) \longrightarrow \mathrm{Aut}(Y)$ restricts to an isogeny $\pi_{*|H} : H \to \mathrm{Aut}^o(Y)$ for any quasi-complement H as above.

Proof. The assertion that G is central in $\mathrm{Aut}^o(X)$ and admits a quasi-complement follows from [Ro, Corollary, p. 434].

By Proposition 4.3 or alternatively Proposition 5.1, G is the neutral component of the kernel of π_*. Thus, the kernel of $\pi_{*|H}$ is finite.

It remains to show that $\pi_{*|H}$ is surjective when k has characteristic 0 and X is smooth. By Lemma 3.2 and Corollary 3.5, we have a G-isomorphism

$$X \cong G \times^F Z \tag{11}$$

for some finite subgroup $F \subset G$ and some closed F-stable subscheme $Z \subset X$ such that $\pi : Z \to Y$ is an F-torsor. Thus, Z is smooth and complete. Replacing Z with a component, and F with the stabiliser of that component, we may assume that F is a variety. Then by Lemma 5.2, the natural homomorphism $\mathrm{Aut}^{F,o}(Z) \to \mathrm{Aut}^o(Y)$ is surjective.

We now claim that $\mathrm{Aut}^F(Z)$ may be identified with a closed subgroup of $\mathrm{Aut}^G(X)$. Indeed, as in the proof of Theorem 4.2, any $\varphi \in \mathrm{Aut}^F(X)$ yields a morphism

$$\phi : G \times Z \longrightarrow G \times Z, \quad (g, z) \longmapsto \big(g, \varphi(z)\big).$$

This is a $G \times F$-automorphism of $X \times Z$, and hence descends to a G-automorphism Φ of X. The assignement $\varphi \mapsto \Phi$ yields the desired identification. This proves the claim and, in turn, the surjectivity of $\pi_{*|H}$. $\square$

Remarks 5.5. (i) With the notation and assumptions of the preceding theorem, the surjectivity of $\pi_{*|H}$ also holds when X (or equivalently Y) is normal. Choose indeed an $\mathrm{Aut}^o(Y)$-equivariant desingularization

$$f : Y' \longrightarrow Y,$$

that is, f is proper and birational, and the action of $\mathrm{Aut}^o(Y)$ on Y lifts to an action on Y' such that f is equivariant (see [EV] for the existence of such desingularizations). Since Y is normal, we have

$f_*(\mathcal{O}_{Y'}) = \mathcal{O}_Y$. In view of Proposition 2.1, this yields a homomorphism

$$f_* : \mathrm{Aut}^o(Y') \longrightarrow \mathrm{Aut}^o(Y)$$

which is injective (on closed points) as f is birational, and surjective by construction. Thus, f_* is an isomorphism. Likewise, the natural map $\mathrm{Aut}^o(X') \to \mathrm{Aut}^o(X)$ is an isomorphism, where $X' := X \times_Y Y'$ is the total space of the pull-back torsor $\pi' : X' \to Y'$. Now the desired surjectivity follows from Theorem 5.4.

We do not know whether $\pi_{*|H}$ is surjective for arbitrary (complete) varieties X, Y. Also, we do not know whether the characteristic-0 assumption can be omitted.

(ii) The preceding theorem may be reformulated in terms of vector fields only: let X, Y be smooth complete varieties over an algebraically closed field of characteristic 0, and $\pi : X \to Y$ a smooth morphism such that the relative tangent bundle T_π is trivial. Then every global vector on Y lifts to a global vector field on X.

Consider indeed the Stein factorization of π,

$$X \xrightarrow{\pi'} X' \xrightarrow{p} Y.$$

Then one easily checks that p is étale; thus, X' is smooth and π' is smooth with trivial relative tangent bundle. Also, every global vector field on Y lifts to a global vector field on X', as follows e.g. from Lemma 5.2. Thus, we may replace π with π', and hence assume that the fibers of π are connected. Then these fibers are just the orbits of $G := \mathrm{Aut}^o_Y(X)$, an abelian variety. Moreover, for F and Z as in (11), the restriction $\pi_{|Z}$ is smooth, since so is π. Thus, $\pi_{|Z}$ is an F-torsor. So the claim follows again from Theorem 5.4.

Finally, using the factorization (6) and combining Lemma 5.2 and Theorem 5.4, we obtain the following:

Corollary 5.6. *Let G be a proper algebraic group and $\pi : X \to Y$ a G-torsor, where Y is a complete variety over an algebraically closed field of characteristic* 0. *Then there exists a closed connected subgroup $H \subset \mathrm{Aut}^G(X)$ which is isogenous to* $\mathrm{Aut}^o(Y)$ *via* $\pi_* : \mathrm{Aut}^G(X) \to \mathrm{Aut}(Y)$.

Here the assumption that G is proper cannot be omitted. For example, let Y be an abelian variety, so that $\mathrm{Aut}^o(Y)$ is the group of translations. Let also G be the multiplicative group $\mathbb{G}_m$, so that G-torsors $\pi : X \to Y$ correspond bijectively to invertible sheaves $\mathcal{L}$ on Y. Then $\mathrm{Aut}^o(Y)$ lifts to an isomorphic (resp. isogenous) subgroup of $\mathrm{Aut}^G(X)$ if and only if $\mathcal{L}$ is trivial (resp. of finite order). Also, the image of π_* contains $\mathrm{Aut}^o(Y)$ if and only if $\mathcal{L}$ is algebraically trivial (see [Mu] for these results).

This is the starting point of the theory of homogeneous bundles over abelian varieties, to be developed in [Br2].

References

[Ak] D. N. Akhiezer, *Lie group actions in complex analysis*, Aspects of Mathematics **E 27**, Vieweg, Braunschweig/Wiesbaden, 1995.

[Bi] A. Bialynicki-Birula, *On induced actions of algebraic groups*, Ann. Inst. Fourier (Grenoble) **43** (1993), no. 2, 365–368.

[Bl] A. Blanchard, *Sur les variétés analytiques complexes*, Ann. Sci. École Norm. Sup. (3) **73** (1956), 157–202.

[Br1] M. Brion, *Some basic results on actions of non-affine algebraic groups*, in: Symmetry and Spaces (in honor of Gerry Schwarz), 1–20, Progr. Math. **278**, Birkhäuser, Boston, MA, 2009.

[Br2] M. Brion, *Homogeneous bundles over abelian varieties*, in preparation.

[Co] B. Conrad, *A modern proof of Chevalley's theorem on algebraic groups*, J. Ramanujam Math. Soc. **17** (2002), 1–18.

[DG] M. Demazure, P. Gabriel, *Groupes algébriques*, Masson, Paris, 1970.

[EV] S. Encinas, O. Villamayor, *A course on constructive desingularization and equivariance*, in: Resolution of singularities (Obergurgl, 1997), 147–227, Progr. Math. **181**, Birkhäuser, Basel, 2000.

[KM] S. Keel, S. Mori, *Quotients by groupoids*, Ann. of Math. (2) **145** (1997), no. 1, 193–213.

[Ma] H. Matsumura, *On algebraic groups of birational transformations*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8) **34** (1963), 151–155.

[MO] H. Matsumura, F. Oort, *Representability of group functors, and automorphisms of algebraic schemes*, Invent. Math. **4** (1967), 1–25.

[Mo] A. Morimoto, *Sur le groupe d'automorphismes d'un espace fibré principal analytique complexe*, Nagoya Math. J. **13** (1958), 157–168.

[Mu] D. Mumford, *Abelian Varieties*, Oxford University Press, Oxford, 1970.

[MFK] D. Mumford, J. Fogarty, F. Kirwan, *Geometric Invariant Theory. Third Edition*, Ergeb. der Math., Springer, 1994.

[Ra] M. Raynaud, *Faisceaux amples sur les schémas en groupes et les espaces homogènes*, Lecture Note Math. **119**, Springer-Verlag, New York, 1970.

[Ro] M. Rosenlicht, *Some basic theorems on algebraic groups*, Amer. J. Math. **78** (1956), 401–443.

[SGA1] *Revêtements étales et groupe fondamental*, Séminaire de géométrie algébrique du Bois Marie 1960–61 dirigé par A. Grothendieck, Documents Mathématiques **3**, Soc. Math. France, 2003.

[SS] C. Sancho de Salas, F. Sancho de Salas, *Principal bundles, quasi-abelian varieties and structure of algebraic groups*, J. Algebra **322** (2009), 2751–2772.

Université de Grenoble I,, Institut Fourier, CNRS UMR 5582, B.P. 74, 38402 Saint-Martin d'Hères Cedex, France

E-mail address: `Michel.Brion@ujf-grenoble.fr`

ON THE FOCUSING OF CRAMÉR - VON MISES TEST.

ALEJANDRA CABAÑA AND ENRIQUE M. CABAÑA

Abstract. The statistical bibliography frequently refers to *omnibus tests* intended to be sensitive to all or at least a wide variety of alternatives, and *focused or directional tests* directed to detect efficiently some specific alternatives.

In fact, the apparent opposition between omnibus and focused is artificial, and, for instance, K-S test is focused on changes in position of Double Exponential distribution, as well as Cramér - von Mises is focused on changes in position of the distribution with density $f(t) = 1/(2\cosh(\pi t/2))$.

We provide in this article a simple proof of this latter fact.

1. Introduction

In the statistical literature refering to a test as being omnibus or directional often implies opposite categories.

Omnibus tests are able to detect a wide bunch of alternatives, and no special ability to detect any particular one is intended.

When statistical practitioners wish to detect specific alternatives they can use directional tests. These ones focus their power in the direction of the interesting alternatives.

The former tests are not expected to be efficient in the detection of particular alternatives. On the other hand, it is generally claimed that the second ones have the drawback that they have a poor power against alternatives other that the ones on which they were focused.

Research partially supported by TIN2008-06582-C03-02/TIN, Ministerio de Ciencia y Tecnología.

Partially supported by CSIC-Udelar,Uruguay, Centre de Recerca Matemàatica, Barcelona, Spain and Carolina Foundation, Spain.

Notwhithstanding, it is well established that a test can be both omnibus and focused: this is the case of the well known omnibus Kolmogorov - Smirnov goodness-of-fit test, that is also focused to detect changes in position of samples of the Double - Exponential Distribution as shown by J. Capon ([3]) by computing lower bounds for the asymptotic efficiency of the test for several alternatives.

In this short note, we show that the well known Cramér - von Mises goodness-of-fit test, also reputed to be an omnibus test, is also focused to detect changes in position of random samples of another family of distributions obtained by changes in location and scale from the distribution with probability density

$$g(t) = \frac{1}{2\cosh(\pi t/2)}. \tag{1}$$

It is known (see [8]) that there is one direction with the highest asymptotic power that is possible for Cramér - von Mises test. We present here a straightforward computation of such direction.

The principal result is that the asymptotic power of the Cramér - von Mises test for those alternatives is almost optimal. This statement is made precise in §4, where the power of the test is compared with the power of the two-sided test based on the likelihood ratio.

This kind of quasi-optimal behaviour characterises several tests of goodness-of-fit developed by the authors in which a quadratic statistic of Watson type is employed in such a way that the resulting tests are consistent against any alternative, and also have a near optimum efficiency for some alternative of focusing arbitrarily selected by the user (see [1], [2] and references therein).

The tuning on the interesting alternatives is a part of the design of our tests, but the quasi-optimum efficiency is inherent to the statistic in use.

The efficiency of our tests is described in the already cited articles. But the fact that the efficiency of the classical Cramér - von Mises test share such kind of properties does not appear to us to be widely discussed in the statistical literature, and motivates this article.

The power of Cramér - von Mises test has been analysed by several authors, and is fully described by Durbin and Knott ([5]), for instance. We describe it from scratch in order to facilitate the reading, before identifying in §3 the alternatives for optimum power.

2. The Cramér - von Mises goodness-of-fit test.

The Cramér - von Mises statistic $\omega_n^2 = n \int_{-\infty}^{\infty} (F_n(t) - F_0(t))^2 dF_0(t)$ quantifies a quadratic distance between the probability distribution function F_0 and the empirical distribution function $F_n(t) = \sum_{i=1}^n \mathbf{1}_{\{X_i \le t\}}$ of the sample of i.i.d. random variables $X_1, X_2, \dots, X_n$ with probability distribution F.

By introducing the *empirical process* $b_n(t) = \sqrt{n}(F_n(t) - F_0(t))$, ω_n^2 is written as

$$\omega_n^2 = \int_{-\infty}^{\infty} b_n^2(t) dF_0(t).$$

We shall assume that F_0 is continuous, with density f_0, finite first- and second-order moments, and, with no loss of generality that $\int t dF_0(t) = 0$, $\int t^2 dF_0(t) = 1$.

Let the probability distribution of ω_n^2 be denoted by $P(t, F, n) = \mathbf{P}\{\omega_n^2 \le t\}$.

The Cramér - von Mises test of the null hypothesis $\mathcal{H}_0$: "$F = F_0$", with confidence level α, rejects $\mathcal{H}_0$ when $\omega_n^2 > c_n(\alpha)$, where $c_n(\alpha)$ solves the equation $P(c_n(\alpha), F_0, n) = 1 - \alpha$, and its power for the alternative F is $1 - P(c_n(\alpha), F, n)$.

2.1. **The asymptotic law of ω_n^2 under $\mathcal{H}_0$.** Since b_n converges in law to a brownian bridge associated to F_0, that is, to a Gaussian centred process b^{F_0} with covariances $\mathbf{E} b^{F_0}(s) b^{F_0}(t) = F_0(s \wedge t) - F_0(s)F_0(t)$, then ω_n^2 has the asymptotic law of $\int (b^{F_0}(t))^2 dF_0(t) \sim \int_0^1 b^2(u) du$, where b denotes a standard Brownian bridge, because b^{F_0} has the same law as $b \circ F_0$.

In order to obtain the distribution of $Q_0 = \int_0^1 b^2(u) du = \|b\|^2$, the L^2 squared norm of the standard Brownian bridge b in $L^2(([0,1])$ with the Lebesgue measure, let us follow Durbin ([4]) and compute the Fourier expansion

$$b(u) = \sum_{j=1}^{\infty} \left(\int_0^1 b(v) \psi_j(v) dv \right) \psi_j(u) \tag{2}$$

of b in terms of the complete orthonormal system $\{\psi_j(u) = \sqrt{2}\sin j\pi u : j = 1, 2, \dots\}$ of eigenfunctions of the covariance kernel which admits the expansion

$$\mathbf{E}b(u)b(v) = u \wedge v - uv = \sum_{j=1}^{\infty} \frac{1}{j^2\pi^2}\psi_j(u)\psi_j(v).$$

The random coefficients in (2) are independent centred Gaussian variables vith variances

$$\mathbf{E}\left(\int_0^1 b(u)\psi(u)du\right)^2 = \int_0^1\int_0^1 (u \wedge v - uv)\psi(u)\psi(v)\,du\,dv = \frac{1}{j^2\pi^2}$$

and hence we may rewrite (2) as $b(u) = \sum_{j=1}^{\infty} \frac{B_j}{j\pi}\psi_j(u)$, by introducing the i.i.d. standard Gaussian variables $B_j = j\pi\int_0^1 b(u)\psi_j(u)du$, leading us to conclude

$$Q_0 = \|b\|^2 = \sum_{j=1}^{\infty} \frac{B_j^2}{j^2\pi^2}. \tag{3}$$

2.2. **The limiting law of ω_n under sequences of contiguous alternatives.** Let us assume now that for each n, the sample has a probability law $F^{(n)}$ with density $f_n(t)$ satisfying

$$\sqrt{\frac{f_n(t)}{f_0(t)}} = 1 + \frac{\delta k_n(t)}{2\sqrt{n}}$$

for a sequence of functions k_n such that

$$\int_{-\infty}^{\infty}(k_n(t) - k(t))^2 dF_0(t) \to 0, \quad \int_{-\infty}^{\infty} k^2(t)dF_0(t) = 1.$$

When this happens, we shall say that the alternative $\mathcal{H}(k,\delta)$ holds. These alternatives are contiguous to the null hypothesis (see [9]) and therefore the asymptotic law of b_n under $\mathcal{H}(k,\delta)$ is the same one corresponding to $\mathcal{H}_0 = \mathcal{H}(k,0)$ plus a deterministic term, according to Le Cam Third Lemma ([6], [7]).

The limiting distribution of the empirical process under $\mathcal{H}(k,\delta)$, is obtained by noticing that the first term in the decomposition

$$b_n(t) = \sqrt{n}(F_n(t) - F^{(n)}(t)) + \sqrt{n}(F^{(n)}(t) - F_0(t)).$$

tends to $b^{(F_0)}(t)$, and the second one is written as

$$\sqrt{n}\int_{-\infty}^{t}(f_n(s)-f_0(s))ds = \sqrt{n}\int_{-\infty}^{t}\frac{\delta k_n(s)}{\sqrt{n}}dF_0(s) \to \delta\int_{-\infty}^{t}k(s)dF_0(s)$$

so that, with the change of variables $u = F_0(t)$ and the new function K defined by

$$K(u) = \int_0^u \kappa(v)dv, \kappa(F_0(t)) = k(t),$$

we get

(4)

$$b_n(t) \xrightarrow{\mathcal{L}} b^{(F_0)}(t)+\delta\int_{-\infty}^{t}k(s)dF_0(s) = b(u)+\delta\int_0^u \kappa(v)dv = b(u)+\delta K(u).$$

The assumptions on k imply that κ satisfies $\int_0^1 \kappa(u)du = 0$, $\int_0^1 \kappa^2(u)du = 1$, and, in particular, $K(0) = K(1) = 0$. The function κ shall be called *standardized shape* of the alternative $\mathcal{H}(k,\delta)$.

From (4), we obtain

$$\omega_n^2 \xrightarrow{\mathcal{L}} \int_0^1 (b(u) + \delta K(u))^2 du.$$

Let us notice that this expression of the limit law of ω_n leads to conclude that when the null hypothesis is replaced by $\mathcal{H}(k,\delta)$, then the asymptotic expectation of ω_n increases in the amount

$$\Delta(\delta) = \delta^2 \int_0^1 K^2(u)du. \tag{5}$$

It is reasonable to expect that larger values of $\Delta(\delta)$ be associated with larger powers of the tests comparing $\mathcal{H}_0$ with $\mathcal{H}(\delta,k)$. Therefore, we search in the next section the function K that maximises $\Delta(\delta)$ for given δ.

3. The focused alternatives.

3.1. **The standardized shape κ of the alternative that produces the largest increment in the asymptotic expectation of ω_n.** We shall obtain the function $K(u) = \int_0^u \kappa(s)ds$ that maximises $\int_0^1 K^2(u)du$ with the restrictions

$$\int_0^1 \kappa^2(u)du = 1, \int_0^1 \kappa(u)du = 0.$$

The associated Euler equations express that for each continuously differentiable g such that

$$g(0) = g(1) = 0, \int_0^1 K'(u)g'(u) = 0$$

the condition

$$\int_0^1 K(u)g(u)du = 0$$

must hold.

The condition $\int_0^1 K'(u)g'(u) = 0$ holds for every g such that $g(0) = g(1) = 0$ provided

$$\int_0^1 K'(u)g'(u)du = [g(u)K'(u)]_0^1 - \int_0^1 K''(u)g(u)du = 0.$$

Since the integrated term in the right-hand side vanishes, we find that when g is orthogonal to K'' in $L^2([0,1])$, it is also orthogonal to K, and this means that K and K'' are proportional, that is, for some constant $\pm\lambda^2$, K solves the differential equation $K'' = \pm\lambda^2 K$.

The solutions of $K'' = \pm\lambda^2 K$ in $[0,1]$ with border conditions $K(0) = K(1) = 0$, satisfying $\int_0^1 (K'(u))^2 du = 1$ are

$$K(u) = \frac{\sqrt{2}}{j\pi} \sin j\pi u, j = 1, 2, \ldots.$$

The solution with maximum norm is the one with $j = 1$, hence

$$\kappa(u) = \sqrt{2} \cos \pi u. \tag{6}$$

This is the standardized shape of the alternative that maximises (5) for given δ. The corresponding function $K(u) = \int_0^u \kappa(s)ds$ is proportional to the first function in the orthonormal system introduced in 2.1, that is, $K(u) = \psi_1(u)/\pi$.

3.2. **Alternatives of change in location.** When the alternative distributions specify a change in location

$$f_n(t) = f_0(t + \delta c/\sqrt{n})$$

we have

$$\sqrt{\frac{f_n(t)}{f_0(t)}} = 1 + \frac{\delta c}{2\sqrt{n}} \frac{f_0'(t)}{f_0(t)} + o(\tfrac{1}{\sqrt{n}})$$

so that $k(t) = c\frac{f_0'(t)}{f_0(t)}$. The constant c is introduced in order to be able to impose $\|k\|^2 = 1$.

It follows that $\kappa(u) = c\frac{f_0'(F_0^{-1}(u))}{f_0(F_0^{-1}(u))}$ and Equation (6) shows that the alternative shall be detected by the Cramér - von Mises statistic with maximum asymptotic increment of the expectation when

$$c\frac{f_0'(F_0^{-1}(u))}{f_0(F_0^{-1}(u))} = \sqrt{2}\cos\pi u.$$

In order to solve this differential equation in F_0, we return to the variable $t = F_0^{-1}(u)$, and get

$$cf_0'(t) = \sqrt{2}f_0(t)\cos\pi F_0(t),$$

which, integrated in $(-\infty, t]$ gives

$$cf_0(t) = \frac{\sqrt{2}}{\pi}\sin\pi F_0(t).$$

A further integration leads to

$$\frac{\sqrt{2}t}{c\pi} = \int_0^t \frac{dF_0(s)}{\sin\pi F_0(s)} = \int_{F_0(0)}^{F_0(t)} \frac{du}{\sin\pi u}$$

$$= \frac{1}{2\pi}\log\frac{(\cos\pi F_0(t) - 1)(\cos\pi F_0(0) + 1)}{(\cos\pi F_0(t) + 1)(\cos\pi F_0(0) - 1)}.$$

By imposing with no loss of generality that F_0 is centred in 0, follows the simpler expression

$$\gamma t = \log \frac{1 - \cos \pi F_0(t)}{1 + \cos \pi F_0(t)},$$

in which the parameter $\gamma = \frac{2\sqrt{2}}{c}$ determines the dispersion.

By solving in F_0 and choosing $\gamma = \pi$ to get a distribution with variance equal one, we conclude

$$F_0(t) = \frac{1}{\pi} \arccos \frac{1 - e^{\pi t}}{1 + e^{\pi t}}, \quad f_0(t) = \frac{1}{2\cosh(\gamma t/2)}. \tag{7}$$

3.3. **Asymptotic law of ω_n under changes in location for samples with the law of Equation (7), and power of the test.** The statistic ω_n has the asymptotic law of

$$Q(\delta) = \int_0^1 (b(u) + \delta K(u))^2 du = \int_0^1 \left(b(u) + \frac{\delta}{\pi}\psi_1(u) \right)^2 du.$$

Since $b(u) + \frac{\delta}{\pi}\psi_1(u) = \sum_{j=1}^{\infty} \frac{1}{j\pi} B_j + \frac{\delta}{\pi}\psi_1(u)$, then

$$Q(\delta) = \left\| b + \frac{\delta}{\pi}\psi_1 \right\|^2 = \frac{1}{\pi^2} \left[(B_1 + \delta)^2 + \sum_{j=2}^{\infty} \frac{1}{j^2} B_j^2 \right]$$

Cramér - von Mises test of $F^{(n)}(t) = F_0(t)$ against $F^{(n)}(t) = F_0(t + \frac{2\sqrt{2}\delta}{\pi\sqrt{n}})$ with significance level α is asymptotically equivalent to the test of $\mathcal{H}_0$:"$\delta = 0$" with critical region $Q(\delta) > c(\alpha)$ where $c(\alpha)$ solves $\mathbf{P}\{Q(0) > c(\alpha)\} = \alpha$. The power, that we have computed by a numerical convolution for the purposes discussed in next section, is

$$\Pi(\delta, \alpha) = \mathbf{P}\{Q(\delta) > c(\alpha)\}.$$

4. Comparison with the two-sided test based on Neymann and Pearson statistic.

The Neyman and Pearson test of $\mathcal{H}_0$ against the alternatives $\mathcal{H}_n$ that the true density of the sample distribution is $g_n(t) = f_0(t + \frac{\delta}{c\sqrt{n}})$

has critical region

$$\sum_{i=1}^{n} \log\left(f_0\left(X_i + \tfrac{\delta}{c\sqrt{n}}\right)/f_0(X_i)\right) \geq \text{constant},$$

asymptotically equivalent to

$$\frac{\delta}{\sqrt{n}} \sum_{i=1}^{n} \frac{f_0'(X_i)}{cf_0(X_i)} \geq \text{constant}.$$

When $\mathcal{H}_0$ holds, the variables $f_0'(X_i)/(cf_0(X_i))$ are centred, with variance 1, and therefore the asymptotic law of the statistic $T_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{f_0'(X_i)}{cf_0(X_i)}$ is standard normal.

If the sequence of alternatives $\mathcal{H}_n$ hold, then

$$\mathbf{E}T_n = \sqrt{n}\mathbf{E}f_0'(X_1)/(cf_0(X_1)) = \sqrt{n} \int \frac{f_0'(x)}{cf_0(x)} f_0(x + \tfrac{\delta}{c\sqrt{n}})dx$$

has limit δ, $\mathbf{E}(f_0'(X_i))/(cf_0(X_i)))^2$ tends to 1, hence T_n converges in law to $Z + \delta$, Z standard Gaussian.

As a consequence, the test of $\delta = 0$ against $\delta > 0$ with optimal asymptotic power is the one with critical region $T_n >$ constant.

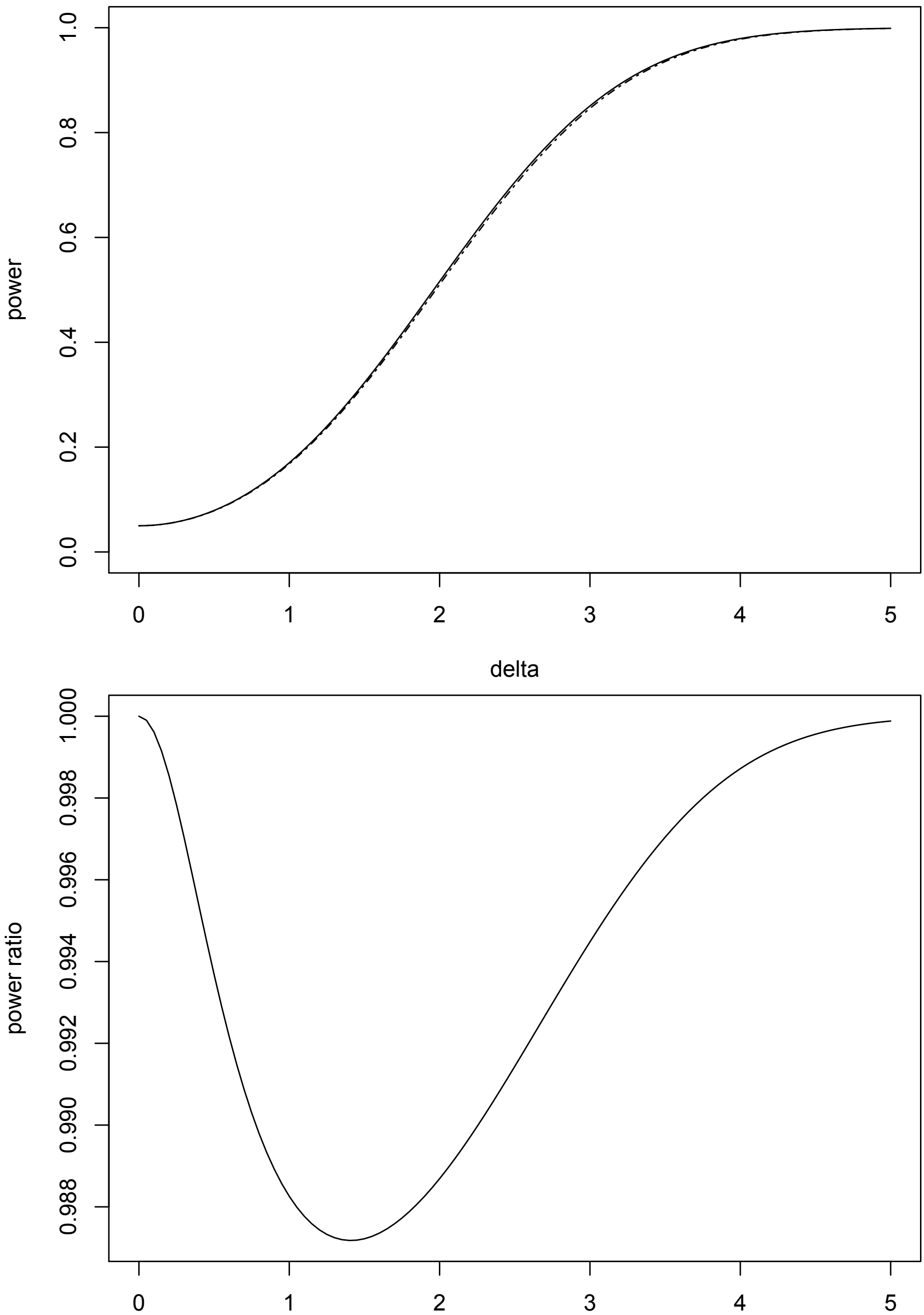
While there is no optimal test for $\delta = 0$ against $\delta \neq 0$, the usual practice if there are not significant differences between the cases $\delta > 0$ or $\delta < 0$ is to reject $\delta = 0$ when $|T_n| >$ constant. In that case, if Φ denotes as usual the standard normal cumulative distribution function, the asymptotic power of the two - sided test with asymptotic level α, is

$$\begin{aligned}\Pi^*(\delta, \alpha) &= \mathbf{P}\{Z + \delta > \Phi^{-1}(1 - \tfrac{\alpha}{2})\} + \mathbf{P}\{Z + \delta < \Phi^{-1}(\tfrac{\alpha}{2})\} \\ &= \Phi(\Phi^{-1}(\tfrac{\alpha}{2}) + \delta) + \Phi(\Phi^{-1}(\tfrac{\alpha}{2}) - \delta).\end{aligned}$$

The practically coincident plots of the functions $\Pi(\delta,.05)$ and $\Pi^*(\delta,.05)$ in Figure 1 show that Cramér - von Mises test against the alternative of displacement of samples with distribution (7) is *almost optimal*, in the sense that its performance is *almost* asymptotically equivalent to the performance of the test with critical region $T_n >$constant.

The relationship between the asymptotic powers (and the intended meaning of "almost optimal") is better shown in the second diagram

FIGURE 1. Almost coincident asymptotic powers $\Pi^*(\delta,.05)$ and $\Pi(\delta,.05)$ of the two-sided test based on the Neymann and Pearson statistic (solid line) and of the Cramér - von Mises test (dotted line), respectively, for alternatives of change in position of a sample with distribution (7) (upper diagram) and ratio $\Pi(\delta,.05)/\Pi^*(\delta,.05)$ (lower diagram).



of Figure 1, where the ratio $\Pi(\delta,.05)/\Pi^*(\delta,.05)$ obtained by numerical computation is plotted.

References

[1] Cabaña, A. and Cabaña, E. M. (2005), *Goodness-of-fit to the exponential distribution, focused on Weibull alternatives*, Communications in Statistics, Simulation and Computation **34**, 711-724.

[2] Cabaña, A. and Cabaña, E. M. (2009), *Goodness-of-fit tests for continuous regression*, Methodology and Computing in Applied Probability, **11** 119-144.

[3] Capon, J. (1965) *On the asymptotic efficiency of the Kolmogorov - Smirnov test.* J. Amer. Statist. Assoc. **60**, 843-853.

[4] Durbin, J. (1973) *Distribution Theory for Tests Based on the Sample Distribution Funcion.* SIAM, 64 p.

[5] Durbin, J. & Knott, M. (1972) *Components of Cramér - von Mises Statistics. I.* J. Royal Statist. Soc. Series B. **34**, 290-307.

[6] Hajek, J. & Sidak, Z. (1967) *Theory of Rank Tests*, Academic Press, New York.

[7] Le Cam L. & Yang, G. L. (1990), *Asymptotics in Statistics. Some basic concepts.* Springer-Verlag, New York.

[8] Neuhaus, G. (1976), *Asymptotic Power Properties of the Cramér-von Mises Test Under Contiguous Alternatives* Jr. Multivariate Analysis **6**, 95-110.

[9] Oosterhoff, J & van Zwet, W. R., (1979), *A note on contiguity and Hellinger distance*, in Contributions to Statistics, 157-166, Reidel, Dordrecht.

Universitat Autònoma de Barcelona, Spain.

Universidad de la República, Montevideo, Uruguay.

FEUILLETAGE DE HIRSCH, MESURES HARMONIQUES ET g-MESURES

BERTRAND DEROIN ET CONSTANTIN VERNICOS

1. Introduction

Un feuilletage lisse $\mathcal{F}$ est la donnée d'une variété M et d'une distribution intégrable $T\mathcal{F} \subset TM$ de dimension p, c'est à dire telle que par tout point de M passe une sous-variété de dimension p tangente à $T\mathcal{F}$. De telles sous-variétés sont appelées feuilles, p la dimension du feuilletage, $q = \dim(M) - p$ sa codimension, et $T\mathcal{F}$ le fibré tangent à $\mathcal{F}$.

Étant donnée une métrique lisse ds^2 sur le fibré tangent de $\mathcal{F}$, on note Δ le laplacien associé à cette métrique. Lucy Garnett a étudié dans [Ga] l'équation de la chaleur feuilletée $\frac{\partial u}{\partial t} = \Delta u$. Elle y développe la théorie ergodique du mouvement brownien le long des feuilles, et montre que les mesures stationnaires de ce processus de Markov sont les mesures de probabilité μ dites *harmoniques*, c'est à dire telles que $\Delta\mu = 0$ au sens faible. L'ensemble des mesures harmoniques forme un compact convexe de l'espace des mesures de probabilité sur la variété ambiante.

Lorsque le feuilletage est lisse et transversalement conforme, et qu'il ne possède pas de mesure transverse invariante, il est démontré dans [DK] qu'il n'existe qu'un nombre fini d'ensembles minimaux [1] supportant chacun une unique mesure harmonique, et que de surcroît, toute mesure harmonique est une combinaison convexe de celles-ci.

Le premier auteur a été financé par les projets ANR-08-JCJC-0130-01 et ANR-09-BLAN-0116, le second par la "Science Foundation Ireland" avec le programme "SFI Stokes lectureship".

1. On entend par *minimal* un ensemble fermé $\mathcal{F}$-saturé dans lequel toute les feuilles sont denses.

Ainsi, l'ensemble des mesures harmoniques sur un tel feuilletage est un simplexe de dimension $n-1$, où n est le nombre de minimaux de $\mathcal{F}$.

Les hypothèses de régularité de ce théorème sont les suivantes : le feuilletage est de classe C^1 transversalement, les feuilles sont des sous-variétés immergées de classe C^∞, et la métrique sur ces feuilles varie de façon höldérienne en fonction du paramètre transverse. Dans cette note, nous démontrons que si l'on affaiblit l'hypothèse de régularité sur la métrique, le résultat n'est plus valable :

Théorème. *Il existe un feuilletage lisse par surfaces d'une variété de dimension* 3*, dont toutes les feuilles sont denses, et une métrique* ds^2 *sur son fibré tangent qui est lisse le long des feuilles, qui dépend continûment du paramètre transverse – mais pas de façon Hölder – et pour laquelle il existe au moins deux mesures harmoniques différentes.*

Le feuilletage que nous considérons a été construit par Hirsch ; c'est en quelque sorte une suspension d'un revêtement du cercle dans lui-même de degré strictement plus grand que 1, voir la partie 2. L'idée principale de notre travail consiste à montrer qu'à partir d'une g-fonction associée au revêtement du cercle dans lui-même, voir [Ke], on peut construire une certaine métrique riemannienne ds^2 sur le fibré tangent du feuilletage de Hirsch, de façon à ce que toute g-mesure sur le cercle donne naissance à une mesure harmonique pour le laplacien feuilleté associé à ds^2, voir la partie 3 pour plus de détails. Notre théorème est alors une conséquence de cette construction et d'un théorème d'Anthony Quas, qui montre l'existence de g-fonctions ayant plusieurs g-mesures, voir [Qu].

Nous avons donc un exemple de feuilletage où la géométrie du convexe des mesures harmoniques varie en fonction de la métrique choisie sur les feuilles. Des exemples de ce type ont été trouvé par Victor Kleptsyn et Samuel Petite en toute régularité, mais en codimension supérieure [KP].

Signalons aussi que le feuilletage que nous construisons est minimal mais pas uniquement ergodique. De tels exemples ont été construit dans [De] en toute régularité en utilisant là encore un feuilletage de type suspension et l'existence d'un difféomorphisme minimal du tore non uniquement ergodique, dûe à Hillel Furstenberg.

Remerciements. Nous remercions le rapporteur pour ses remarques qui ont amélioré l'exposition.

2. Feuilletage de Hirsch

Hirsch a construit un feuilletage lisse de dimension 2 et de codimension 1 d'une variété compacte fermée, associé à un difféomorphisme local T du cercle dans lui-même, de degré topologique $d > 1$. Nous rappelons cette construction dans le cas de la transformation $T(z) = z^2$ du cercle unité dans lui-même.

Considérons un pantalon orienté P et notons ses trois composantes de bord $\partial_i P$, $i = 1, 2, 3$. Soit $\sigma : P \to P$ une involution lisse qui stabilise la composante $\partial_3 P$ de ∂P et échange les composantes $\partial_1 P$ et $\partial_2 P$. Par exemple, on pourra prendre pour P le disque unité de $\mathbf{C}$ privé des disques de rayon $1/4$ centrés en $1/2$ et $-1/2$, et poser $\sigma(x) = -x$. Les composantes $\partial_1 P$, $\partial_2 P$ sont alors respectivement les cercles de rayon $1/4$ de centre $1/2$ et $-1/2$, et $\partial_3 P$ est le cercle unité.

Notons $i(z) = -z$ l'involution du cercle dans lui-même qui consiste à échanger les points des fibres de T. Le quotient $N = (P \times \mathbf{S}^1)/(\sigma \times i)$ possède une structure naturelle de fibration en pantalons $P \to N \to \mathbf{S}^1/i$, et est difféomorphe à un tore solide duquel on a enlevé un tore solide intérieur qui fait deux fois le tour du premier - voir le survol [Gh] dans lequel le lecteur trouvera une jolie figure représentant ce quotient. Le bord de N est formé de deux composantes toriques : la composante intérieure $\partial_i N$ qui s'identifie naturellement avec $\partial_1 P \times \mathbf{S}^1$, et la composante extérieure $\partial_{ext} N$ qui est le quotient de $\partial_3 P \times \mathbf{S}^1$ par le difféomorphisme $\sigma \times i$. On recolle ces deux composantes par le difféomorphisme

$$(x, z) \in \partial_3 P \times \mathbf{S}^1/\sigma \times i \mapsto (\frac{xz}{4} + \frac{1}{2}, z^2) \in \partial_1 P \times \mathbf{S}^1.$$

On obtient une variété compacte fermée M. La fibration horizontale par pantalons sur N induit un feuilletage lisse $\mathcal{F}$ par surfaces ; c'est le feuilletage de Hirsch associé à T.

Pour construire une métrique sur le fibré tangent du feuilletage de Hirsch, il suffit de construire une famille de métriques $\{ds_z^2\}_{z\in\mathbf{S}^1}$ sur un voisinage ouvert U de P dans $\mathbf{C}$, telles que

– Pour tout $z \in \mathbf{S}^1$, on a $\sigma^\star ds_z^2 = ds_{i(z)}^2$.

– Au voisinage de $\partial_3 P$, on a $(\frac{xz}{4}+\frac{1}{2})^\star ds_{z^2}^2 = ds_z^2$ et $(\frac{xz}{4}-\frac{1}{2})^\star ds_{z^2}^2 = ds_{i(z)}^2$.

Une façon simple de construire de telles familles est de considérer des métriques sur P que nous appellerons *admissibles*. Une métrique admissible est une métrique ds^2 sur un voisinage de P dans $\mathbf{C}$ qui, au voisinage de $\partial_3 P$, admet l'expression

$$|ds| = \frac{|dz|}{|z|(2\pi + \log\frac{1}{|z|})},$$

et vérifie de plus $(\frac{x}{4}+\frac{1}{2})^\star ds^2 = ds^2$ et $(\frac{x}{4}-\frac{1}{2})^\star ds^2 = ds^2$. Alors, pour construire une métrique sur le fibré tangent de $\mathcal{F}$, il suffit de construire une famille $\{ds_z^2\}_{z\in\mathbf{S}^1}$ de métriques admissibles qui vérifient de surcroît la condition $\sigma^\star ds_z^2 = ds_{i(z)}^2$. En effet, une métrique admissible est invariante par rotation au voisinage de $\partial_3 P$, ce qui montre que pour toute paire de métriques admissibles ds_j^2, $j=0,1$, et tout z du cercle, on a $(\frac{xz}{4}\pm\frac{1}{2})^\star ds_0^2 = ds_1^2$.

3. g-mesures et mesures harmoniques

On reprend les notations du paragraphe précédent. Une *g-fonction*[2] est une fonction continue $g : \mathbf{S}^1 \to (1,+\infty)$ telle que pour tout point $z \in \mathbf{S}^1$,

$$\frac{1}{g(z)} + \frac{1}{g(i(z))} = 1.$$

Une *g-mesure* est une mesure de probabilité μ sur le cercle telle que la dérivée de Radon-Nikodym de T relativement à μ est la fonction g. Rappelons que cela signifie que, si B est un Borélien du cercle

2. La terminologie est malheureuse mais c'est celle qui est classiquement utilisée.

sur lequel T est injective, alors $\mu(TB) = \int_B g d\mu$. L'existence d'une g-mesure découle du théorème du point fixe de Kakutani, voir [Ke].

Lucy Garnett démontre dans [Ga] qu'il existe toujours une mesure harmonique sur un feuilletage équippé d'une métrique sur son fibré tangent, qui est lisse sur les feuilles, et continue transversalement. Dans ce qui suit, nous construisons explicitement des mesures harmoniques dans le cas particulier du feuilletage de Hirsch. Plus précisément, étant donnée une g-fonction associée à T, nous produisons une métrique riemannienne sur $T\mathcal{F}$, qui est lisse le long des feuilles et admet la même régularité transverse que g, en sorte que toute g-mesure donne lieu à une mesure harmonique sur $\mathcal{F}$.

Proposition 3.1. *Pour tout $\varepsilon > 0$, il existe un voisinage U de P dans $\mathbf{C}$, tel que pour tout couple (L_1, L_2) de réels supérieurs à ε et vérifiant*

$$e^{-L_1} + e^{-L_2} = 1,$$

il existe une métrique riemannienne admissible $ds^2_{L_1,L_2}$ sur U, et une fonction $\Delta_{ds^2_{L_1,L_2}}$-harmonique $\varphi_{L_1,L_2} : U \to \mathbf{R}^{>0}$ qui vérifie les conditions suivantes :

- *Pour tout x dans un voisinage de $\partial_3 P$, on a $\varphi_{L_1,L_2}(x) = 1 + \frac{1}{2\pi} \log \frac{1}{|x|}$.*
- *Pour tout x dans un voisinage de $\partial_3 P$, on a*

$$\varphi_{L_1,L_2}(\frac{x}{4}+1/2) = e^{-L_1}\varphi_{L_1,L_2}(x), \quad et \quad \varphi_{L_1,L_2}(\frac{x}{4}-1/2) = e^{-L_2}\varphi_{L_1,L_2}(x).$$

De plus, on peut supposer que les métriques $ds^2_{L_1,L_2}$ et les fonctions φ_{L_1,L_2} dépendent de façon analytique de L_1 et L_2, et que pour tout (L_1, L_2), on a $\sigma^\star ds^2_{L_1,L_2} = ds^2_{L_2,L_1}$.

Démonstration. Sur les cylindres $C_i = \mathbf{S}^1 \times [0, L_i]$, $i = 1, 2$, considérons la métrique de courbure -1 définie par :

$$e^{2(v-L_i)} du^2 + dv^2.$$

En effet, c'est la métrique qu'on obtient en partant de $\frac{du^2+dy^2}{y^2}$ et en effectuant le changement de variables $y = e^{L_i - v}$. Les bords $\partial_- C_i =$

$\mathbf{S}^1 \times 0$ et $\partial_+ C_i = \mathbf{S}^1 \times L_i$ sont alors des *horocycles* respectivement négatif de longueur e^{-L_i} et positif de longueur 1 [3].

On coupe C_1 et C_2 le long des géodésiques $1 \times [0, \varepsilon]$ et $1 \times [0, \varepsilon]$, et on colle le segment $1_+ \times [0, \varepsilon]$ de C_1 (resp. $1_- \times [0, \varepsilon]$ de C_2) au segment $1_- \times [0, \varepsilon]$ de C_2 (resp. $1_+ \times [0, \varepsilon]$ de C_1) de façon isométrique et en renversant l'orientation. On construit de cette façon un pantalon P_{L_1,L_2} avec une métrique ds^2 de courbure -1 et une singularité conique d'angle 4π. Ce pantalon P_{L_1,L_2} a trois composantes de bord : les composantes $\partial_1 P_{L_1,L_2} = \partial_+ C_1$, $\partial_2 P_{L_1,L_2} = \partial_+ C_2$, qui sont des horocycles positifs de longueur 1, et la composante $\partial_3 P_{L_1,L_2}$ qui est un horocycle négatif de longueur la somme des longueurs des bords $\partial_- C_1$ et $\partial_- C_2$, c'est à dire $e^{-L_1} + e^{-L_2} = 1$.

La métrique avec singularité conique munit P_{L_1,L_2} d'une structure de surface de Riemann lisse - l'atlas des cartes préservant l'orientation dans lesquelles la métrique est conforme à la métrique plate $|dz|$ sur $\mathbf{C}$. La fonction $\varphi_{L_1,L_2} : P_{L_1,L_2} \to \mathbf{R}$ définie sur chaque C_i par e^{-v} est alors une fonction *harmonique* sur P_{L_1,L_2}, qui vaut e^{-L_i} sur $\partial_i P_{L_1,L_2}$ pour $i = 1, 2$, et 1 sur $\partial_3 P_{L_1,L_2}$.

Les bords de P_{L_1,L_2} étant horocycliques de longueur 1, il existe un difféomorphisme $\Phi : P \to P_{L_1,L_2}$ tel que $\Phi^\star ds^2$ est une métrique admissible, à ceci près qu'elle admet une singularité conique. On peut choisir Φ en sorte que cette dernière se situe à l'origine, et que l'on ait en son voisinage $\Phi^\star ds^2 = |x|^2 |dx|^2$. On considère alors une métrique de la forme $ds^2_{L_1,L_2} = \rho \Phi^\star ds^2$, où $\rho : P \setminus \{0\} \to \mathbf{R}^{>0}$ est une fonction lisse, qui vaut identiquement 1 à l'extérieur d'un petit voisinage de l'origine, et qui, dans un voisinage encore plus petit, est de la forme $\rho(x) = \frac{1}{|x|^2}$.

La fonction $\varphi_{L_1,L_2} \circ \Phi$ est alors $\Delta_{ds^2_{L_1,L_2}}$-harmonique, et vérifie les conditions du lemme. □

Nous choisissons ε de sorte que $0 < \varepsilon < \inf_{z \in \mathbf{S}^1} \log g(z)$. Pour chaque point z du cercle, on pose

$$L_1(z) = \log g(z), \quad \text{et} \quad L_2(z) = \log g(i(z)).$$

3. Nous entendons par horocycle positif ou négatif une courbe lisse de courbure signée 1 ou -1.

La famille de métriques admissibles $\{ds_z^2\}_{z\in\mathbf{S}^1}$ définies par $ds_z^2 = ds^2_{L_1(z),L_2(z)}$ définit alors une métrique sur le feuilletage de Hirsch. D'autre part, si μ est une g-mesure sur le cercle, alors la mesure

$$m = \varphi_{L_1(z),L_2(z)}\mathrm{vol}(ds_z^2) \otimes \mu$$

définit une mesure harmonique sur le feuilletage de Hirsch. Pour montrer notre théorème, il nous suffit de prendre une g-fonction continue pour laquelle il existe plusieurs g-mesures différentes, dont l'existence nous est assurée par un théorème de Anthony N. Quas [Qu].

Références

[De] B. Deroin. Non unique-ergodicity of harmonic measures : smoothing Samuel Petite's examples. Differential Geometry. Proc. VIII Intern. col. Santiago de Compostela (2008).

[DK] B. Deroin & V. Kleptsyn. Random conformal dynamical systems. *Geom. Funct. Anal.* Vol. 17 (2007) no. 4, p. 1043-1105.

[Ga] L. Garnett. Foliations, the ergodic theorem and Brownian motion. *J. Funct. Anal.* **51** (1983), no.3, p. 285-311.

[Gh] É. Ghys. Laminations par surfaces de Riemann. Dynamique et géométrie complexes (Lyon, 1997), 4995, Panor. Synthèses, 8, Soc. Math. France, Paris, 1999.

[Ke] M. Keane. Strongly mixing g-measures. *Inventiones Math.* **16** (1972), p. 309-324.

[KP] V. Kleptsyn et S. Petite. Communication personnelle.

[Qu] A. N. Quas. Non-ergodicity for C^1 expanding maps and g-measures. *Ergodic Th. Dyn. Syst.* **16** (1996), p. 531-543.

CNRS UMR 8628, Département de Mathématique d'Orsay, Bâtiment 425, Université de Paris Sud, F–91405 Orsay cedex, France.
Courriel: bertrand.deroin@math.u-psud.fr

Institut de mathématique et de modélisation de Montpellier, Université Montpellier 2, Case Courrier 051, Place Eugène Bataillon, F–34395 Montpellier Cedex, France.
Courriel: constantin.vernicos@math.univ-montp2.fr

ON EXISTENCE OF SMOOTH CRITICAL SUBSOLUTIONS OF THE HAMILTON-JACOBI EQUATION

ALBERT FATHI

ABSTRACT. We give a (necessary and sufficient) condition to obtain C^k critical subsolutions for a Tonelli Hamiltonian on a compact manifold.

1. INTRODUCTION

To make the paper short, we will assume that the reader is familiar with weak KAM theory as it can be found in [10, 6] or in papers like [5, 6, 7, 8, 9, 13, 14, 16]. We will however recall some of the objects or theorems of the theory in this introduction, and, mainly, in section §3 below. Let us recall that weak KAM theory makes the connection between Mather's theory of Lagrangian systems, see [18], and global viscosity solutions of the Hamilton-Jacobi Equation whose existence was established by Lions, Papanicolaou, and Varadhan, see [17]

We consider M a compact connected manifold without boundary. We denote by $\pi : T^*M \to M$ he canonical projection from the cotangent bundle T^*M of M. As usual we will denote a point in T^*M by (x, p), with $x \in M$, and $p \in T^*_xM = \pi^{-1}(x)$. With this notation the canonical projection $\pi : T^*M \to M$ is nothing but $(x, p) \mapsto x$.

In the rest of the paper we will consider a Tonelli Hamiltonian $H : T^*M \to \mathbb{R}$, i.e. the function H satisfies:

1) H is C^2, where k at least 2;
2) $\partial^2 H/\partial p^2(x, p)$ is positive definite, for every $(x, p) \in M$.
3) $H(x, p)/\|p\|_x \to +\infty$ as $\|p\|_x \to +\infty$,

where in the last condition 3) we have used a norm on T_x^*M coming from a Riemannian metric on M. Since M is compact, all Riemannian metrics are equivalent, and this last condition 3) is satisfied by all Riemannian metrics as soon as it is satisfied by one of them.

As is usual now, see [7, 13, 14], we define the Mañé critical value $c[0]$ of H by

$$c[0] = \inf\{\sup_{x\in M} H(x, d_x u) \mid u \in C^1(M, \mathbb{R})\}.$$

Of course by density of $C^\infty(M, \mathbb{R})$ in $C^1(M, \mathbb{R})$ for the C^1 topology, we could have taken the inf on $C^\infty(M, \mathbb{R})$.

Recall, see for example [13, 14], that we say that $u : M \to \mathbb{R}$ is a critical subsolution (of the Hamilton-Jacobi Equation) if it is Lipschitz, and $H(x, d_x u) \leq c[0]$, for (Lebesgue) almost every $x \in M$. Due to the coercivity and convexity of H in p, a critical subsolution is nothing but a (global) viscosity subsolution of the Hamilton-Jacobi Equation

$$H(x, d_x u) = c[0],$$

see [1, Chapter II] or [2, Chapitre 2].

It is not difficult to obtain from the stability of viscosity subsolutions, again see [1, Proposition 2.2 page 35]or [2, Théorème 2.3 page 21], that there exists a critical subsolution. A much stronger result has been obtained by Patrick Bernard: there always exists $\mathrm{C}^{1,1}$ critical subsolutions (as usual a $\mathrm{C}^{1,1}$ function is a C^1 function, whose derivative is locally Lipschitz).

To a critical subsolution u, we can associate a specific compact non-empty subset $\mathcal{I}(u)$ called the projected Aubry set of u, such that $d_x u$ exists at every $x \in \mathcal{I}(u)$, and moreover $x \mapsto d_x u$ is Lipschitz on $\mathcal{I}(u)$. In fact, one has $H(x, d_x u) = c[0]$ for every $x \in \mathcal{I}(u)$; therefore it is very difficult to perturb a critical subsolution near $\mathcal{I}(u)$, while keeping it a critical subsolution. There are several possible descriptions for $\mathcal{I}(u)$, see for example [10, 13]. We will give some of these descriptions in §3.

Here is the main result of this paper. It shows that the problem of existence of smoother critical subsolutions is localized to a neighborhood of Aubry sets.

Theorem 1.1. *Let $u : M \to \mathbb{R}$ be a critical subsolution for the Tonelli Hamiltonian H on M. Suppose we can find an open subset U of M, with $\mathcal{I}(u) \subset U$, and a C^k map $\bar{u} : U \to \mathbb{R}$ such that:*

1) *$\bar{u} = u$ on $\mathcal{I}(u)$,*
2) *$H(x, d_x\bar{u}) \leq c[0]$, for every $x \in U$,*

then there exists a C^k critical subsolution $\tilde{u} : M \to \mathbb{R}$ with $\tilde{u} = u$ on $\mathcal{I}(u)$. Moreover, we can find such a critical subsolution $\tilde{u} : M \to \mathbb{R}$ which is C^∞ outside $\mathcal{I}(u)$, and strict outside $\mathcal{I}(u)$ (i.e. $H(x, d_x\tilde{u}) < c[0]$, for every $x \notin \mathcal{I}(u)$).

We obtain as a corollary the following result, see [3]

Corollary 1.2 (Patrick Bernard)**.** *If the Aubry set $\tilde{\mathcal{A}}^*$ consists of a finite number of hyperbolic orbits then we can find a C^k critical subsolution $u : M \to \mathbb{R}$ which is strict outside the projected Aubry set $\mathcal{A} = \pi(\tilde{\mathcal{A}}^*) \subset M$.*

For the definitions of the Aubry set $\tilde{\mathcal{A}}^* \subset T^*M$, see section §3 below.

2. An obvious way to combine subsolutions

This section contains the simple main idea of this work. It shows how to combine viscosity subsolutions to obtain a new one. Here we only need to assume that $H : T^*M \to \mathbb{R}$ is convex in p. Recall that, under this convexity assumption, a locally Lipschitz function $u : U \to M$, defined on the open subset $U \subset$, is a viscosity subsolution of

$$H(x, d_xu) = c,$$

where $c \in \mathbb{R}$ is fixed, if and only if $H(x, d_xu) \leq c$ (Lebesgue) almost everywhere on U, again see [1, Chapter II] or [2, Chapitre 2].

Proposition 2.1. *Suppose $c \in \mathbb{R}$ is fixed. Let $U \subset M$ be an open subset of M, and $u_1, u_2 : U \to \mathbb{R}$ be two locally Lipschitz maps satisfying $H(x, d_xu_i) \leq c, i = 1, 2$ (Lebesgue) almost everywhere on U. For any Lipschitz function $\rho : \mathbb{R} \to \mathbb{R}$, with ρ non-decreasing, and $\mathrm{Lip}(\rho) \leq 1$, the function $u_\rho = u_1 + \rho \circ (u_2 - u_1)$ also satisfies $H(x, d_xu_\rho) \leq c$ (Lebesgue) almost everywhere on U.*

Proof. We first consider the case where ρ is differentiable everywhere (for example C^1). The set

$$U' = \{x \in U \mid d_x u_1, d_x u_2 \text{ both exist, and } H(x, d_x u_i) \le c, i = 1, 2\}$$

is of full Lebesgue measure in U. For every $x \in U'$, we have

$$\begin{aligned} d_x u_\rho &= d_x u_1 + \rho'[(u_2 - u_1)(x)](d_x u_2 - d_x u_1) \\ &= \big(1 - \rho'[(u_2 - u_1)(x)]\big) d_x u_1 + \rho'[(u_2 - u_1)(x)] d_x u_2. \end{aligned}$$

Since ρ is non-decreasing we have $\rho'(t) \ge 0$. Moreover, since $\mathrm{Lip}(\rho) \le 1$, we also have $\rho'(t) \le 1$. Therefore $d_x u_\rho$ is a convex combination of $d_x u_1$ and $d_x u_2$. The convexity of $H(x, p)$ in p implies that $H(x, d_x u_\rho) \le c$, for every $x \in U'$, hence for (Lebesgue) almost every $x \in U$, since U' is of full Lebesgue measure in U.

Suppose now that we do not assume that ρ is differentiable everywhere. Choose $\theta_n : \mathbb{R} \to [0, +\infty[, n \in \mathbb{N}$, a C^∞ approximation of the identity for the convolution, then $\rho_n = \rho * \theta_n$ is also non-decreasing and has Lipschitz constant ≤ 1. The function ρ_n is C^∞, and $\rho_n \to \rho$ uniformly on any compact subset of $\mathbb{R}$. By the first part of the proof $u_{\rho_n} = u_1 + \rho_n \circ (u_2 - u_1)$ is locally Lipschitz and satisfies $H(x, d_x u_{\rho_n}) = c$ almost everywhere on U. Therefore u_{ρ_n} is a viscosity subsolution of $H(x, d_x u) = c$ on U. Since $u_{\rho_n} \to u_\rho$ uniformly on compact subsets of U, we obtain that the limit u_ρ is also a viscosity subsolution of $H(x, d_x u) = c$, see [1, Proposition 2.2 page 35]or [2, Théorème 2.3 page 21]. Therefore at each point x where the derivative of the locally Lipschitz function u_ρ exists, we have $H(x, d_x u_\rho) \le c$, see [1, Proposition 1.9 page 31] or [2, Corollaire 2.1 page 17]. □

3. Background

As said in the introduction details for this section can be found in [10, 6] or in papers like [5, 6, 7, 8, 9, 13, 14, 16].

We will assume in this section that $H : T^*M \to \mathbb{R}$ is a Tonelli Hamiltonian. We will call φ_t^H the Hamiltonian flow of H. This flow is the flow defined by the ODE

$$\dot{x} = \frac{\partial H}{\partial p}(x,p),$$
$$\dot{p} = -\frac{\partial H}{\partial x}(x,p).$$

If $u : M \to \mathbb{R}$ is a Lipschitz function, the derivative $d_x u$ exists for (Lebesgue) almost every $x \in M$, and we set

$$\mathrm{Graph}(du) = \{(x, d_x u) \mid x \text{ where } d_x u \text{ exists}\}.$$

If $u : M \to \mathbb{R}$ is a critical subsolution, we can define the Aubry set $\tilde{\mathcal{I}}^*(u)$ by

$$\tilde{\mathcal{I}}^*(u) = \bigcap_{t \in \mathbb{R}} \varphi_t^H[\mathrm{Graph}(du) \cap H^{-1}(c[0])].$$

Although this is not obvious this set is compact and non-empty. By its definition it is invariant under the Hamiltonian flow of H. The projected Aubry set of u is $\mathcal{I}(u) = \pi(\tilde{\mathcal{I}}^*(u))$, where $\pi : T^*M \to M$ is the canonical projection.

The Aubry set $\tilde{\mathcal{A}}^*$ of H is

$$\tilde{\mathcal{A}}^* = \bigcap \{\tilde{\mathcal{I}}^*(u) \mid u : M \to \mathbb{R} \text{ is a critical subsolution}\},$$

and the projected Aubry set $\mathcal{A}$ is $\pi(\tilde{\mathcal{A}}^*)$. It can be shown that $\mathcal{A}$ is the intersection, over all u critical subsolution, of the projected Aubry sets $\mathcal{I}(u)$. It can also be shown that there exists a critical subsolution u such that $\tilde{\mathcal{A}}^* = \tilde{\mathcal{I}}^*(u)$, and $\mathcal{A} = \mathcal{I}(u)$.

We gave the definition above of Aubry sets, because it is the quickest to give. We will develop the connection with the more usual definition of Aubry sets as given in [10] (or in [9] where they are rather called Peierls sets).

We recall that the Lagrangian $L : TM \to \mathbb{R}$ is given by

$$L(x,v) = \sup_{p \in T_x^*M} p(v) - H(x,p).$$

The map L is as smooth as the Tonelli Hamiltonian H. Moreover, it satisfies the analogous of the properties 2) and 3) of a Tonelli

Hamiltonian. The Legendre transform $\mathcal{L} : TM \to T^*M$ defined by

$$\mathcal{L}(x, v) = \big(x, \frac{\partial L}{\partial v}(x, v)\big),$$

is a global diffeomorphism whose inverse is given by

$$\mathcal{L}^{-1}(x, p) = \big(x, \frac{\partial H}{\partial p}(x, p)\big).$$

One has the Fenchel inequality

$$\forall x \in M, \forall v \in T_xM, \forall p \in T_x^*M, p(v) \leq L(x, v).$$

The Fenchel inequality is an equality if and only if $(x, p) = \mathcal{L}(x, v)$ ($\Leftrightarrow p = \partial L/\partial v(x, v) \Leftrightarrow v = \partial H/\partial p(x, p)$).

Suppose now that u is a strict subsolution, and fix $(x, p) \in \tilde{\mathcal{I}}^*(u)$. If we write $\varphi_t^H(x, p) = (\gamma(t), p(t))$ then $p(t) = d_{\gamma(t)}u$, and $H(\gamma(t), d_{\gamma(t)}u) = c[0]$. Since the Legendre transform exchanges speed curves of extremals of the Lagrangian l and orbits of φ_t^H, we have $d_{\gamma(t)}u = p(t) = \partial L/\partial v(\gamma(t), \dot\gamma(t))$. Therefore using the equality case in the Fenchel inequality, we get

$$d_{\gamma(t)}u(\dot\gamma(t)) = L(\gamma(t), \dot\gamma(t)) + c[0].$$

By integration this implies that $\gamma :] - \infty, +\infty[\to M$ is $(u, L, c[0])$-calibrated. Recall that a curve $\gamma : I \to M$, where I in an interval in $\mathbb{R}$, is said to be $(u, L, c[0])$-calibrated if for all $t, t' \in I$, with $t \leq t'$, we have

$$u(\gamma(t')) - u(\gamma(t)) = \int_t^{t'} L(\gamma(s), \dot\gamma(s))\, ds + c[0](t' - t).$$

Conversely, if $\gamma :] - \infty, +\infty[\to M$ is $(u, L, c[0])$-calibrated, using the properties of calibrated curves, we have that $d_{\gamma(t)}u$ exists, and

$$d_{\gamma(t)}u = \frac{\partial L}{\partial v}(\gamma(t), \dot\gamma(t)) \text{ and } H(\gamma(t), d_{\gamma(t)}u) = c[0].$$

Since a calibrated curve is an extremal, it follows that $t \mapsto (\gamma(t), d_{\gamma(t)}u)$ is an orbit of φ_t^H, contained in $\tilde{\mathcal{I}}^*(u)$. Hence $\tilde{\mathcal{I}}^*(u) = \mathcal{L}(\tilde{\mathcal{I}}(u))$, where

$$\tilde{\mathcal{I}}(u) = \{(\gamma(0), \dot\gamma(0)) \mid \gamma :]{-\infty}, +\infty[\to M \text{ is } (u, L, c[0])\text{-calibrated}\} \subset TM.$$

This makes the connection with a more usual definition of $\mathcal{I}(u)$ as the projection of $\tilde{\mathcal{I}}(u)$ on M, or

$$\mathcal{I}(u) = \{x \in M \mid \exists \gamma :]-\infty, +\infty[\to M \ (u, L, c[0])\text{-calibrated, and } \gamma(0) = x\}.$$

We now recall the definition of the Lax-Oleinik semi-groups

$$T_t^-, T_t^+ : \mathcal{C}^0(M, \mathbb{R}) \to \mathcal{C}^0(M, \mathbb{R}).$$

If $t \geq 0$ and $u \in \mathcal{C}^0(M, \mathbb{R})$, we have

$$T_t^-(u)(x) = \inf_\gamma u(\gamma(-t)) + \int_{-t}^0 L(\gamma(s), \dot{\gamma}(s))\, ds,$$

where the infimum is taken over all curves $\gamma : [-t, 0] \to M$, with $\gamma(0) = x$. In the same way

$$T_t^+(u)(x) = \sup_\gamma u(\gamma(t)) - \int_0^t L(\gamma(s), \dot{\gamma}(s))\, ds,$$

where the supremum is taken over all curves $\gamma : [0, t] \to M$, with $\gamma(0) = x$. The function $u : M \to \mathbb{R}$ is a critical subsolution if and only if $u \leq T_t^-(u) + c[0]t$, for every $t \geq 0$ ($\Leftrightarrow u \geq T_t^+(u) - c[0]t$, for every $t \geq 0$).

A negative (resp. positive) weak KAM solution is a function $u_- : M \to \mathbb{R}$ (resp. $u_+ : M \to \mathbb{R}$) such that $u_- = T_t^- u_- + c[0]t$ (resp. $u_+ = T_t^+ u_+ - c[0]t$, for every $t \geq 0$. By what we said above weak KAM solutions are automatically critical subsolutions.

Given a critical subsolution u, then $T_t^- u + c[0]t$ (resp. $T_t^+(u) + c[0]t$) is non-increasing (resp. non-decreasing) in t, and converges uniformly to a negative (resp. positive) weak KAM solution $u_- \geq u$ (resp. $u_+ \leq u$). For proof of the convergence see [10], or arguments in [9]. In particular, we have $u_+ \leq u \leq u_-$. For a given $x \in M$, we have $u(x) = u_-(x)$ if and only if $u(x) = T_t^- u(x) + c[0]t$, for every $t \geq 0$. It follows that $u_-(x) = u(x)$ if and only if there exists a $(u, L, c[0])$-calibrated curve $\gamma :] - \infty, 0] \to M$ with $\gamma(0) = x$. In the same way $u_+(x) = u(x)$ if and only if there exists a $(u, L, c[0])$-calibrated curve $\gamma : [0, \infty[\to M$ with $\gamma(0) = x$. Since $u_+ \leq u \leq u_-$, we obtain that $u_+(x) = u_-(x)$, if and only if we can find $\gamma :] - \infty, \infty[\to M$

$(u, L, c[0])$-calibrated with $\gamma(0) = x$. This implies

$$\mathcal{I}(u) = \{x \in M \mid u_+(x) = u_-(x)\}.$$

In particular $u_- - u_+ > 0$ on $M \setminus \mathcal{I}(u)$. It can be shown that $\mathcal{I}(u) = \mathcal{I}(u_-) = \mathcal{I}(u_+)$.

It is useful to introduce the concept for a critical subsolution of being strict on an open subset, see [13, 14]. Since we will use this concept when the function is at least C^1 on the open set, we can give the following definition: we will say that the critical subsolution $u : M \to \mathbb{R}$ is strict on the open subset $U \subset M$ if it is C^1 on U and

$$\forall x \in U, H(x, d_x u) < c[0].$$

We will need the following density theorem. For a proof see for example [13, §7]and [14, §6].

Theorem 3.1. *If $u : M \to \mathbb{R}$ is a critical subsolution, and $\epsilon > 0$ is given, we can find a critical subsolution $\tilde{u} : M \to \mathbb{R}$ such that:*

1) *the function $\tilde{u}$ is C^∞ and strict on $M \setminus \mathcal{A}$;*
2) *$\|\tilde{u} - u\|_\infty < \epsilon$.*

4. Proof of Theorem 1.1

In this section we will assume that $u : M \to \mathbb{R}$ is a critical subsolution. By what was recalled in the previous section §3, we can find a pair (u_-, u_+) of negative and positive weak KAM solutions, with $u_- \geq u \geq u_+$ and $\mathcal{I}(u) = \{x \mid u_+(x) = u_-(x)\}$.

We will further assume that $\bar{u} : U \to \mathbb{R}$ is a C^k function such that:

1) U is an open subset of M containing $\mathcal{I}(u)$,
2) $\bar{u} = u (= u_- = u_+)$ on $\mathcal{I}(u)$,
3) $H(x, d_x \bar{u}) \leq c[0]$ on U.

Lemma 4.1. *Let K be a compact subset of $M \setminus \mathcal{I}(u)$. We can find a C^k function $u_1 : U \to \mathbb{R}$, such that $H(x, d_x u_1) \leq c[0]$, for every $x \in U$, $u_1 = \bar{u}$ on a neighborhood of $\mathcal{I}(u)$,and $u_1 < u_-$ on $U \cap K$.*

Proof. Since $u_- > u_+$ on the compact set K, we can choose $\alpha > 0$ such that $u_+ + 4\alpha < u_-$, on K. Since u_+ is a critical subsolution, by Theorem 3.1, we can find a global critical subsolution $\tilde{u}_+ : M \to \mathbb{R}$

which is C^∞ outside of $\mathcal{A} \subset \mathcal{I}(u) = \mathcal{I}(u_+)$, and $\|u_+ - \tilde{u}_+\|_\infty \leq \alpha$. Since $u_+ = \bar{u} = u$ on $\mathcal{I}(u)$, we can find an open neighborhood $V \subset U$ of $\mathcal{I}(u)$ such that $\bar{u} - u_+ \leq \alpha$ on V. Therefore $\bar{u} - \tilde{u}_+ \leq 2\alpha$ on V. Let $\theta : \mathbb{R} \to [0,1]$ be a C^∞ function such that $\theta = 1$ on $]-\infty, 2\alpha]$ and $\theta = 0$ on $[3\alpha, +\infty[$. We now define $\rho : \mathbb{R} \to \mathbb{R}$ by

$$\rho(t) = \int_0^t \theta(s)\, ds.$$

The function ρ is clearly C^∞, non-decreasing, and its Lipschitz constant is 1. Since θ is non-negative, bounded by 1, and is identically 0 on $[3\alpha, +\infty[$, we get

$$\max \rho = \int_0^{3\alpha} \theta(s)\, ds \leq \int_0^{3\alpha} ds = 3\alpha.$$

By Proposition 2.1, the function $u_1 = \tilde{u}_+ + \rho(\bar{u} - \tilde{u}_+)$ satisfies $H(x, d_x u_\rho) \leq c[0]$ on U. We also have $u_1 \leq \tilde{u}_+ + \max \rho \leq \tilde{u}_+ + 3\alpha \leq u_+ + 4\alpha$. Therefore, by the choice of α, we obtain $u_1 < u_-$ on $K \cap U$. Note that u_1 is C^k outside of $\mathcal{I}(u)$. On the open set $V \supset \mathcal{I}(u)$, we have $\bar{u} - \tilde{u}_+ \leq 2\alpha$. Since on $]-\infty, 2\alpha]$ the derivative ρ' is identically 1, we have $\rho(t) = t$, for every $t \in]-\infty, 2\alpha]$. On V, we therefore get $\rho(\bar{u} - \tilde{u}_+) = \bar{u} - \tilde{u}_+$, and $u_1 = \bar{u}$. In particular, the function u_1 is also C^k on V, hence on $U = V \cup (U \setminus \mathcal{I}(u))$. □

Lemma 4.2. *For any neighborhood W of $\mathcal{I}(u)$, we can find a C^k function $u_2 : M \to \mathbb{R}$ such that*

1) *$u_2 = \bar{u}$ in a neighborhood of $\mathcal{I}(u)$,*
2) *u_2 is a critical subsolution,*
3) *u_2 is a strict critical subsolution outside of W, i.e. $H(x, d_x u_2) < c[0]$, for every $x \in M \setminus W$.*

Proof. We can assume $\bar{W} \subset U$. Moreover, by Lemma 4.1, applied with $K = M \setminus W$, replacing $\bar{u}$ by u_1 if necessary, we can also assume $\bar{u} < u_-$ on $U \setminus W$. Choose U' a neighborhood of $\bar{W}$ with $\bar{W} \subset U' \subset \bar{U}' \subset U$. Define β by

$$3\beta = \inf_{\bar{U}' \setminus W} u_- - \bar{u}.$$

Note that $\beta > 0$, since $\bar{U}' \setminus W$ is a compact subset of $U \setminus W$, on which the continuous function $u_- - \bar{u}$ is > 0. Since u_- is a critical subsolution, by Theorem 3.1, we can choose a critical subsolution $\tilde{u}_- : M \to \mathbb{R}$ which is C^∞, strict outside $\mathcal{A} \subset \mathcal{I}(u) = \mathcal{I}(u_-)$, and satisfies $\|\tilde{u}_- - u_-\|_\infty \leq \beta$. Call V a neighborhood of $\mathcal{I}(u)$ such that $u_- - \bar{u} \leq \beta$ on V. This is possible because $u_- = \bar{u} = u$ on $\mathcal{I}(u)$. We obtain $\tilde{u}_- - \bar{u} \leq 2\beta$ on V. Let $\theta : \mathbb{R} \to [0,1]$ be a C^∞ function such that $\theta = 0$ on $]-\infty, 2\beta]$ and $\theta = 1$ on $[3\beta, +\infty[$. We define $\rho : \mathbb{R} \to \mathbb{R}$ by

$$\rho(t) = \int_0^t \theta(t)\, dt.$$

The function ρ is C^∞, non-decreasing, and has Lipschitz constant 1. By the choice of θ, the function ρ is identically 0 on $]-\infty, 2\beta]$, and $\rho(t) = t + \rho(3\beta) - 3\beta$ on $[3\beta, +\infty[$. The function $u_\rho = \bar{u} + \rho(\tilde{u}_- - \bar{u})$ is defined on U and C^k on $U \setminus \mathcal{I}(u)$. By Proposition 2.1, the function u_ρ satisfies $H(x, d_x u_\rho) \leq c[0]$. By the properties of ρ, it also satisfies $u_\rho = \bar{u}$ on V and $u_\rho = \tilde{u}_- + \rho(3\beta) - 3\beta$ on $\bar{U}' \setminus W$. In particular u_ρ is C^k on the whole of U. Since $\bar{W} \subset U'$, we can define a C^k function $u_2 : M \to \mathbb{R}$ such that $u_2 = u_\rho$ on U' and $u_2 = \tilde{u}_- + \rho(3\beta) - 3\beta$ on $M \setminus W$. Note that u_2 is a critical subsolution which is strict outside W like $\tilde{u}_-$. □

Proof of Theorem 1.1. Choose a sequence $V_n, n \geq 0$, of neighborhoods of $\mathcal{I}(u)$ such that $V_{n+1} \subset V_n$ and $\cap_{n \in \mathbb{N}} V_n = \mathcal{I}(u)$. By Lemma 4.2, we can find a sequence $u_n : M \to \mathbb{R}$ of C^k critical subsolutions, such that $u_n = u$ on $\mathcal{I}(u)$, and u_n is strict outside V_n. We can pick a converging series $\epsilon_n > 0, n \geq 0$, such that $\sum_{n \in \mathbb{N}} \epsilon_n u_n$ converges in the C^k topology (this is easy to show see for example [12, Lemma 3.3, page 722]. Changing a finite number of terms, we can assume $\sum_{n \in \mathbb{N}} \epsilon_n = 1$. By the convexity of H in p, the sum $\sum_{n \in \mathbb{N}} \epsilon_n u_n$ is also a critical subsolution. It is strict outside $V_n, n \geq 0$, since $\epsilon_n > 0$. Hence it is strict outside $\cap_{n \in \mathbb{N}} V_n = \mathcal{I}(u)$. Of course $\sum_{n \in \mathbb{N}} \epsilon_n u_n = u$ on $\mathcal{I}(u)$.

To make $\tilde{u} = \sum_{n \in \mathbb{N}} \epsilon_n u_n$ of class C^∞ outside $\mathcal{I}(u)$, we can now take an appropriate C^k approximation of $\tilde{u}$ on $M \setminus \mathcal{I}(u)$ in the strong (or Whitney) topology which is C^∞ on that subset, see [15] for the strong topology and the approximation theorems. □

Proof of Corollary 1.2. As observed in [3, §6], under the hypothesis of the corollary, we can find a weak KAM solution u_- such that $\mathcal{I}(u_-) = \mathcal{A}$, and u_- is C^k is a neighborhood of $\mathcal{A}$. It therefore suffices to apply Theorem 1.1 with $u = \bar{u} = u_-$ to obtain the corollary. □

Acknowledgement. I would like to thank my daughter Lola for TEXing my hand-written notes, I would also like to thank Pierre Pageault and Maxime Zavidovique for corrections.

References

[1] M. Bardi and I. Cappuzzo-Dolceta. Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations. Systems & Control: Foundations and Applications, Birkhäuser Boston Inc., Boston, MA, 1997.

[2] G. Barles. Solutions de viscosité des équations de Hamilton-Jacobi. Mathématiques et Applications 17, Springer-Verlag, Paris, 1994.

[3] P. Bernard. Smooth critical sub-solutions of the Hamilton-Jacobi equation. *Math. Res. Lett.* 14(3):503–511, 2007.

[4] P. Bernard. Existence of $C^{1,1}$ critical sub-solutions of the Hamilton-Jacobi equation on compact manifolds. *Ann. Sci. École Norm. Sup. (4)* 40(3):445–452, 2007.

[5] P. Bernard. The dynamics of pseudographs in convex Hamiltonian systems. *J. Amer. Math. Soc.* 21(3):615–669, 2008

[6] G. Contreras & R. Iturriaga Global minimizers of autonomous Lagrangians. 22° Colquio Brasileiro de Matemtica. [22nd Brazilian Mathematics Colloquium] Instituto de Matemtica Pura e Aplicada (IMPA), Rio de Janeiro, 1999.

[7] G. Contreras, R. Iturriaga, G.P. Paternain & M.Paternain, Lagrangian graphs, minimizing measures and Ma's criticalvalues. *Geom. Funct. Anal.* 8(5):788–809, 1988.

[8] A. Fathi. Théorème KAM faible et théorie de Mather sur les systèmes lagrangiens. *C. R. Acad. Sci. Paris Sér. I Math.*, 324(9):1043–1046, 1997.

[9] A. Fathi. Solutions KAM faibles conjuguées et barrières de Peierls. *C. R. Acad. Sci. Paris Sér. I Math.*, 325(6):649–652, 1997.

[10] A. Fathi. Weak KAM Theorem and Lagrangian Dynamics. `http://www.math.u-bordeaux1.fr/~thieulle/Recherche/KamFaible/Publications/Fathi2008_01.pdf`

[11] A. Fathi. Denjoy-Schwartz and Hamilton-Jacobi. Viscosity Solutions of Differential Equations and Related Topics 2008/06/25–2008/06/27 *RIMS Kôkyûroku* 1651, 2009.

[12] A. Fathi. Partitions of unity for countable covers. *Amer. Math. Monthly*, 104(8):720–723, 1997.

[13] A. Fathi and A. Siconolfi. Existence of C^1 critical subsolutions of the Hamilton-Jacobi equation. *Invent. math.*, 155(2):363–388, 2004.

[14] A. Fathi and A. Siconolfi. PDE aspects of Aubry-Mather theory for quasiconvex Hamiltonians *Calc. Var. Partial Differential Equations*, 22(2):185-228, 2005.

[15] M. Hirsch. Differential topology. Graduate Texts in Mathematics, 33, Springer, (1976)

[16] V.Y. Kaloshin. Mather theory, weak KAM theory, and viscosity solutions of Hamilton-Jacobi PDE's. EQUADIFF 2003, 39–48, World Sci. Publ., Hackensack, NJ, 2005.

[17] P.-L. Lions, G. Papanicolau and S.R.S. Varadhan. Homogenization of Hamilton-Jacobi equation. Unpublished preprint, 1987.

[18] J. N. Mather. Action minimizing measures for positive definite Lagrangian systems. *Math. Z.*, 207:169–207, 1991.

PATHS TOWARDS ADAPTIVE ESTIMATION FOR INSTRUMENTAL VARIABLE REGRESSION

JEAN-MICHEL LOUBES AND CLÉMENT MARTEAU

Abstract. We tackle the problem of estimating a regression function observed in an instrumental regression framework. This model is an inverse problem with unknown operator. We provide a spectral cut-off estimation procedure which enables to derive oracle inequalities which warrants that our estimate, built without any prior knowledge, behaves as well as, up to log term, if the best model were known.

Introduction

An economic relationship between a response variable Y and a vector of explanatory variables X is often represented by an equation

$$Y = \varphi(X) + U,$$

where φ is the parameter of interest which models the relationship while U is an error term. Contrary to usual statistical regression models, the error term is correlated with the explanatory variables X, hence $\mathbf{E}(U|X) \neq 0$, preventing direct estimation of φ. To overcome the endogeneity of X, we assume that there exists an observed random variable W, called the instrument, which decorrelates the effects of the two variables X and Y in the sense that $\mathbf{E}(U|W) = 0$. It is often the case in economics, where the practical construction of instrumental variables play an important part, for instance for practical situations where prices of goods and quantity in goods can be explained using an instrument. This situation is also encountered

2000 *Mathematics Subject Classification.* 62G05, 62G20.

Key words and phrases. Inverse Problems, Instrumental Variables, Model Selection, Econometrics.

when dealing with simultaneous equations, error-in-variable models, treatment model with endogenous effects. It defines the so-called instrumental variable regression model which has received a growing interest among the last decade and turned to be a challenging issue in statistics. In particular, we refer to [NP03] for general references on the use of instrumental variables in economics while [CFR06] deal with the statistical estimation problem.

More precisely, we aim at estimating a function φ from the observations of (Y, X, W) satisfying the following condition

$$Y = \varphi(X) + U, \quad \begin{cases} \mathbf{E}(U|X) & \neq 0 \\ \mathbf{E}(U|W) & = 0 \end{cases} \tag{1}$$

Hence, the model (1) can be rewritten as an inverse problem using the expectation conditional operator with respect to W, which will be denoted T, as follows :

$$r := \mathbf{E}(Y|W) = \mathbf{E}(\varphi(X)|W) = T\varphi. \tag{2}$$

The function r is not known and only an observation $\hat{r}$ is available, leading to the usual inverse problem settings

$$\hat{r} = T\varphi + \delta, \tag{3}$$

where φ is defined as the solution of a noisy Fredholm equation of the first order which may generate an ill-posed inverse problem. The literature on inverse problems in statistics is large, but contrary to most of the problems tackled in the literature on inverse problems (see [EHN96], [MR96], [CGPT02], [CHR03], [LL08] and [O'S86] for general references), the operator T is unknown either, which transforms the model into an inverse problem with unknown operator. Few results exist in this settings and only very recently new methods have arisen. In particular [CH05], [Mar06, Mar09], or [EK01] and [HR08] in a more general case, construct estimators which enable to estimate inverse problem with unknown operators in an adaptive way, i.e getting optimal rates of convergence without prior knowledge of the regularity of the functional parameter of interest.

In this work, we are facing an even more difficult situation since both r and the operator T have to be estimated from the same sample. Some attention has been paid to this estimation issue, by estimating the joint density with different kinds of technics such as kernel based Tikhonov regularization [CFR06], regularization in Hilbert scales, finite dimensional sieve minimum distance estimator [NP03], with different rates and different smoothness assumptions, providing sometimes minimax rates of convergence. But, to our knowledge, all the proposed estimators rely on prior knowledge on the regularity of the function φ expressed through an embedding condition into a smoothness space or an Hilbert scale, or a condition linking the regularity of φ to the regularity of the operator, namely a link condition or source condition (see [CR08] for general comments and insightful comments on such assumptions). In a first part, we explain how to use a general penalized approach to turn any regularization scheme into an adaptive procedure when the operator is known. But the extension of this method to the case of IV regression fails, hence we provide under some conditions for the SVD decomposition, an adaptive estimation procedure of the function φ which converges, without prior regularity assumption, at the optimal rate of convergence, up to a logarithmic term. Moreover, we derive an oracle inequality which ensures optimality among the different choices of estimators.

Hence, the objective of this work is twofold; first extending the estimation procedure for inverse problem with unknown operator to the case of correlated data, and yet obtaining an oracle inequality; then providing a tractable adaptive estimator to some cases of instrumental variable regression.

1. A statistical framework for instrumental variable (IV) regression

1.1. **Mathematical model.** We observe an i.i.d sample (Y_i, X_i, W_i) for $i = 1, \ldots, n$ with unknown distribution f. Define the following

Hilbert spaces

$$L^2_X = \{h : \mathbb{R} \to \mathbb{R},\ \|h\|^2_X := \mathbf{E}(h^2(X)) < +\infty\}$$
$$L^2_W = \{g : \mathbb{R} \to \mathbb{R},\ \|g\|^2_W := \mathbf{E}(g^2(W)) < +\infty\},$$

with the corresponding scalar product $\langle .,. \rangle_X$ and $\langle .,. \rangle_W$. For sake of convenience, we only consider in this paper the case where φ is univariate. The approach presented in this paper may be certainly extended to the multivariate case (i.e. with a variable X of dimension $d > 1$).

Then the conditional expectation operator of X with respect to W is defined as an operator T

$$\begin{aligned} T : \quad & L^2_X \to L^2_W \\ & g \to \mathbf{E}[g(X)|W = .] \quad . \end{aligned}$$

The model (1) can be written, as discussed in [CR08], as

$$\begin{aligned} Y_i &= \varphi(X_i) + \mathbf{E}[\varphi(X_i)|W_i] - \mathbf{E}[\varphi(X_i)|W_i] + U_i \\ &= \mathbf{E}[\varphi(X_i)|W_i] + V_i \\ &= T\varphi(W_i) + V_i, \end{aligned} \tag{4}$$

where $V_i = \varphi(X_i) - \mathbf{E}[\varphi(X_i)|W_i] + U_i$, is such that $\mathbf{E}(V|W) = 0$. The parameter of interest is the unknown function φ. Hence, the observation model turns to be an inverse problem with unknown operator T with a correlated noise V. Solving this issue amounts to deal with the estimation of the operator and then controlling the correlation with respect to the noise.

The operator T is unknown since it depends on the unknown distribution of the observed variables Y, X, W denoted $f_{(Y,X,W)}$. The estimation of this operator can be performed either by directly using an estimate of $f_{(Y,X,W)}$, or if exists, by estimating the spectral value decomposition of the operator.

Assume that T is compact and admits a singular value decomposition (SVD) $(\lambda_j, \phi_j, \psi_j)_{j\geqslant 1}$, which provides a natural basis adapted to the operator for representing the function φ, see for instance [EHN96].

More precisely, let T^* be the adjoint operator of T, then T^*T is a compact operator on L^2_X with eigenvalues λ_j^2, $j \geqslant 1$ associated to the corresponding eigenfunctions ϕ_j, while ψ_j are defined by $\psi_j = \frac{T\phi_j}{\|T\phi_j\|}$. So we obtain

$$T\phi_j = \lambda_j \psi_j, \quad T^*\psi_j = \lambda_j \phi_j.$$

The decay of the eigenvalues defines the difficulty of the inverse problem. Hereafter, we only consider the case of mildly ill-posed inverse problems, i.e when the eigenvalues decay at a polynomial rate.

IP: Degree of ill-posedness: We assume that there exists t, called the degree of ill-posedness of the operator which controls the decay of the eigenvalues of the operator T. More precisely, there are constants λ_L, λ_U such that

$$\lambda_L k^{-t} \leqslant \lambda_k \leqslant \lambda_U k^{-t}, \ \forall k \geqslant 1 \tag{5}$$

We assume some conditions on the observations errors in order to obtain Hoeffding-type concentration bounds. Other equivalent conditions can be used.

Exponential Moment conditions:: The observation Y satisfy to the following moment condition. There exists some positive numbers $v \geqslant \mathbf{E}(Y_j^2)$ and c such that

$$\forall j \geqslant 1, \ \forall k \geqslant 2, \quad \mathbf{E}(Y_j^k) < \frac{k!}{2} v c^{k-2}. \tag{6}$$

1.2. **An econometric example.** Instrumental variable regression in econometrics are used when modeling a relationship between correlated variables. It occurs usually when considering the econometric problem of the estimation of price and demand of goods. If Q is a quantity of a good with price P observed over the years, the usual linear regression model

$$\begin{aligned} \log Q_i &= \beta_0 + \beta_1 \log P_i + U_i \\ &= f(P_i) + u_i \end{aligned}$$

where the coefficients β_0 and β_1 are the elasticity, faces the difficulty that $\mathbf{E}(f(P|U)) \neq 0$. Hence it turns necessary to to de-correlate the effects using an auxiliary variable which should be highly correlated with P but uncorrelated with the error term U. This variable is called an instrument.

Examples are numerous when studying the variation of price and demand. For example consider Q the annual sales of wheat and P the prices. An instrument could be in that case PL the rain level in the production region. It is obvious that the level of rain does not change the demand, hence $\mathrm{Corr}(PL, U) = 0$ while the lack of rain decreases the production which in turn increases the prices, so $\mathrm{Corr}(PL, \log P) \neq 0$.
However, to solve this practical example, the specific link between the covariates and the instrumental variable is required.

Assume that the link between X and the instrument W is of the form $X = \mathcal{L}(W, Z)$ with Z an independent random variable with distribution $\mathbf{P}_Z$. Then the operator has the following form

$$T\varphi(w) = \int \varphi \circ \mathcal{L}(w, Z) d\mathbf{P}_Z(Z) = \int \varphi(x) K_{\mathcal{L}}(x, w) dx$$

with a change of variable under some differentiability conditions on $\mathcal{L}$. Under technical assumptions, the operator defines a Fredholm integral operator with kernel $K_{\mathcal{L}}$ depending on the the link function $\mathcal{L}$ and the distribution of Z. Such operators are well studied in and, in many cases, the SVD decomposition will be available, which enables to use the estimation procedure developed in this paper.

As a practical example, one may be interested in the particular case, where X is uniform on $[0, 1]$ and $W = X + Z$ where Z is a random variable independent of X with unknown density g_Z. We point out that the model $Y_i = \varphi(W_i - Z_i) + U_i$ is also at the core of curve registration issues when curves are warped through random shifts W_i's.

In this example, both φ and g_Z are supposed to be 1-periodic. The conditional operator $T : L^2(X) \to L^2(W)$ can be written as

$$Tf(w) = \mathbf{E}(f(X)|W = w) = \mathbf{E}(f(w - Z)) = \int_0^1 f(w - z) g(z) dz,$$

with adjoint

$$T^{\star}h(x) = \mathbf{E}(h(W)|X = x) = \int_0^1 h(z + x)g(z)dz,$$

for all periodic functions f, g belonging respectively in $L^2(X)$ and $L^2(W)$. Hence, T is a deconvolution type operator, up to some change of variable. Let $(\phi_k)_{k\in\mathbb{N}}$ be the usual real trigonometric basis on $[0, 1]$:

$$\phi_1(t) \equiv 1,\ \phi_{2p}(t) = \sqrt{2}\cos(2\pi pt),\ \phi_{2p+1}(t) = \sqrt{2}\sin(2\pi pt),\ p \in \mathbb{N}.$$

Since X is uniform on $[0, 1]$, $(\phi_k)_{k\in\mathbb{N}}$ is an orthonormal basis of $L^2(X)$. With simple algebra, it is possible to prove that this sequence corresponds to the eigenvectors of $T^{\star}T$, see for instance [Cav08]. The corresponding eigenvalues are related to the Fourier coefficients of the density g_Z. The eigenvalues are obviously unknown but may be easily estimated using the procedure presented above.

Another example is obtained when considering the case where

$$(X, W) \sim \mathcal{N}\left(\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left(\begin{array}{cc} 1 & \rho \\ \rho & 1 \end{array}\right)\right),$$

Then, the corresponding operator T is a self adjoint operator with eigenvalues ρ^j and eigenfunctions the Hermite polynomials. This example is not covered by the procedure developed in this paper since the eigenvectors are exponentially decreasing (see assumption (27) bellow). Nevertheless, this proves that some assumptions on (X, W) may provide some hint on the corresponding SVD.

2. Adaptivity with complexity regularization

Model (4) is very similar to the well known model

$$y = T\varphi + v.$$

When the operator T is known, adaptive estimation can be achieved using penalized regularization technics as described in [LL10, LL08]. Assume that we are equipped with a sequence of nested linear subspaces whose union is dense in $\mathcal{Y}$, $Y_1 \subset Y_2 \ldots \subset Y_m \ldots \subset \mathcal{Y}$, with $dim(Y_m) = d_m$. We are interested in a subcollection of these spaces generated by a set of indices $\mathcal{M}_n$. In this paper, we will use these

approximation spaces as projection spaces in order to study the data. So, denote the projection of any space W over any subspace Z by $\Pi_Z W$. Let $\Pi^n_{Y_m}$ stands for the projection in the empirical norm. Set also the corresponding projected operator $T_m = \Pi^n_{Y_m} T$.
In this part, we impose some smoothness condition on the function to be estimated, namely

SC source condition:

There exists $\nu > 0$ such that $\varphi \in Range((T^*T)^\nu) := \mathcal{R}((T^*T)^\nu)$

This condition, well used in the field of inverse problems, links the smoothness of the function to the regularity of the operator. The relationships with other kind of regularity assumptions are described in [LR09].

Using a sieve of the space $\mathcal{Y}$, we consider the corresponding approximation spaces in the space $\mathcal{X}$, defined as $X_m = T^*_m \mathcal{Y}$. By construction

$$\Pi_{X_m} = (\Pi^n_{Y_m} T)^+ \Pi^n_{Y_m} T.$$

Hereafter, we consider a class of regularized estimators built using a projection and a regularization procedure. Hence the first step is to project the data onto a well chosen space. Namely let Y_{m_0} be a big enough space in the sense that m_0 is such that

$$\|(I - \Pi_{X_{m_0}})\varphi\| \leq \inf_{m \in \mathcal{M}_n} [\|(I - \Pi_{X_m})\varphi\| + \sqrt{\frac{d_m}{n} \frac{1}{\gamma_m}}],$$

with

$$\gamma_m := \inf_{v \in Y_m, \|g\|=1} \|T^*_m g\|,$$

which expresses the effect of operator T^*_m over the approximating subspace Y_m. This quantity can be chosen so as not to depend on the unknown regularity of the solution φ, but only on the ill-posedness of the inverse problem, namely $\gamma_m = O(d_m^{-t})$ as shown in [LL08, LL10]. This bound leads to the usual optimal rate of convergence for inverse problems. Under assumption **SC** the above inequality is satisfied if the dimension of the set is such that

$$d_{m_0}^{2\nu t} \geq n^{\frac{2\nu t}{4\nu t + 2t + 1}}.$$

Thus it is enough to choose m_0 such that $d_{m_0} \geq n \geq n^{1/(2t+1)}$.

The second step is obtained by, for $\mathcal{K}_n$ a set of indices, considering $\{\tilde{R}_k,\ k \in \mathcal{K}_n\}$ a collection of regularization operators which depend on different values of the smoothing parameters. For instance consider Tikhonov regularization operators which rely on the choice of a smoothing sequence, Landweber iteration operators which rely on the choice of a stopping index, or other general smoothing operators described in [EHN96]. Consider the corresponding estimators

$$\hat{\varphi}_k := \tilde{R}_k \Pi^n_{Y_{m_0}} y = R_k y, \tag{7}$$

where we have written $R_k := \tilde{R}_k \Pi^n_{Y_{m_0}}$. The behavior of such general estimators depends on the choice of the regularization sequence. From the theory of inverse problems, we know that it is possible to choose a regularization operator for which the corresponding estimator achieves the optimal rate of convergence, but this choice depends on ν defined in **SC**, which characterizes the regularity of the solution.

Our aim is building a method that picks, according to the data, an optimal R_k, among all the R_k, $k \in \mathcal{K}_n$ in such a way that optimal rates are maintained. This choice must also not depend on a priori regularity assumptions. We point out that selecting the optimal smoothing parameter in a collection of sequences, belongs to model selection theory since it is equivalent as selecting a good model among a collection of sets.

For this consider the following penalized procedure. For a given constant $r > 2$ and weights L_k, $k \in \mathcal{K}_n$ to be chosen, define the penalty as

$$\mathrm{pen}(k) := r\sigma^2(1 + L_k)[Tr(R_k^t R_k) + \rho^2(R_k)],$$

where $Tr(R_k^t R_k)$ is the trace and $\rho(R_k^t R_k) = \rho^2(R_k)$ is the spectral radius. Finally $\hat{k}$ is selected as the solution of

$$\hat{k} := \arg\min_{k \in \mathcal{K}_n} \left\{ \|R_k(y - T(\hat{\varphi}_k))\|^2 + \mathrm{pen}(k) \right\}, \tag{8}$$

which defines the estimator $\hat{\varphi}_{\hat{k}} = R_{\hat{k}} y$. Let $R_k T\varphi$ be the regularized true function, which measures the accuracy of the estimation procedure without observation noise. The following result states the asymptotic behaviour of the estimator $\hat{\varphi}_{\hat{k}}$.

Theorem 2.1. *Under some technical conditions, there exists a constant C which depends on r and on T, such that the following inequality holds true*

(9)
$$\mathbf{E}\|\hat{\varphi}_{\hat{k}}-\varphi\|^2 \leq 2\|(I-\Pi_{X_{m_0}})\varphi\|^2+C\inf_{k\in\mathcal{K}_n}\left[\|R_kT\varphi-\varphi\|^2+2\mathrm{pen}(k)\right]+\frac{\Sigma(d)}{n},$$

where we have set

$$\Sigma(d)=\sum_{k\in\mathcal{K}_n}2\left[\sqrt{\frac{dTr(R_k^tR_k)}{\rho^2(R_k)}}+1\right][\frac{d}{\rho^2(nR_k)}]^{-1}e^{-\sqrt{dL_k[Tr(R_k^tR_k)+\rho^2(R_k)]/\rho^2(R_k)}},$$

for d properly chosen.

Hence, the estimator is optimal in the sense that the adaptive estimator achieves the best rate of convergence among all the regularized estimators, up to an error of order $\mathrm{pen}(k)$ and $\Sigma(d)/n$. This bound is non asymptotic and the rate of convergence depends on both previous terms.

We also point out that $\rho^2(nR_k)$ and $\mathrm{Tr}(R_k^tR_k)/\rho^2(R_k)$ do not depend on n.

The main ingredients of the proof can be found in [LL08, LL10].

When the operator T is unknown, one could be tempted by using the same ideas, just replacing T by an estimator $\hat{T}$. However, the whole procedure turns more difficult since the term in $R_k\hat{T}$ can not be bounded as easily as previously. Recent results on concentration for random matrices provide some hopes to extend this general adaptive procedure to these cases but work is still under progress. However, in the following section, we provide a general methodology built using the SVD decomposition of the operator.

3. An oracle inequality with partially known SVD

In this part, we assume that the SVD is partially known in the sense that the basis of eigenvectors (ϕ_j) is known but that the eigenvalues, λ_j's, are not observed. This assumption, yet restrictive, still enables to handle some useful cases. It will be discussed in details at the end of section 4.

3.1. **General estimation approach.** This case is inspired by the pioneering work by [CH05]. It is fully described in [LM09]

We can write the following decompositions

$$r(w) = \mathbf{E}(Y|W = w) = T\varphi(w) = \sum_{j \geqslant 1} \lambda_j \langle \varphi, \phi_j \rangle_X \psi_j(w), \tag{10}$$

$$\text{and} \quad r(w) = \sum_{j \geqslant 1} r_j \psi_j(w), \tag{11}$$

with $r_j = \langle Y, \psi_j \rangle_W$ that can thus be estimated by

$$\hat{r}_j = \frac{1}{n} \sum_{i=1}^{n} Y_i \psi_j(W_i).$$

Hence the noisy observations are the $\hat{r}_j$'s which will be used to estimate the regression function φ in an inverse problem framework.

Note first that, if the operator were known we could provide an estimator using the spectral decomposition of the function φ as follows. For a given decomposition level m, define the projection estimator (also called spectral cut-off [EHN96])

$$\hat{\varphi}_m^0 = \sum_{j=1}^{m} \frac{\hat{r}_j}{\lambda_j} \phi_j \tag{12}$$

Since the λ_j's are unknown, our first task is to build an estimator of the eigenvalues. For this, using the decomposition (10), we obtain

$$\begin{aligned} \lambda_j =& < T\phi_j, \psi_j >_W = \mathbf{E}[T\phi_j(W)\psi_j(W)] \\ &= \mathbf{E}[\mathbf{E}[\phi_j(X)|W]\psi_j(W)] \\ &= \mathbf{E}[\phi_j(X)\psi_j(W)]. \end{aligned} \tag{13}$$

So, following (13), a natural estimator for the eigenvalue λ_j is given by

$$\hat{\lambda}_j = \frac{1}{n}\sum_{i=1}^{n} \psi_j(W_i)\phi_j(X_i). \tag{14}$$

As studied in [CH05], replacing directly the eigenvalues by their estimates in (12) does not yield a consistent estimator, hence using their same strategy we define an upper bound for the resolution level

$$M = \inf\left\{k \leqslant N : |\hat{\lambda}_k| \leqslant \frac{1}{\sqrt{n}}\log n\right\} - 1, \tag{15}$$

for N any integer chosen greater than n. The parameter N provides an upper bound for M in order to ensure that M is not too large. The main idea behind this definition is that when the estimates of the eigenvalues are too small with respect to the observation noise, trying to still provide an estimation of the inverse λ_k^{-1} only amplificates the estimation error. To avoid this trouble, we truncate the sequence of the estimated eigenvalues when their estimate is too small, i.e smaller than the noise level. We point out that this parameter M is a random variable which we will have to control. More precisely, define two deterministic lower and upper bounds M_0, M_1 as

$$M_0 = \inf\left\{k : |\lambda_k| \leqslant \frac{1}{\sqrt{n}}\log^2 n\right\} - 1, \tag{16}$$

and

$$M_1 = \inf\left\{k : |\lambda_k| \leqslant \frac{1}{\sqrt{n}}\log^{3/4} n\right\}, \tag{17}$$

we can show that with high probability $M_0 \leqslant M < M_1$ as proved in Lemma 5.1. Note that if in the definition (15) the set is empty, we set $M = 0$. However, from the remark above, this case happens with very small probability.

Now, thresholding the spectral decomposition in (12) leads to the following estimator

$$\hat{\varphi}_m = \sum_{j=1}^{m} \frac{\hat{r}_j}{\hat{\lambda}_j} 1_{j \leqslant M} \phi_j. \tag{18}$$

The asymptotic behaviour of this estimate depends on the choice of m. In the next section, we provide an optimal procedure to select the parameter m that gives rise to an adaptive estimator $\varphi^\star$ and an oracle inequality.

3.2. **Oracle inequality.**

All the estimation errors will be given with respect to the L^2_X norm which is a natural choice for this kind of problems. Another possibility would have been to place the issue in $L^2([0,1])$.

First, let $R_0(m, \varphi)$ be the quadratic estimation risk for the naive estimator $\hat{\varphi}^0_m$ (12), defined for all $m \in \mathbb{N}$, by

$$\begin{aligned} R_0(m,\varphi) &= \mathbf{E}\|\hat{\varphi}^0_m - \varphi\|^2_X \\ &= \sum_{k>m} \varphi_k^2 + \frac{1}{n}\sum_{k=1}^{m} \lambda_k^{-2}\sigma_k^2, \ \forall m \in \mathbb{N}, \end{aligned}$$

with $\sigma_k^2 = \mathrm{Var}(Y\psi_k(W))$. The best model would be obtained by choosing a minimizer of this quantity, namely

$$m_0 = \arg\min_m R_0(m, \varphi). \tag{19}$$

This risk depends on the unknown function φ hence m_0 is referred to as the oracle. We aim at constructing an estimator of $R_0(m, \varphi)$ which, by minimization, could give rise to a convenient choice for m, i.e as close as possible to m_0. The first step would be to replace φ_k

by their estimates $\hat{\lambda}_k^{-1}\hat{r}_k$ and take for estimator of σ_k^2, $\hat{\sigma}_k^2$, defined by

$$\begin{aligned}\hat{\sigma}_k^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(Y_i\psi_k(W_i) - \frac{1}{n}\sum_{i=1}^{n} Y_i\psi_k(W_i)\right)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(Y_i\psi_k(W_i) - \hat{r}_k\right)^2.\end{aligned}$$

This would lead us to consider the empirical risk for any $m \leqslant M$, the cut-off which warrants a good behaviour for the $\hat{\lambda}_j$'s

$$U_0(m, r, \lambda) = -\sum_{k=1}^{m} \hat{\lambda}_k^{-2}\hat{r}_k^2 + \frac{c}{n}\sum_{k=1}^{m} \hat{\lambda}_k^{-2}\hat{\sigma}_k^2, \ \forall m \in \mathbb{N},$$

for a well chosen constant c. The corresponding random oracle within the range of models which are considered would be

$$m_1 = \arg\min_{m \leqslant M} R_0(m, \varphi). \tag{20}$$

Unfortunately, the correlation between the errors V_i and the observations Y_i prevents an estimator defined as a minimizer of $U_0(m, r, \lambda)$ to achieve the quadratic risk $R_0(m, \varphi)$. Indeed, we have to use a stronger penalty, leading to an extra error in the estimation that shall be discussed later in the paper. More precisely, c in the penalty is not a constant anymore but is allowed to depend on the number of observations n.

Hence, now define $R(m, \varphi)$ the penalized estimation risk as

$$R(m, \varphi) = \sum_{k>m} \varphi_k^2 + \frac{\log^2 n}{n}\sum_{k=1}^{m} \lambda_k^{-2}\sigma_k^2, \ \forall m \in \mathbb{N}. \tag{21}$$

The best choice for m would be a minimizer of this quantity, which yet depends on the unknown regression function φ. Hence, to mimic this risk, define the following empirical criterion

$$U(m, r, \lambda) = -\sum_{k=1}^{m} \hat{\lambda}_k^{-2}\hat{r}_k^2 + \frac{\log^2 n}{n}\sum_{k=1}^{m} \hat{\lambda}_k^{-2}\hat{\sigma}_k^2, \ \forall m \in \mathbb{N}. \tag{22}$$

Then, the best estimator is selected by minimizing this quantity as follows

$$m^\star := \arg\min_{m\leqslant M} U(m,r,\lambda), \tag{23}$$

Finally, the corresponding adaptive estimator $\varphi^\star$ is defined as:

$$\varphi^\star = \sum_{k=1}^{m^\star} \hat{\lambda}_k^{-1}\hat{r}_k\phi_k. \tag{24}$$

The performances of $\varphi^\star$ are presented in the following theorem.

Theorem 3.1. *Let $\varphi^\star$ the projection estimator defined in (24). Then, there exist B_0, B_1, B_2 and τ positive constants independent of n such that:*

$$\begin{aligned}\mathbb{E}\|\varphi^\star-\varphi\|_X^2 \;\; &\leqslant \;\; B_0\log^2(n).\left[\inf_m R(m,\varphi)\right] + \frac{B_1}{n}\left(\log(n).\|\varphi\|_X^2\right)^{2t} \\ &\qquad + \Omega + \log^2(n).\Gamma(\varphi),\end{aligned}$$

where $\Omega \leqslant B_2(1+\|\varphi\|_X^2)\exp\left\{-\log^{1+\tau} n\right\}$, m_0 denotes the oracle bandwidth and

$$\Gamma(\varphi) = \sum_{k=\min(M_0,m_0)}^{m_0} \left[\varphi_k^2 + \frac{1}{n}\lambda_k^{-2}\sigma_k^2\right], \tag{25}$$

with the convention $\sum_a^b = 0$ if $a = b$.

We obtain a non asymptotic inequality which guarantees that the estimator achieves the optimal bound, up to a logarithmic factor, among all the estimators that could be constructed. We point out that we lose a $\log^2(n)$ factor when compared with the bound obtained in [CH05]. This loss comes partly from the fact that the error on the operator is not deterministic nor even due to a independent noisy observation of the eigenvalues. Here, the λ_k's have to be estimated using the available data by $\hat{\lambda}_k$. In the econometric model, both the operator and the regression function are estimated on the same sample, which leads to high correlation effects that are made explicit in Model (4), hampering the rate of convergence of the corresponding estimator.

An oracle inequality only provides some information on the asymptotic behaviour of the estimator if the remainder term $\Gamma(\varphi)$ is of smaller order than the risk of the oracle. This remainder term models the error made when truncating the eigenvalues, i.e the error when selecting a model close to the random oracle $m_1 \leqslant M$ and not the true oracle m_0. In the next section, we prove that, under some assumptions, this extra term is smaller than the risk of the estimator.

Proof. The full proof of this result can be found in [LM09]. We provide here the general ideas. First, the decay of the eigenvalues and of the estimated eigenvalues is controlled in probability as follows. Set $\mathcal{M} = \{M_0 \leqslant M < M_1\}$, where M, M_0, M_1 are respectively defined in (15), (16) and (17). Then, for all $n \geqslant 1$,

$$P(\mathcal{M}^c) \leqslant C M_0 e^{-\log^{1+\tau} n},$$

where C and τ denote positive constants independent of n, as proved in Lemma 5.1.
Then, the proof of our main result can be decomposed into four steps. In a first time, we prove that the quadratic risk of $\varphi^\star$ is close, up to some residual terms, to $\mathbf{E}\bar{R}(m^\star, \varphi)$ where

$$\bar{R}(m,\varphi) = \sum_{k>m} \varphi_k^2 + \frac{\log^2 n}{n} \sum_{k=1}^{m} \hat{\lambda}_k^{-2} \sigma_k^2, \ \forall m \in \mathbb{N}. \tag{26}$$

This result is uniform in m and justifies our choice of $\bar{R}(m,\varphi)$ as a criterion for the bandwidth selection.

In a second time, we show that $\mathbf{E}\bar{R}(m^\star, \varphi)$ and $\mathbf{E}U(m^\star, r, \varphi)$ are in some sense comparable. Then, according to the definition of $m^\star$ in (23),

$$U(m^\star, r, \varphi) \leqslant U(m, r, \varphi), \forall m \leqslant M.$$

We will conclude the proof by proving that for all $m \leqslant M$, $\mathbf{E}U(m, r, \varphi) = \mathbf{E}\|\hat{\varphi}_m - \varphi\|^2$, up to a log term and some residual terms.

Some additional assumptions are required on both the data Y_i, $i = 1, \dots, n$ and the eigenfunctions ϕ_k and ψ_k for $k \geqslant 1$.

Bounded SVD functions:: There exists a finite constant C_1 such that

$$\forall j \geqslant 1, \quad \|\phi_j\|_\infty < C_1, \quad \|\psi_j\|_\infty < C_1 \tag{27}$$

Requiring bounded SVD functions may be seen as a restrictive condition. Yet it is met when the eigenvectors are trigonometric functions. However, this condition can be also be turned into a moment condition if we replace the concentration bound by a Bernstein type inequality. Note also that the moment conditions on Y amounts to require a bounded regression function φ and equivalent moment conditions on the errors U_j.

Enough ill-posedness : : Let $\sigma_j^2 = \mathrm{Var}(Y\psi_j(W))$. We assume that there exist two positive constants σ_L^2 and σ_U^2 such that

$$\forall j \geqslant 1, \quad \sigma_L^2 \leqslant \sigma_j^2 \leqslant \sigma_U^2. \tag{28}$$

Note that Condition (6) implies the upper bound of Condition (28); which is also a direct consequence of Assumption A.2 in [HH05]. Both the upper and lower bound is similar to the assumption 4.1 and the variance condition in Assumption 3.1 in [CR08]. We also point out that this condition is not needed when building an estimator for the regression function. However it turns necessary when obtaining the lower bound to get a minimax result, or when obtaining an oracle inequality. □

3.3. **Rate of convergence.** To get a rate of convergence for the estimator, we need to specify the regularity of the unknown function φ and compare it with the degree of ill-posedness of the operator T, following the usual conditions in the statistical literature on inverse problems, see for example [MR96] or [CT02], [BHMR07] for some examples.

Regularity Condition: Assume that the function φ is such that there exists s and a constant C such that

$$\sum_{k \geqslant 1} k^{2s} \varphi_k^2 < C \tag{29}$$

This Assumption corresponds to functions whose regularity is governed by the smoothness index s. This parameter is unknown and yet governs the rate of convergence. In the special cases where the eigenfunctions are the Fourier basis, this set corresponds to Sobolev classes. We prove that our estimator achieves the optimal rate of convergence without prior assumption on s.

Corollary 3.2. *Let $\varphi^\star$ be the model selection estimator defined in* (24)*. Then, under the Sobolev embedding assumption* (29)*, we get the following rate of convergence*

$$\mathbb{E}\|\varphi^\star - \varphi\|_X^2 = O\left(\left(\frac{n}{\log^{2\gamma} n}\right)^{\frac{-2s}{2s+2t+1}}\right),$$

with $\gamma = 2 + 2s + 2t$.

We point out that $\varphi^\star$ is constructed without prior knowledge of the unknown regularity s of φ, yet achieving the optimal rate of convergence, up to some logarithmic terms. In this sense, our estimator is said to be asymptotically adaptive. The rate we obtain is similar to the minimax rate obtained in [CR08]. Following these previous authors, we point out that Hall and Horowitz in [HH05] also obtain another minimax optimal rate of convergence in a similar settings but under different regularity assumptions.

Remark 3.3. In an equivalent way, we could have imposed a super smooth assumption, on the function φ, i.e assuming that for given γ, t and constant C,

$$\sum_{k=1}^{\infty} \exp(2\gamma k^t)\varphi_k^2 < C.$$

Following the guidelines of the proof of Corollary 3.2 and Theorem 2.1, we obtain that $M_0 > m_0 \sim (a2\gamma \log n)^{1/t}$ with $2a\gamma > 1$, leading to the optimal recovery rate for super smooth functions in inverse problems.

4. Conclusion and comments

In conclusion, this work shows that provided the eigenvectors are known, for smooth functions φ, estimating the eigenvalues and using a threshold suffices to get a good estimator of the regression function in the instrumental variable framework. The price to pay for not knowing the operator is only an extra $\log^2 n$ with respect to usual inverse problems and is only due to the correlation induced by the V_i's. Remark that this log term could be avoided by splitting the data. One may use a training set for the construction of the bandwidth $m^\star$ and the remaining data for the recovery of φ. In this case, the quadratic risks of both our estimator and the oracle are comparable, up to some computable constant. Nevertheless, this approach is not satisfying from a mathematical point of view since the underlying problem of adaptation is hidden.

One could object that the knowledge of the eigenvectors is a huge hint and thus, the operator is not totally unknown. Still, in the following examples, we present a class of cases where this situation happens, mainly when the relationship between the variable X and the instrument W has a particular form. However, some papers have considered the case of completely unknown operators, using functional approach, see for instance [CFR06], but their estimate clearly rely on smoothness assumptions for the regression. Hence the two approaches are complementary since we provide more refined adaptive result under stronger assumptions. Nevertheless, using similar techniques to develop a fully adaptive estimation procedure would be a next step toward a full understanding of the IV regression model.

To our knowledge, we provide the first adaptive estimation procedure for IV regression in some particular cases which yet present some interest from an econometric point of view. We are aware that we do not handle the estimation problem in the general case but this work only claims to be a first step towards an adaptive estimation procedure for this difficult problem.

5. Appendix

Lemma 5.1. *Set $\mathcal{M} = \{M_0 \leqslant M < M_1\}$, where M, M_0, M_1 are respectively defined in (15), (16) and (17). Then, for all $n \geqslant 1$,*

$$P(\mathcal{M}^c) \leqslant CM_0 e^{-\log^{1+\tau} n},$$

where C and τ denote positive constants independent of n.

PROOF. It is easy to see that:

$$P(\mathcal{M}^c) = P\left(\{M < M_0\} \cup \{M \geqslant M_1\}\right) \leqslant P(M < M_0) + P(M \geqslant M_1).$$

Using (15) and (17),

$$P(M \geqslant M_1) = P\left(\bigcap_{k=1}^{M_1} \left\{|\hat{\lambda}_k| \geqslant \frac{1}{\sqrt{n}} \log n\right\}\right) \leqslant P\left(|\hat{\lambda}_{M_1}| \geqslant \frac{1}{\sqrt{n}} \log n\right).$$

The definition of $\hat{\lambda}_{M_1}$ yields

$$\begin{aligned}
P(M \geqslant M_1) &\leqslant P\left(\left|\hat{\lambda}_{M_1} - \lambda_{M_1} + \lambda_{M_1}\right| \geqslant \frac{1}{\sqrt{n}} \log n\right), \\
&\leqslant P\left(\left|\hat{\lambda}_{M_1} - \lambda_{M_1}\right| \geqslant \frac{1}{\sqrt{n}} \log n - |\lambda_{M_1}|\right), \\
&\leqslant P\left(\left|\frac{1}{n}\sum_{i=1}^{n} \phi_{M_1}(X_i)\psi_{M_1}(W_i) - \mathbf{E}[\phi_{M_1}(X)\psi_{M_1}(W)]\right| \geqslant b_n\right),
\end{aligned}$$

where $b_n = n^{-1/2}\log n - |\lambda_{M_1}|$ for all $n \in \mathbb{N}$. Let $k \in \mathbb{N}$ and $x \in [0,1]$ be fixed. Assumption (27) and Hoeffding inequality yield

$$\begin{aligned}
P(|\hat{\lambda}_k - \lambda_k| > x) &\leqslant 2\exp\left\{-\frac{(nx)^2}{2\sum_{i=1}^n \mathrm{Var}(\phi_{M_1}(X_i)\psi_{M_1}(W_i)) + 2nCx/3}\right\}, \\
&= 2\exp\left\{-\frac{nx^2}{2\mathrm{Var}(\phi_{M_1}(X)\psi_{M_1}(W)) + 2Cx/3}\right\}.
\end{aligned}$$

Using again the assumption (27) on the bases $(\phi_k)_{k\in\mathbb{N}}$ and $(\psi_k)_{k\in\mathbb{N}}$,

$$\mathrm{Var}(\phi_{M_1}(X)\psi_{M_1}(W)) \leqslant \mathbb{E}[\phi_{M_1}^2(X)\psi_{M_1}^2(W)] \leqslant C_1^4.$$

Hence,

$$P(|\hat{\lambda}_k - \lambda_k| > x) \leqslant 2\exp\left(-Cnx^2\right), \ \forall x \in [0,1], \tag{30}$$

for some constant C depending on C_1 but independent of n. Using (17), we obtain $1 > b_n > 0$ for all $n \in \mathbb{N}$. Therefore, using (30) with $x = b_n$, we obtain:

$$\begin{aligned} P(M \geqslant M_1) \leqslant 2\exp\left\{-Cnb_n^2\right\} & \leqslant 2\exp\left\{-C(\log n - \log^{3/4} n)^2\right\}, \\ & \leqslant C\exp\left\{-\log^{1+\tau} n\right\}, \end{aligned}$$

where C and τ denote positive constants independent of n.

The bound of $P(M < M_0)$ follows the same lines:

$$\begin{aligned} P(M < M_0) = P\left(\bigcup_{j=1}^{M_0}\left\{|\hat{\lambda}_j| \leqslant \frac{\log n}{\sqrt{n}}\right\}\right) & \leqslant \sum_{j=1}^{M_0} P\left(|\hat{\lambda}_j| \leqslant \frac{\log n}{\sqrt{n}}\right), \\ & \leqslant \sum_{j=1}^{M_0} P\left(\hat{\lambda}_j \leqslant \frac{\log n}{\sqrt{n}}\right). \end{aligned}$$

Let $j \in \{1, \ldots, M_0\}$ be fixed.

$$P\left(\hat{\lambda}_j \leqslant \frac{\log n}{\sqrt{n}}\right) = P\left(\hat{\lambda}_j - \lambda_j \leqslant \tilde{b}_{n,j}\right),$$

where $\tilde{b}_{n,j} = n^{-1/2}\log n - \lambda_j$ for all $n \in \mathbb{N}$. Thanks to (16), $\tilde{b}_{n,j} < 0$ for all $n \in \mathbb{N}$. Using (30) with $x = -\tilde{b}_{n,j}$, we get

$$P\left(\hat{\lambda}_j \leqslant \frac{\log n}{\sqrt{n}}\right) \leqslant \exp\left\{-Cn\tilde{b}_{n,j}^2\right\} \leqslant C\exp\left\{-\log^{1+\tau} n\right\},$$

for some $C, \tau > 0$. This concludes the proof of Lemma 5.1.

□

Proof of Corollary 3.2 We start by recalling the oracle inequality obtained for the estimator $\varphi^\star$.

$$\begin{aligned} \mathbb{E}\|\varphi^\star - \varphi\|^2 & \leqslant C_0\log^2(n).\left[\inf_m R(m,\varphi)\right] + \frac{C_1}{n}\left(\log(n).\|\varphi\|^2\right)^{2\beta} \\ & \quad +\Omega + \log^2(n).\Gamma(\varphi), \end{aligned}$$

We have to bound the risk under the regularity condition and the extra term $\log^2(n)\Gamma(\varphi)$. Recall that the risk is given by

$$R(m,\varphi) = \sum_{k>m} \varphi_k^2 + \frac{\log^2 n}{n} \sum_{k=1}^{m} \lambda_k^{-2} \sigma_k^2.$$

Hence under (29), we obtain both upper bounds for two constants C_1 and C_2

$$\sum_{k>m} \varphi_k^2 \leqslant m^{-2s} C_1,$$

$$\frac{\log^2 n}{n} \sum_{k=1}^{m} \lambda_k^{-2} \sigma_k^2 \leqslant C_2 \frac{\log^2 n}{n} \sigma_U^2 m^{2t+1}.$$

An optimal choice is given by $m = [(n/\log n)^{\frac{1}{1+2s+2t}}]$, leading to the desired rate of convergence.

Now consider the remainder term $\Gamma(\varphi)$. Under Assumption [IP], $M_0 \geqslant [n^{1/2s}/\log^2 n]$, but since $m_0 = [n^{\frac{1}{1+2s+2t}}]$ we get clearly that $m_0 \leqslant M_0$, which entails that $\Gamma(\varphi) = 0$.

References

[BHMR07] N. Bissantz, T. Hohage, A. Munk, and F. Ruymgaart. Convergence rates of general regularization methods for statistical inverse problems and applications. *SIAM J. Numer. Anal.*, 45(6):2610–2636 (electronic), 2007.

[Cav08] L. Cavalier. Nonparametric statistical inverse problems. *Inverse Problems*, 24(3):034004, 19, 2008.

[CFR06] M. Carrasco, J-P. Florens, and E. Renault. *Linear Inverse Problems in Structural Econometrics: Estimation Based on Spectral Decomposition and Regularization*, volume 6. North Holland, 2006.

[CGPT02] L. Cavalier, G. K. Golubev, D. Picard, and A. B. Tsybakov. Oracle inequalities for inverse problems. *Ann. Statist.*, 30(3):843–874, 2002. Dedicated to the memory of Lucien Le Cam.

[CH05] L. Cavalier and N. W. Hengartner. Adaptive estimation for inverse problems with noisy operators. *Inverse Problems*, 21(4):1345–1361, 2005.

[CHR03] A. Cohen, M. Hoffmann, and M. Reiss. Adaptive wavelet galerkin methods for linear inverse problems. *SIAM*, 1(3):323–354, 2003.

[CR08] X. Chen and M. Reiss. On rate optimality for ill-posed inverse problems in econometrics. *Cowles Foundation Discussion Paper No. 1626*, 2008.

[CT02] L. Cavalier and A. Tsybakov. Sharp adaptation for inverse problems with random noise. *Probab. Theory Related Fields*, 123(3):323–354, 2002.

[EHN96] H. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.

[EK01] S. Efromovich and V. Koltchinskii. On inverse problems with unknown operators. *IEEE Trans. Inform. Theory*, 47(7):2876–2894, 2001.

[HH05] P. Hall and J. L. Horowitz. Nonparametric methods for inference in the presence of instrumental variables. *Ann. Statist.*, 33(6):2904–2929, 2005.

[HR08] M. Hoffmann and M. Reiss. Nonlinear estimation for linear inverse problems with error in the operator. *Ann. Statist.*, 36(1):310–336, 2008.

[LL08] J-M. Loubes and C. Ludena. Adaptive complexity regularization for inverse problems. *Electronic Journal Of Statistics*, 2:661–677, 2008.

[LL10] J-M. Loubes and C. Ludena. Adaptive complexity regularization for inverse problems. *ESAIM Probab. Statist.*, 2:661–677, 2010.

[LM09] Jean-Michel Loubes and Clément Marteau. Oracle Inequality for Instrumental Variable Regression. hal-00356428, 2009. 62G05; 62G20.

[LR09] Jean-Michel Loubes and Vincent Rivoirard. Review of rates of convergence and regularity conditions for inverse problems. *Int. J. Tomogr. Stat.*, 11(W09):61–82, 2009.

[Mar06] C. Marteau. Regularization of inverse problems with unknown operator. *Math. Methods Statist.*, 15(4):415–443, 2006.

[Mar09] C. Marteau. On the stability of the risk hull method. *Journal of Statistical Planning and Inference*, (139):1821–1835, 2009.

[MR96] B. Mair and F. Ruymgaart. Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.*, 56(5):1424–1444, 1996.

[NP03] W. K. Newey and J. L. Powell. Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578, 2003.

[O'S86] F. O'Sullivan. A statistical perspective on ill-posed inverse problems. *Statist. Sci.*, 1(4):502–527, 1986. With comments and a rejoinder by the author.

Equipe de probabilités et statistique, Institut de Mathématique de Toulouse, UMR5219, Université de Toulouse, 31000 Toulouse France

E-mail address: Jean-Michel.Loubes@math.univ-toulouse.fr

Equipe de probabilités et statistique, Institut de Mathématique de Toulouse, UMR5219, Université de Toulouse, 31000 Toulouse France

E-mail address: Clement.Marteau@math.univ-toulouse.fr

SEMISIMPLE HOPF ALGEBRAS AND THEIR REPRESENTATIONS

SONIA NATALE

Abstract. We review some recent results and constructions concerning fusion categories, focusing on the classification problem of semisimple Hopf algebras. We discuss several invariants of these structures, as well as some open questions.

Contents

1.	Introduction	123
2.	Semisimple Hopf algebras and fusion categories	125
2.1.	$\operatorname{Rep} H$ as a tensor category	125
3.	Hopf algebra extensions	129
3.1.	Abelian extensions	133
4.	Extensions of fusion categories by finite groups	135
4.1.	G-extensions	136
4.2.	G-equivariantization	137
4.3.	G-actions on $\operatorname{Rep} H$ and cocentral extensions	138
4.4.	Weakly group-theoretical fusion categories	140
5.	Exact sequences of fusion categories	146
5.1.	Extensions and Hopf monads	148
5.2.	Extensions and commutative central algebras	149
5.3.	Some examples of $\operatorname{Rep} G$-extensions	151
5.4.	Exact sequences in the braided context	152
6.	Some invariants of a fusion category	153
6.1.	Grothendieck ring	154

2000 *Mathematics Subject Classification.* 16T05; 17B37.

Partially supported by Alexander von Humboldt Foundation, ANPCyT, CONICET, SeCYT (UNC) and FAMAF (República Argentina).

6.2. Module categories 156
6.3. Frobenius-Schur indicators 157
6.4. Exponent 157
7. Some further questions 159
References 161

1. Introduction

The most basic examples of Hopf algebras are the algebras of (regular, etc.) functions on a (algebraic, finite, etc.) group, and the enveloping algebras of Lie algebras. An important class of noncommutative examples is provided by the quantum groups of Drinfeld and Jimbo [Dr] and their generalizations, which since their discovery were a source of new developments in the theory of Hopf algebras.

Semisimple Hopf algebras can be thought of as noncommutative generalizations of finite groups. A finite dimensional commutative Hopf algebra is necessarily semisimple and isomorphic to the algebra of functions on a finite group, that is, to the dual of a group algebra.

The structure of Hopf algebras is naturally related to the study of symmetries of distinct mathematical objects: quasitriangular Hopf algebras constitute a tool for the systematic construction of solutions of the Yang-Baxter equation; in topology, they are related to the construction of knots and 3-manifolds; in the theory of operator algebras, certain semisimple Hopf algebras appear as invariants or "Galois groups" in the study of inclusions of subfactors. One of the main features of Hopf algebras is that they give rise to tensor categories through their categories of representations.

Let k be an algebraically closed field of characteristic zero.

Recall that a finite tensor category over k is a k-linear abelian rigid monoidal category $\mathcal{C}$ such that Hom spaces are finite dimensional and objects have finite length, and such that the tensor product $\otimes : \mathcal{C} \times \mathcal{C} \to \mathcal{C}$ is k-bilinear and the unit object $\mathbf{1}$ is simple. A fusion category over k is a semisimple tensor category with finitely many isomorphism classes of simple objects. This kind of categories possess remarkable properties of symmetry that generalizes that of finite groups. For results on monoidal, tensor and fusion categories and related structures, the reader can see [Ks, Mj, ENO, Mg3] and references therein.

The representation category $\mathcal{C} = \operatorname{Rep} H$ of a finite dimensional semisimple (quasi)-Hopf algebra H is a fusion category. Such fusion categories are characterized by the fact that they are endowed with a (quasi)*fiber functor* $\mathcal{C} \to \mathrm{Vec}_k$, that is, a k-linear exact (quasi)tensor functor, to the category Vec_k of finite dimensional vector spaces over k.

Several known classes of non-trivial semisimple Hopf algebras are obtained by a process of *extension* from smaller (often trivial) ones. Underlying the study of extensions of Hopf algebras, there is an interesting cohomological description, very important in the classification problem [Ka, M3].

Several classification results for semisimple Hopf algebras are known. We refer the reader to the papers [Mo3, Mo4, M5, Bu2] for more information on this subject.

In view of a result of A. Masuoka [M2], if $\dim H = p^n$, with p a prime number, then H contains a non-trivial central group-like element: this reduces the classification in these dimensions to the problem of classifying extensions. In a sequel of papers, Masuoka gave also the classification in dimension p^2, p^3, 6, 8 and 18. The full classification is also known in dimension pq ($p \neq q$, prime numbers; indeed, there are no non-trivial examples in this case), after results of Etingof, Gelaki and Westreich. The classification in dimension pq^2 and pqr, where p, q, r are distinct prime numbers, has been recently completed by Etingof, Nikshych and Ostrik in [ENO2]. For Kac algebras (C^*-semisimple Hopf algebras, after G. I. Kac), related to the theory of subfactors, several classification results in low dimension where obtained by Izumi and Kosaki [IK].

In this note we present an overview of some recent results and constructions concerning fusion categories, focusing on the classification problem of semisimple Hopf algebras. We separate our exposition distinguishing three main streams: Hopf algebra extensions, extensions of fusion categories by finite groups as studied in [GNi, ENO2, DGNO], and a recent notion of exact sequence of fusion categories developed in [BrN]. We put emphasis in showing the connections between these different notions, as well as their rôle in classification questions.

We also discuss in Section 6 several invariants of these structures: Grothendieck rings, module categories, Frobenius-Schur indicators and exponents. Without being exhaustive in our references, we mention some known results about these invariants, specially in relation with the main examples discussed in the paper.

In the final section, we present a series of open questions regarding the concepts discussed previously.

Throughout, k will denote an algebraically closed field of characteristic zero. The symbols Hom, $\otimes$, etc., will mean Hom_k, $\otimes_k$, etc. Our references for the theory of Hopf algebras are [Mo, Sc3].

For an algebra A, we shall use the notation $\operatorname{Rep} A$ to indicate the category of finite dimensional (left) A-modules. Similarly, for a coalgebra C, the notation $\operatorname{Corep} C$ (respectively, $C - \mathrm{Corep}$) will indicate the category of finite dimensional left (respectively, left) C-comodules. So that, if A is a finite dimensional algebra, we have $\operatorname{Rep} A = \operatorname{Corep} C^*$.

2. Semisimple Hopf algebras and fusion categories

A Hopf algebra is called *semisimple* (respectively, *cosemisimple*) if it is semisimple as an algebra (respectively, if it is cosemisimple as a coalgebra). A semisimple Hopf algebra is automatically finite dimensional. Let H be a finite-dimensional Hopf algebra over k. By a result of Larson and Radford, it is known that H is semisimple if and only if H is cosemisimple, if and only if $\mathcal{S}^2 = \mathrm{id}$ [LR, LR2].

2.1. $\operatorname{Rep} H$ as a tensor category.

Let H be a finite dimensional Hopf algebra. The category $\operatorname{Rep} H$ of its finite dimensional representations is a finite *tensor category* with tensor product given by the diagonal action of H and unit object k. The antipode implements the H-action on the dual vector space.

Finite tensor categories of the form $\operatorname{Rep} H$ are characterized, using tannakian reconstruction arguments, as those possessing a fiber functor with values in the category of vector spaces over k. The forgetful functor $\operatorname{Rep} H \to \mathrm{Vec}_k$ is a fiber functor and other fiber functors correspond to *twisting* the comultiplication of H in the following sense.

Definition 2.1. A *twist* in H is an invertible element $J \in H \otimes H$ satisfying:

$$(\Delta \otimes \mathrm{id})(J)(J \otimes 1) = (\mathrm{id} \otimes \Delta)(J)(1 \otimes J), \tag{2.1}$$

$$(\epsilon \otimes \mathrm{id})(J) = 1 = (\mathrm{id} \otimes \epsilon)(J). \tag{2.2}$$

Dually, an invertible normalized 2-cocycle on H is a convolution invertible linear map $\sigma : H \otimes H \to k$, such that, for all $g, h, t \in H$,

$$\sigma(h_{(1)}, g_{(1)})\sigma(t, h_{(2)}g_{(2)}) = \sigma(t_{(1)}, h_{(1)})\sigma(t_{(2)}h_{(2)}, g), \tag{2.3}$$

$$\sigma(h, 1) = \epsilon(h) = \sigma(1, h). \tag{2.4}$$

If $J \in H \otimes H$ is a twist, then (H^J, m, Δ^J, S^J) is a Hopf algebra with $H^J = H$ as algebras, $\Delta^J(h) = J^{-1}\Delta(h)J$, and $S^J(h) = v^{-1}S(h)v$, for all $h \in H$, where $v = m(S \otimes \mathrm{id})(J)$.

The Hopf algebras H and H' are called *twist equivalent* if $H' \simeq H^J$. This type of deformation was originally introduced by Drinfeld [Dr2] in the context of quasi-Hopf algebras.

The following theorem is a consequence of a more general result of Schauenburg [S]. An analogous statement for finite dimensional quasi-Hopf algebras has been proved by Etingof and Gelaki.

Theorem 2.2. *The finite dimensional Hopf algebras H and H' are twist equivalent if and only if* $\operatorname{Rep} H \simeq \operatorname{Rep} H'$ *as tensor categories.*

In particular, properties like (quasi)triangularity, semisimplicity or the structure of the Grothendieck ring are preserved under twisting deformations.

Dually, if $\sigma : H \otimes H \to k$ is an invertible normalized 2-cocycle on H, then $(H^\sigma, m^\sigma, \Delta, \mathcal{S}^\sigma)$ is a Hopf algebra, where $H^\sigma = H$ as coalgebras with multiplication and antipode

$$h._\sigma g = \sigma(h_1, g_1)h_2g_2\sigma^{-1}(h_3, g_3), \quad \mathcal{S}^\sigma(h) = u^{-1}(h_{(1)})\mathcal{S}(h_{(2)})u(h_{(3)}),$$

for all $h, g \in H$, where $u(h) = \sigma(\mathcal{S}(h_{(1)}), h_{(2)})$, $h \in H$.

The Hopf algebra H^σ thus defined is called a *cocycle twist* of H [D]. Equivalently, the dual Hopf algebra $(H^\sigma)^*$ is a twisting deformation of H^* via the twist $\sigma \in H^* \otimes H^*$.

Let H, H', be finite dimensional Hopf algebras over k. Then, by the dual version of Theorem 2.2, the tensor categories $H -$ Corep and $H' -$ Corep of finite dimensional corepresentations of H and H', respectively, are equivalent as tensor categories if and only if $H' = H^\sigma$ is a cocycle deformation of H.

The 2-cocycle $\sigma : H \otimes H \to k$ gives rise to a fiber functor U_σ : $H -$ Corep $\to \mathrm{Vec}_k$ whose underlying functor is the forgetful functor $H-$Corep $\to \mathrm{Vec}_k$ with monoidal structure $f : U_\sigma \circ \otimes \to \otimes \circ (U_\sigma \times U_\sigma)$ induced by σ. That is,

$$f(u \otimes v) = \sigma(u_{(-1)} \otimes v_{(-1)})u_{(0)} \otimes v_{(0)}, \tag{2.5}$$

for all $u \in U$, $v \in V$, $U, V \in H -$ Corep, where $u \mapsto u_{(-1)} \otimes u_{(0)}$, denotes the H-coaction on $u \in U$.

Using tannakian reconstruction, one recovers the Hopf algebra H^σ as the endomorphisms of the fiber functor U_σ: $H^\sigma \simeq \mathrm{End}(U_\sigma)$.

This defines a bijective correspondence between equivalence classes of invertible 2-cocycles on H and isomorphism classes of fiber functors on $H -$ Corep.

By results of Ulbrich, generalizing ideas of Grothendieck, isomorphism classes of fiber functors on $H-$Corep correspond bijectively to isomorphism classes of H-Galois extensions of k, also called *H-Galois objects*.

Recall that the extension of k-algebras $B \subseteq A$ is called a right *H-Galois* extension if A is a right H-comodule algebra such that $B = A^{\mathrm{co}\, H}$ and the canonical map

$$\mathrm{can} : A \otimes_B A \to A \otimes H, \quad x \otimes y \mapsto xy_{(0)} \otimes y_{(1)},$$

is bijective. Here, $\rho : A \to A \otimes H$, $\rho(a) = a_{(0)} \otimes a_{(1)}$, denotes the H-coaction on A. Left H-Galois extensions and left H-Galois objects are defined similarly.

The right H-Galois object A corresponds to the fiber functor $U_A := A\Box_H - : H - \text{Corep} \to \text{Vec}_k$, where $\Box_H$ denotes the cotensor product of H-comodules.

Let H, H' be Hopf algebras. An (H', H)-bigalois object is an (H', H)-bicomodule algebra A which is simultaneously a left H'-Galois object and right H-Galois object.

For instance, the Hopf algebra H is itself an (H, H)-bigalois object with respect to the left and right H-coactions given by the comultiplication $\Delta : H \to H \otimes H$.

More generally, let $\sigma : H \otimes H \to k$ be an invertible 2-cocycle. Then the crossed product ${}_\sigma H = k\#_\sigma H$ is a right H-Galois object. When H is finite dimensional, every right H-Galois object is of this form.

For any right H-Galois object A there is an associated Hopf algebra $H' = L(A, H)$, called the *left Galois* Hopf algebra, such that A is in a natural way an (H', H)-bigalois object. By results of Schauenburg, $A\Box_L - : L - \text{Corep} \to H - \text{Corep}$ defines an equivalence of tensor categories, and every equivalence of tensor categories arises in this way, up to isomorphisms, for a unique (H', H)-bigalois object A [S2].

The Hopf algebra $L(A, H)$ is isomorphic to the cocycle deformation H^σ of H, described before. See [S2, Proposition 3.1.6].

Example 2.3. The classification of Hopf Galois objects for finite groups was given by Movshev and Davydov [Da, Mv]. If G is a finite group, then isomorphism classes of k^G-Galois objects are in one-to-one correspondence with conjugacy classes of pairs (S, α), where S is a subgroup of G and $\alpha \in H^2(S, k^\times)$ is a non-degenerate 2-cocycle. The k^G-Galois object corresponding to the pair (S, α) can be constructed as the algebra of S-invariant functions

$$A(G, S, \alpha) = \{f : G \to k_\alpha S : \quad f(sg) = s \triangleright f(g)\},$$

where $k_\alpha S$ is the twisted group algebra, with action $s \triangleright x_t = x_s x_t x_s^{-1}$, $s, t \in S$, and the G-action on A is $(g.f)(h) = f(hg)$.

Note that if $J \in K \otimes K$ is a twist for the Hopf subalgebra $K \subset H$, then $J \in H \otimes H$ is also a twist for H. We shall say that such J is *lifted* from the Hopf subalgebra K [EV, V].

For instance, let G be a finite group, $H = kG$ the group algebra of G and $K = k\Gamma$, where Γ is an abelian subgroup of G. Equivalence classes of twists for K are in one-to-one correspondence with the group $H^2(\Gamma, k^\times)$ [Mv, Proposition 3]. Hence, every 2-cocycle $c \in H^2(\Gamma, k^\times)$ defines a twist $J_c \in kG \otimes kG$. This twist has the form

$$J = \sum_{\alpha,\beta\in\widehat{\Gamma}} c(\alpha,\beta) e_\alpha \otimes e_\beta, \tag{2.6}$$

where $e_\chi = \frac{1}{|G|}\sum_{h\in G}\chi(h^{-1})h$, $\chi \in \widehat{\Gamma}$, is a basis of orthogonal central idempotents of $k\Gamma$.

3. Hopf algebra extensions

In the context of finite dimensional Hopf algebras there is a notion of extension, generalizing the corresponding notion for finite groups. This is a basic tool in the construction of nontrivial (that is, not commutative and not cocommutative) examples, and also to deal with classification problems.

Definition 3.1. An exact sequence of finite dimensional Hopf algebras is a sequence of Hopf algebra maps

$$k \to K \xrightarrow{i} H \xrightarrow{\pi} \overline{H} \to k, \tag{3.1}$$

where K, H and $\overline{H}$ are finite dimensional, such that

(a) i is injective and π is surjective,
(b) $\pi \circ i = \epsilon_K 1$, where ϵ_K denotes the counit of K.
(c) $\ker\pi = HK^+$, or equivalently,
(c') $K = H^{\mathrm{co}\,\pi} = \{h \in H : (\mathrm{id}\otimes\pi)\Delta(h) = h \otimes 1\}$.

A Hopf subalgebra K of H is called *normal* if it is stable under the left adjoint action of H on itself, defined by

$$\mathrm{ad}_l(h)(x) = h_1 x \mathcal{S}(h_2),$$

for all $h, x \in H$. (In general, one should also require K to be stable under the right adjoint action: $\mathrm{ad}_r(h)(x) = \mathcal{S}(h_1)xh_2$, which holds automatically in the finite dimensional context.)

If $K \subseteq H$ is a normal Hopf subalgebra, then $HK^+ = K^+H$ is a Hopf ideal of H and the canonical map $H \to \overline{H} := H/HK^+$ is a Hopf algebra map. Hence there is an exact sequence (3.1) where all maps are canonical. Moreover, in this case H is isomorphic to a bicrossed product $H \simeq K^\tau \#_\sigma \overline{H}$ as a Hopf algebra, with respect to appropriate compatible data: this follows from cleftness of such an exact sequence, which does hold in the finite dimensional case.

A Hopf algebra is called *simple* if it contains no proper normal Hopf subalgebra. For instance, the group algebra of a finite simple group is an example of a (trivial) simple Hopf algebra. The notion of simplicity is self-dual, that is, H is simple if and only if H^* is.

Example 3.2. (Kobayashi-Masuoka.) Let H be a semisimple Hopf algebra and K a Hopf subalgebra. Suppose that the index of K in H, that is, the quotient $\dim H/\dim K$, is the smallest prime number dividing $\dim H$. Then K is normal in H. This generalizes a well-known fact for finite groups.

The concept of solvability of groups translates into the notion of semisolvability of Hopf algebras, due to Montgomery and Witherspoon, in such a way that if H is semisolvable, then it can be obtained from group algebras and their duals via a finite number of extensions.

A related, although not comparable, notion of solvability of a fusion category is introduced and studied by Etingof, Nikshych and Ostrik in [ENO2]. This notion will be discussed later on in Section 4.

Definition 3.3. [MW]. A *lower normal series* for H is a series of Hopf subalgebras $H_n = k \subseteq H_{n-1} \subseteq \cdots \subseteq H_1 \subseteq H_0 = H$, where H_{i+1} is normal in H_i, for all i. The *factors* are the quotients $\overline{H}_i = H_i/H_iH_{i+1}^+$.

An *upper normal series* is inductively defined as follows. Let $H_{(0)} = H$. Let H_i be a normal Hopf subalgebra of $H_{(i-1)}$ and define $H_{(i)} = H_{(i-1)}/H_{(i-1)}H_i^+$. Assume that $H_n = H_{(n-1)}$, for some positive integer n, so that $H_{(n)} = k$. The *factors* are the Hopf subalgebras H_i of the quotients $H_{(i-1)}$.

The Hopf algebra H is called *lower* (respectively, *upper*) *semisolvable* if it possesses a lower (respectively, upper) normal series such that all factors are commutative or cocommutative. If H is both lower and upper semisolvable, then it is called *semisolvable.*

We have that H is upper semisolvable if and only if H^* is lower semisolvable [MW].

Remark 3.4. Note that, equivalently, an upper normal series can be defined as a sequence of quotient Hopf algebra maps $H_{(0)} = H \to H_{(1)} \to \cdots \to H_{(n)} = k$ such that each of the maps $H_{(i-1)} \to H_{(i)}$ is normal. In this case, the factors are $H_i := H_{(i-1)}^{\operatorname{co}\pi_i} = {}^{\operatorname{co}\pi_i}H_{(i-1)}$, where $H_{(i-1)}^{\operatorname{co}\pi_i}$, ${}^{\operatorname{co}\pi_i}H_{(i-1)}$ are the spaces of (right, respectively left) coinvariants of the map π_i. They coincide and form a Hopf subalgebra of $H_{(i-1)}$, by normality of the map π_i.

A result due to Masuoka [M2] says that a semisimple Hopf algebra of dimension p^n, p prime, contains a nontrivial central group-like element g. This implies that group algebra $k\langle g\rangle$ is a central Hopf subalgebra of H. Inductively, this implies that H is semisolvable [MW].

It was shown in [N4] that in dimension < 60 every semisimple Hopf algebra is obtained, up to a twisting deformation, from group algebras and their duals through iterated extensions. That is, they are (either upper or lower) semisolvable, except possibly after a twisting deformation. This result answered a question formulated by S. Montgomery in [Mo3].

Some nontrivial examples of semisimple Hopf algebras which are simple as Hopf algebras arise as twisting deformations of simple groups.

For instance, the twisting $H = (k\mathbb{A}_5)^J$ of the alternating group $\mathbb{A}_5$, where J is the twist lifted from the unique nontrivial 2-cocycle in a Klein subgoup $\Gamma \subseteq \mathbb{A}_5$ is a nontrivial simple Hopf algebra of dimension 60. This example is due to Nikshych [Nk]. The Hopf algebra H is *not* (upper or lower) semisolvable.

In the papers [GN, GN2] certain twisting deformations of a family of supersolvable groups which are simple Hopf algebras were constructed. These groups are direct products of two generalized dihedral subgroups.

Let p, q and r be prime numbers such that q divides $p-1$ and $r-1$. Let $G_1 = \mathbb{Z}_p \rtimes \mathbb{Z}_q$ and $G_2 = \mathbb{Z}_r \rtimes \mathbb{Z}_q$ be the only nonabelian groups of orders pq and rq, respectively. Let $G = G_1 \times G_2$ and let $\mathbb{Z}_q \times \mathbb{Z}_q \simeq \Gamma \subseteq G$ a subgroup of order q^2.

Let also $1 \neq \alpha \in H^2(\widehat{\Gamma}, k^\times)$, $J \in kG \otimes kG$ the twist lifted from Γ corresponding to α.

Theorem 3.5. [GN]. *The Hopf algebra $H = (kG)^J$ is a nontrivial Hopf algebra of dimension prq^2 which is simple as a Hopf algebra.*

Examples of this construction appear in dimension 60 and p^2q^2, for prime numbers p and q such that q divides $p-1$. In particular, the nontrivial simple Hopf algebra of dimension 36 obtained in this way is the smallest example of a semisimple Hopf algebra which is *not* semisolvable.

Semisimple Hopf algebras of dimension 60 which are simple as Hopf algebras were classified in [BN, N8]: such a Hopf algebra is necessarily isomorphic to the twisting of $\mathbb{A}_5$ constructed in [Nk] or to its dual, or to the (self-dual) example arising from Theorem 3.5.

Other simple twisting deformations of finite groups were constructed in [GN, GN2], for instance, from symmetric groups.

The theorem implies that there exists a semisimple Hopf algebra of dimension p^2q^2 which is simple as a Hopf algebra. This proves the following result, answering an open question of S. Montgomery (2000); see [A, Question 4.17].

Corollary 3.6. *The analogue of Burnside's p^aq^b-Theorem for finite groups does not hold for semisimple Hopf algebras.*

We point out that, as shown in [ENO2], an analogue of Burnside's Theorem is valid for fusion categories in the sense of the definition of solvability given in *loc. cit.*

Remark 3.7. Note that if G is a finite group, then the character table of G allows to determine all normal subgroups and their orders, the center $Z(G)$, and in particular, it determines whether or not the group is simple, nilpotent, or solvable. On the other hand, the character table provides the same information about G as does the Grothendieck ring of the tensor category $\operatorname{Rep} G$. Hence these properties are determined by the tensor category $\operatorname{Rep} G$.

As a consequence of [GN], we get that the notion of simplicity or (semi) solvability of a semisimple Hopf algebra is *not* determined by its tensor category of representations.

In contrast with this result, every twisting deformation of a nilpotent group is semisolvable. See a discussion on nilpotent fusion categories in Section 4.

3.1. **Abelian extensions.** An important class of extensions is that of abelian extensions: these are those for which the 'kernel' is a commutative Hopf algebra while the 'cokernel' is a cocommutative Hopf algebra. Reduction to abelian extensions has allowed to obtain classification results in dimensions pq^2, p^2, p^3, etc.

We refer the reader to [M3, M4] for the study of the cohomology theory underlying an abelian exact sequence.

Suppose that $L = F\Gamma$ is an exact factorization of the finite group L, where Γ and F are subgroups of L. Equivalently, F and Γ form a matched pair of finite groups with the actions $\triangleleft\colon \Gamma \times F \to \Gamma$, $\triangleright\colon \Gamma \times F \to F$, defined by $sx = (x \triangleleft s)(x \triangleright s)$, $x \in F$, $s \in \Gamma$.

Let $\sigma : F \times F \to (k^\Gamma)^\times$, $\sigma(x, y) = \sum_s \sigma_s(x, y)e_s$, and $\tau : \Gamma \times \Gamma \to (k^F)^\times$, $\tau(s, t) = \sum_x \tau_x(s, t)e_x$, be normalized 2-cocycles with the respect to the actions afforded, respectively, by $\triangleleft$ and $\triangleright$, subject to appropriate compatibility conditions [M3]. Here, $e_y \in k^F$, $y \in F$, are the canonical idempotents defined by $e_y(x) = \delta_{x,y}$, and similarly for $e_s \in k^\Gamma$.

The bicrossed product $H = k^{\Gamma}\,{}^{\tau}\#_{\sigma} kF$ associated to this data is a Hopf algebra, with multiplication and comultiplication determined, for all $s, t \in \Gamma$, $x, y \in F$, by

$$(e_s\#x)(e_t\#y) = \delta_{s \triangleleft x, t}\, \sigma_s(x, y)\, e_s\#xy, \tag{3.2}$$

$$\Delta(e_s\#x) = \sum_{gh=s} \tau_x(g, h)\, e_g\#(h \triangleright x) \otimes e_h\#x, \tag{3.3}$$

The Hopf algebra H fits into an *abelian* exact sequence $k \to k^{\Gamma} \to H \to kF \to k$. Moreover, every Hopf algebra fitting into such exact sequence can be described in this way. This gives a bijective correspondence between the equivalence classes of these Hopf algebra extensions and a certain abelian group $\operatorname{Opext}(k^{\Gamma}, kF)$ associated to the matched pair (F, Γ).

The class of an element of $\operatorname{Opext}(k^{\Gamma}, kF)$ can be represented by a pair of compatible cocycles (τ, σ). The group $\operatorname{Opext}(k^{\Gamma}, kF)$ can also be described as the first cohomology group of a certain double complex [M3, Proposition 5.2].

The following result is proved in [ENO2]. This leads to the full classification of semisimple Hopf algebras of the prescribed dimensions.

Theorem 3.8. *Let p, q, r be distinct prime numbers. Then every semisimple Hopf algebra of dimension pqr or pq^2 is an abelian extension.*

Indeed, it is shown in [ENO2, Corollary 9.7] that a semisimple Hopf algebra of dimension pq^2 is either an abelian extension or a twist of a group algebra or the dual of such a twist. But in the last cases, H and H^* are of Frobenius type, hence H is an abelian extension, by the results in [N9, N10].

More generally, the conclusion in Theorem 3.8 is true for any *group-theoretical* semisimple Hopf algebra (see Subsection 4.4 below) of square-free dimension, as shown in [ENO2, Lemma 9.5].

A result of G. I. Kac [Ka, M4] says that there is a long exact sequence, called the *Kac exact sequence*,

$$
\begin{aligned}
0 \to H^1(G, k^\times) &\xrightarrow{\text{res}} H^1(F, k^\times) \oplus H^1(\Gamma, k^\times) \to \operatorname{Aut}(k^\Gamma \# kF) \\
\to H^2(G, k^\times) &\xrightarrow{\text{res}} H^2(F, k^\times) \oplus H^2(\Gamma, k^\times) \to \operatorname{Opext}(k^\Gamma, kF) \\
\xrightarrow{\kappa} H^3(G, k^\times) &\xrightarrow{\text{res}} H^3(F, k^\times) \oplus H^3(\Gamma, k^\times) \to \dots
\end{aligned}
$$

This is an important tool in calculations related to the Opext group. See [S3] for a generalization, as well as a conceptual explanation of the Kac exact sequence in terms of related monoidal categories.

A consequence of the results of [S3] is the following description, given in [N2], of the representation category of an abelian extension, in terms of the map $\kappa : \operatorname{Opext}(k^\Gamma, kF) \to H^3(G, k^\times)$ in the Kac exact sequence.

Let $k \to k^\Gamma \to H \to kF \to k$ be an abelian exact sequence associated to an exact factorization $G = F\Gamma$. Let $\omega \in H^3(G, k^\times)$ be the 3-cocycle corresponding to H via the map κ.

Let us denote by $\mathcal{C}(G, \omega, F)$ the category of kF-bimodules in the tensor category $\mathcal{C}(G, \omega)$ of G-graded vector spaces with associativity given by ω (this is a special case of a group-theoretical fusion category, discussed in Subsection 4.4).

Theorem 3.9. *There is an equivalence of fusion categories* $\operatorname{Rep} H \simeq \mathcal{C}(G, \omega, F)$.

Also, there is an equivalence of fusion categories $\operatorname{Rep} D(H) \simeq \operatorname{Rep} D^\omega(G)$. *In other words, the Drinfeld double* $D(H)$ *is twist equivalent to the twisted quantum double* $D^\omega(G)$.

Here, the *twisted quantum double* $D^\omega(G)$, $\omega \in H^3(G, k^\times)$, is the quasi-Hopf algebra introduced by Dijkgraaf, Pasquier and Roche [DPR]. For the case of split extensions, that is when $(\tau, \sigma) = 1$ and hence $\omega = 1$, this result was obtained previously in [BGM].

4. Extensions of fusion categories by finite groups

In this section we review some important recent results from [GNi, DGNO, ENO2]. These concern certain classes of extensions of a fusion category by finite groups. We also discuss some connections with the results of Section 3.

4.1. G**-extensions.** Let G be a finite group. A G-grading of a fusion category $\mathcal{C}$ is a decomposition of $\mathcal{C}$ as a direct sum of full abelian subcategories $\mathcal{C} = \oplus_{g\in G}\mathcal{C}_g$, such that $\mathcal{C}_g^* = \mathcal{C}_{g^{-1}}$ and the tensor product $\otimes : \mathcal{C} \times \mathcal{C} \to \mathcal{C}$ maps $\mathcal{C}_g \times \mathcal{C}_h$ to $\mathcal{C}_{gh}$. The neutral component $\mathcal{C}_e$ is thus a full fusion subcategory of $\mathcal{C}$.

The grading is called *faithful* if $\mathcal{C}_g \neq 0$, for all $g \in G$. In this case, $\mathcal{C}$ is called a *G-extension* of $\mathcal{C}_e$ [ENO2].

Proposition 4.1. *Let $\mathcal{C} = \operatorname{Rep} H$ be the representation category of a semisimple Hopf algebra. Then a faithful G-grading on $\mathcal{C}$ corresponds to a* central *exact sequence of Hopf algebras $k \to k^G \to H \to \overline{H} \to k$, such that $\operatorname{Rep} \overline{H} = \mathcal{C}_e$.*

Dually, a faithful G-grading on $\mathcal{C} = \operatorname{Corep} H$ corresponds to a cocentral *exact sequence of Hopf algebras $k \to K \to H \to kG \to k$, such that $\operatorname{Corep} K = \mathcal{C}_e$.*

Here, the sequence $k \to K \to H \to kG \to k$ is called cocentral if the dual sequence is central (see Subsection 4.3 below).

Proof. See [GNi, Proof of Theorem 3.8] for the statement on $\operatorname{Rep} H$. The dual statement follows from this, since H fits into a cocentral extension $k \to K \to H \to kG \to k$ if and only if the dual Hopf algebra H^* fits into a central extension $k \to k^G \to H^* \to K^* \to k$, if and only if the category $\operatorname{Rep} H^* = \operatorname{Corep} H$ is a G-extension of $\operatorname{Rep} K^* = \operatorname{Corep} K$. $\square$

Let $\mathcal{C}$ be a fusion category and let $\mathcal{C}_{\mathrm{ad}}$ be the adjoint subcategory of $\mathcal{C}$. That is, $\mathcal{C}_{\mathrm{ad}}$ is the full fusion subcategory of $\mathcal{C}$ generated by the subobjects of $X \otimes X^*$, where X runs through the simple objects of $\mathcal{C}$.

It is shown in [GNi] that there is a canonical faithful grading on $\mathcal{C}$: $\mathcal{C} = \oplus_{g\in U(\mathcal{C})}\mathcal{C}_g$, called the *universal grading*, such that $\mathcal{C}_e = \mathcal{C}_{\mathrm{ad}}$. The group $U(\mathcal{C})$ is called the *universal grading group* of $\mathcal{C}$.

In the case where $\mathcal{C} = \operatorname{Rep} H$, for a semisimple Hopf algebra H, $K = k^{U(\mathcal{C})}$ is the maximal central Hopf subalgebra of H and $\mathcal{C}_{\mathrm{ad}} = \operatorname{Rep} H/HK^+$.

4.2. ***G*-equivariantization.** Related to the notion of G-extension, is that of an equivariantization of a fusion category under a group action. This appears in different places, like [AG, FW, Nk2, T2, ENO2]. A characterization of this type of extensions, as well as a generalization using Hopf monads, was given in [BrN].

Let G be a finite group and let $\mathcal{C}$ be a tensor category. The group G is regarded as a monoidal category, denoted by $\underline{G}$, whose objects are the elements of G, arrows are identities and tensor product is the multiplication in G. Similarly, let $\underline{\operatorname{Aut}_{\otimes}}\mathcal{C}$ denote the monoidal category whose objects are tensor autoequivalences of $\mathcal{C}$, morphisms are isomorphisms of tensor functors and tensor product is given by composition of functors.

An *action* of G on $\mathcal{C}$ is a monoidal functor

$$T : \underline{G} \to \underline{\operatorname{Aut}_{\otimes}}\mathcal{C}, \quad \text{with} \quad f^V_{g,h} : T_g(V) \otimes T_h(V) \xrightarrow{\simeq} T_{gh}(V). \tag{4.1}$$

Given an action of G on $\mathcal{C}$, the G-*equivariantization* of $\mathcal{C}$, denoted $\mathcal{C}^G$, is the category of G-equivariant objects and G-equivariant morphisms, defined as follows. A G-equivariant object in $\mathcal{C}$ is a pair $(V, (u^V_g)_{g \in G})$, where V is an object of $\mathcal{C}$ and $u^V_g : T_g(V) \to V$, $g \in G$, are isomorphisms such that, for all $g, h \in G$,

$$u^V_g T_g(u^V_h) = u^V_{gh} f^V_{g,h}. \tag{4.2}$$

A G-equivariant morphism $\phi : (U, u^U_g) \to (V, u^V_g)$ is a morphism $\phi : U \to V$ in $\mathcal{C}$ such that $\phi u^U_g = u^V_g \phi$, for all $g \in G$.

This is a tensor category with tensor product defined as $(U, u^U_g) \otimes (V, u^V_g) = (U \otimes V, (u^U_g \otimes u^V_g) j_g|_{U,V})$, where $j_g|_{U,V} : T_g(U \otimes V) \to T_g(U) \otimes T_g(V)$ are the isomorphisms giving the monoidal structure on T_g.

The category $\mathcal{C}^G$ is Morita equivalent, in the sense of Müger, to a certain G-extension $\mathcal{C} \rtimes G$ of $\mathcal{C}$ with respect to the indecomposable module category $\mathcal{C}$ [Nk2].

In the case of the representation category of a semisimple Hopf algebra K, there is a duality between G-graded fusion categories with trivial component $\operatorname{Rep} K$ and G-equivariantizations of $\operatorname{Rep} K$.

Suppose H fits into a cocentral extension $k \to K \to H \to kG \to k$. Then $\operatorname{Rep} H \simeq (\operatorname{Rep} K)^G$; see Subsection 4.3 below. On the other hand, $\operatorname{Corep} H$ is a G-extension of $\operatorname{Corep} K$, by Proposition 4.1.

4.3. **G-actions on $\operatorname{Rep} H$ and cocentral extensions.** An exact sequence of finite dimensional Hopf algebras

$$k \to H \to \tilde{H} \xrightarrow{\pi} kG \to k$$

is called *cocentral* if $\pi(h_1) \otimes h_2 = \pi(h_2) \otimes h_1$, for all $h \in \tilde{H}$ (equivalently, the dual inclusion $\pi^* : (kG)^* \to \tilde{H}^*$ is central).

In [N7, Proposition 3.5] we showed that every such cocentral exact sequence gave rise to a G-action on $\operatorname{Rep} H$ such that $\operatorname{Rep} \tilde{H} \simeq (\operatorname{Rep} H)^G$ as tensor categories. In this subsection we shall show that the converse is also true, up to twisting deformations. This gives a characterization of cocentral extensions in terms of equivariantizations.

Let G be a finite group and let H be a semisimple Hopf algebra (although the semisimplicity of H is not crucial in our arguments). Consider an action of G on $\operatorname{Rep} H$ by tensor autoequivalences $T : \underline{G} \to \underline{\operatorname{Aut}}_{\otimes} \operatorname{Rep} H$.

For each $g \in G$, consider the tensor functor $(T_g, j_g) : \operatorname{Rep} H \to \operatorname{Rep} H$. By [S] there exist a twist $J(g) \in H \otimes H$ and a Hopf algebra isomorphism $\phi_g : H \to H^{J(g)}$ such that (T_g, j_g) is isomorphic as a tensor functor to $(\phi_g^*, J(g)^{-1})$, where ϕ_g^* is the direct image functor and, by abuse of notation, $J(g)^{-1} : \phi_g^*(U \otimes V) \to \phi_g^*(U) \otimes \phi_g^*(V)$ is the isomorphism given by the action of $J(g)^{-1}$. In particular, ϕ_g are algebra automorphisms of H, for all $g \in G$, and the map $J : G \to H \otimes H$ is invertible. Let us denote $g.a = \phi_g(a)$, $g \in G$, $a \in H$.

In particular, the following hold, for all $g \in G$, $a, b \in H$:

$$g.(ab) = (g.a)(g.b), \quad g.1 = 1, \tag{4.3}$$

$$\Delta(g.a) = J(g)(g.a_1 \otimes g.a_2)J(g)^{-1}. \tag{4.4}$$

For each $g \in G$, let $\lambda(g) : (T_g, j_g) \to (\phi_g^*, J(g)^{-1})$ be an isomorphism of tensor functors. Then we have natural isomorphisms of tensor functors $f'_{g,h} : \phi_g^* \circ \phi_h^* \to \phi_{gh}^*$, for all $g, h \in G$, defined for an H-module X as

$$(f'_{g,h})_X = \lambda(gh)_X (f_{g,h})_X T_g(\lambda(h)_X^{-1})\lambda(g)^{-1}_{\phi_h^*(X)}.$$

The isomorphisms $f'_{g,h}$ determine an invertible map $\sigma : G \times G \to H$ such that $\sigma(g,h)^{-1}|_X = f'_{g,h}|_X$, for all $g, h \in G$, and for all H-module X.

The data ϕ, σ, J satisfy the following conditions:

$$(g.\sigma(h,t))\sigma(g,ht) = \sigma(gh,t)\sigma(g,h), \quad \sigma(1,g) = \sigma(g,1) = 1, \tag{4.5}$$

$$g.(h.a) = \sigma(g,h)(gh.a)\sigma(g,h)^{-1}, \tag{4.6}$$

$$\Delta(\sigma(g,h))J(gh) = J(g)(g.J(h))\,(\sigma(g,h) \otimes \sigma(g,h)), \tag{4.7}$$

for all $g, h, t \in G$, $a, b \in H$. Indeed, (4.5) and (4.6) are equivalent to $\sigma(g,h)$ being isomorphisms of k-linear functors, and (4.7) is equivalent to $\sigma(g,h)$ being morphisms of tensor functors.

Conditions (4.3)–(4.7), together with the twist conditions for $J(g)$, imply that the vector space $H' = H^J \#_\sigma kG$ is a Hopf algebra (the bicrossed product associated to the weak σ-action $\phi : kG \otimes H \to H$, 2-cocycle $\sigma : kG \otimes kG \to H$, and dual cocycle $J : kG \to H \otimes H$) with multiplication and comultiplication given by

$$(a\#g)(b\#h) = a(g.b)\sigma(g,h)\#gh, \tag{4.8}$$

$$\Delta(a\#g) = \Delta(a)J(g)(g \otimes g), \tag{4.9}$$

for all $a, b \in H$, $g, h \in G$. Moreover, the obvious maps give a cocentral exact sequence of Hopf algebras $k \to H \to H' \xrightarrow{\pi} kG \to k$.

Proposition 4.2. *Let G be a finite group and let $\tilde{H}$ be a semisimple Hopf algebra. Then the following are equivalent:*

(i) *There exists a semisimple Hopf algebra H and an action of G on* $\operatorname{Rep} H$ *by tensor autoequivalences, such that* $\operatorname{Rep}\tilde{H} \simeq (\operatorname{Rep} H)^G$.

(ii) $\tilde{H}$ *is twist equivalent to a Hopf algebra H' that fits into a cocentral exact sequence $k \to H \to H' \to kG \to k$.*

Proof. The implication (ii) $\Rightarrow$ (i) is [N7, Proposition 3.5]. We shall show that (i) $\Rightarrow$ (ii). The proof is based on the relation between G-actions and Hopf monads, as studied in [BrN], see Subsection 5.1. Keep the notation above. Let T_G be the Hopf monad on $\operatorname{Rep} H$ corresponding to the given action of G. Consider the bicrossed product Hopf algebra $H' = H^J \#_\sigma kG$ as above. The weak action of G on H given by $g.a = \phi_g(a)$ together with σ and J, induce an action of G on $\operatorname{Rep} H$ by tensor autoequivalences such that the corresponding equivariantization is equivalent to $\operatorname{Rep}\tilde{H}$ [N7, Proposition 3.5]. This action determines in turn a Hopf monad T'_G on $\operatorname{Rep} H$.

By definition of ϕ and σ, there exists an isomorphism of Hopf monads (that is, a morphism of monads which is monoidal) $\lambda = \oplus_g \lambda(g) : T_G \to T'_G$, where $\lambda(g) : (T_g, j_g) \to (\phi_g^*, J(g)^{-1})$ are the given isomorphisms of tensor functors. Indeed, by definition of the isomorphisms $f'_{g,h}$, λ is a morphism of monads, and it is comonoidal because $\lambda(g)$ is an isomorphism of tensor functors, for all $g \in G$.

Hence, $(\operatorname{Rep} H)^{T_G} \simeq (\operatorname{Rep} H)^{T'_G}$ as tensor categories [BV]. The proposition follows from the fact that $\operatorname{Rep}\tilde{H} \simeq (\operatorname{Rep} H)^{T_G}$, while $(\operatorname{Rep} H)^{T'_G} \simeq \operatorname{Rep} H'$, by [N7]. □

4.4. **Weakly group-theoretical fusion categories.** The concepts of G-extension and G-equivariantization discussed previously lead to the notions of nilpotent and solvable fusion categories.

Definition 4.3. [ENO2, GNi]. A fusion category $\mathcal{C}$ is called (cyclically) *nilpotent* if there is a sequence of fusion categories

$$\mathcal{C}_0 = \operatorname{Vec}_k, \mathcal{C}_1, \dots, \mathcal{C}_n = \mathcal{C},$$

such that $\mathcal{C}_i$ is a G_i-extension of $\mathcal{C}_{i-1}$, for some finite (cyclic) groups $G_1, \dots, G_n$.

This definition extends the definition of nilpotency of a finite group, that is, the group G is nilpotent if and only if $\operatorname{Rep} G$ is a nilpotent fusion category. We have in addition:

Proposition 4.4. *Let $\mathcal{C} = \operatorname{Rep} H$, where H is a semisimple Hopf algebra. Then $\mathcal{C}$ is nilpotent if and only if there is a sequence of (normal) quotient Hopf algebras*

$$H_{(n)} = H \to H_{(n-1)} \to \cdots \to H_{(0)} = k,$$

such that $H_i = H^{\operatorname{co} H_{(i-1)}} \simeq k^{G_i}$ is a central Hopf subalgebra of $H_{(i)}$, for all $i = 1, \dots, n$.

Dually, the category $\operatorname{Corep} H$ is nilpotent if and only if there is a sequence of (normal) Hopf subalgebras

$$k = H_0 \subseteq H_1 \cdots \subseteq H_n = H,$$

such that $\overline{H_i} = H_i / H_i H_{i-1}^{+} \simeq kG_i$ is a cocentral Hopf algebra quotient of H_i, for all $i = 1, \dots, n$.

Proof. Suppose $\mathcal{C} = \operatorname{Rep} H$ is nilpotent. Let $\mathcal{C}_0 = \operatorname{Vec}_k, \mathcal{C}_1, \dots, \mathcal{C}_n = \mathcal{C}$, be a sequence of fusion categories such that $\mathcal{C}_i$ is a G_i-extension of $\mathcal{C}_{i-1}$, where $G_1, \dots, G_n$ are finite groups. In particular, $\mathcal{C}_{n-1}$ is isomorphic to a full fusion subcategory of $\mathcal{C}_n = \operatorname{Rep} H$ (the trivial component with respect to the G_n-grading), hence $\mathcal{C}_{n-1} \simeq \operatorname{Rep} H_{(n-1)}$, for some quotient Hopf algebra $H = H_{(n)} \to H_{(n-1)}$. Furthermore, by Proposition 4.1 there is a central exact sequence $k \to k^{G_n} \to H_{(n)} \to H_{(n-1)} \to k$. The claim follows by induction on n. Note that each of the factors of the resulting series is one the Hopf subalgebras $H_{(i)}^{\operatorname{co} H_{(i-1)}} = k^{G_i}$, which is a central Hopf subalgebra of $H_{(i)}$, $i = 1, \dots, n$.

For the statement on the category of corepresentations, apply the above to $\operatorname{Rep} H^* = \operatorname{Corep} H$. □

Corollary 4.5. *Let H be a semisimple Hopf algebra. Then we have:*

(i) $\operatorname{Rep} H$ *is nilpotent if and only if H is upper semisolvable with central factors k^{G_i}.*

(ii) $\operatorname{Corep} H$ *is nilpotent if and only if H is lower semisolvable with cocentral factors kG_i.*

Let us say that a semisimple Hopf algebra H is *nilpotent* if the category $\operatorname{Rep} H$ is nilpotent.

In view of the results of Masuoka [M2], every semisimple Hopf algebra of dimension p^n, p a prime number is nilpotent. More generally, every fusion category of dimension p^n is nilpotent [GNi, Example 4.5].

Remark 4.6. Nilpotency of a semisimple Hopf algebra is not a self-dual notion. Indeed, if G is a finite group, then the Hopf algebra k^G is always nilpotent. However, the group algebra kG is nilpotent if and only if G is a nilpotent group.

Example 4.7. The universal grading group of a group-theoretical category $\mathcal{C} = \mathcal{C}(G, \omega, S, \alpha)$ is computed in [GNa]. It is shown in [GNa, Corollary 4.3] that $\mathcal{C}$ is a nilpotent fusion category, if and only if the normal closure of S in G is nilpotent.

Let G be a finite group and let $A = A(S, \alpha)$ be the k^G-Galois object corresponding to a subgroup $S \subseteq G$ and a nondegenerate 2-cocycle $\alpha \in H^2(S, k^\times)$, as described in Example 2.3. Associated to A there is a semisimple Hopf algebra H which is a cocycle twisting of k^G, and such that $\operatorname{Corep} H \simeq \operatorname{Corep} k^G = \operatorname{Rep} G$. We have in addition $\operatorname{Rep} H \simeq \mathcal{C}(G, S, \alpha)$, by [O1, Theorem 4.2].

Then H is a nilpotent Hopf algebra if and only $\mathcal{C} = \mathcal{C}(G, S, \alpha)$ is a nilpotent fusion category, if and only if the normal closure of S in G is nilpotent.

Note that $H^* = (kG)^J$ is a twisting of kG, so that H^* is nilpotent if and only if G is nilpotent. In particular, in this case, H is nilpotent if H^* is, but it may happen that H is nilpotent and H^* is not.

In the paper [ENO2], the authors define a fusion category to be *simple* if it contains no proper fusion subcategories.

When $\mathcal{C} = \operatorname{Rep} H$ for a semisimple Hopf algebra H, $\mathcal{C}$ is simple if and only if H has no Hopf algebra quotients at all (normal or not). In particular, if G is a finite group $\operatorname{Rep} G$ is simple if and only if G is a simple group, but the category $\mathcal{C}(G)$ of G-graded vector spaces is simple if and only if G is a cyclic group of prime order (that is, G

has no proper subgroups). A different notion of simplicity of a tensor category, discussed later on in Section 5, is given in [BrN].

The following corollary can be seen as a consequence of the results of [N4].

Corollary 4.8. *Let H be a semisimple Hopf algebra of dimension < 60. If* $\operatorname{Rep} H$ *is simple in the sense of* [ENO2], *then $H \simeq k\mathbb{Z}_p$, p prime.*

More generally, by [ENO2, 9.5] the only simple fusion categories with integer Frobenius-Perron dimension < 60 are the categories $\mathcal{C}(G, \omega)$, where G is a cyclic group of prime order and $\omega \in H^3(G, k^\times)$. Indeed, it follows from the results *loc. cit.* that a fusion category of dimension < 60 is always solvable (dimension $p^a q^b$) or group-theoretical (dimension pqr).

On the other hand, a simple fusion category of (Frobenius-Perron) dimension 60 is necessarily isomorphic to the representation category $\operatorname{Rep} \mathbb{A}_5$ [ENO2, Theorem 9.12]. In particular, a semisimple Hopf algebra H of dimension 60 such that $\operatorname{Rep} H$ is simple in the sense of [ENO2] is a twisting of the alternating group $\mathbb{A}_5$.

Definition 4.9. [ENO2]. A fusion category $\mathcal{C}$ is called weakly group-theoretical if there exists an indecomposable algebra A in $\mathcal{C}$ such that ${}_A\mathcal{C}_A$ is a nilpotent fusion category. In the case where ${}_A\mathcal{C}_A$ is a cyclically nilpotent fusion category, then $\mathcal{C}$ is called solvable.

Here, ${}_A\mathcal{C}_A$ is the category of A-bimodules in $\mathcal{C}$ with tensor product $\otimes_A$. This definition can be rephrased saying that $\mathcal{C}$ is *Morita equivalent* to a nilpotent fusion category in the sense of Müger [Mg2].

Solvable fusion categories can be alternatively defined as follows [ENO2, Proposition 4.4]: $\mathcal{C}$ is solvable if and only if there is a sequence of fusion categories

$$\mathcal{C}_0 = \operatorname{Vec}_k, \mathcal{C}_1, \dots, \mathcal{C}_n = \mathcal{C},$$

such that $\mathcal{C}_i$ is obtained from $\mathcal{C}_{i-1}$ either by a G_i-equivariantization or as a G_i-extension, where $G_1, \dots, G_n$ are *cyclic groups of prime order.*

If G is a finite group and $\omega \in H^3(G, k^\times)$, we have that the categories $\mathcal{C}(G, \omega)$ and $\operatorname{Rep} G$ are solvable if and only if G is solvable. In addition, the following facts are pointed out in [ENO2, Remark 4.6]:

• Let G be a non-solvable group. Then the category $\mathcal{C}(G)$ of G-graded vector spaces is nilpotent, but *not* solvable.

• Let G be a solvable group and let $J \in kG \otimes kG$ be a twist such that $H = (kG)^J$ is simple as a Hopf algebra, as in [GN]. Then the category $\operatorname{Rep} H \simeq \operatorname{Rep} G$ is solvable, but H is *not* semisolvable.

• Let H be an abelian extension $k \to k^{\mathbb{A}_4} \to H \to k\mathbb{Z}_5 \to k$, corresponding to the exact factorization $\mathbb{A}_5 = \mathbb{A}_4.\mathbb{Z}_5$. Then H is semisolvable, but $\operatorname{Rep} H$ is *not* solvable. (Indeed, the class of solvable fusion categories is closed under Morita equivalence [ENO2, Proposition 4.5], and we have that the category $\operatorname{Corep} H$ is Morita equivalent to $\mathcal{C}(\mathbb{A}_5, \omega)$, where ω is the Kac 3-cocycle.)

It is shown in [ENO2] that a fusion category $\mathcal{C}$ is weakly group-theoretical in several cases, allowing to give general classification results for fusion categories of certain specific dimensions.

We have the following analogue of Burnside's $p^a q^b$-Theorem.

Theorem 4.10. [ENO2, Theorem 1.6]. *Let p, q prime numbers and a, b nonnegative integers. Then every fusion category of Frobenius-Perron dimension $p^a q^b$ is solvable.*

Let G be a finite group and let ω be a 3-cocycle of G. Let $\mathcal{C} = \mathcal{C}(G, \omega)$ be the category of G-graded vector spaces with associativity isomorphism given by ω. Let also F be a subgroup of G and α a 2-cocycle on F. Suppose that the restriction of ω to F is trivial, so that the twisted group algebra $k_\alpha F$ is an algebra in $\mathcal{C}$. Then the category $\mathcal{C}(G, \omega, F, \alpha)$ of $k_\alpha F$-bimodules in $\mathcal{C}$ is a tensor category.

Definition 4.11. [ENO, 8.8]. $\mathcal{C}(G, \omega, F, \alpha)$ is called a *group-theoretical* category. A (quasi-)Hopf algebra is called group-theoretical if its category of representations is group-theoretical.

Thus every group-theoretical fusion category is weakly group-theoretical, but the converse is not true. Examples of semisimple quasi-Hopf algebras which are not group-theoretical arise from the construction of Tambara and Yamagami [TY] and correspond to nondegenerate bilinear forms of elliptic type on certain finite abelian groups.

An important problem related to the classification of semisimple Hopf algebras over k was the question raised in the paper [ENO], whether all semisimple Hopf algebras over k are group-theoretical.

Every abelian extension is group-theoretical [N2]. Some other positive answers to this question have been obtained for certain cases: semisimple Hopf algebras of dimension p^n and pq^2, integral fusion categories of dimension pqr, where p, q, r are prime numbers, are group-theoretical [DGNO, ENO2].

In the general case, the question was answered negatively by Nikshych [Nk2]. In fact, a family of semisimple Hopf algebras which are not group-theoretical is constructed in [Nk2]: a Hopf algebra H in this family fits into an exact sequence

$$k \to k^{\mathbb{Z}_2} \to H \to (kG)^J \to k, \tag{4.10}$$

where G is a certain finite group and $J \in kG \otimes kG$ is an invertible twist. It turns out that these examples are all semisolvable. The smallest such example has dimension 36, and according to [N7] this is the smallest possible dimension that a non-group-theoretical semisimple Hopf algebra can have.

Note in addition that the results of Nikshych also imply the following:

Corollary 4.12. *The class of group-theoretical Hopf algebras is not closed under Hopf algebra extensions.* □

Remark 4.13. In the paper [N7] we considered semisimple Hopf algebras which fit into an exact sequence

$$k \to (kG)^J \to H \to k\mathbb{Z}_2 \to k, \tag{4.11}$$

where G is a finite group and $J \in kG \otimes kG$ is an invertible twist.

Contrary to the situation for the extensions (4.10), those in (4.11) are always group-theoretical, despite of the symmetry in the form of the extensions. Moreover, a Hopf algebra H as in (4.11) is twist equivalent to an *abelian* extension.

5. Exact sequences of fusion categories

The results in this section are contained in [BrN]. Our exposition here concentrates mainly in the context of fusion (in particular, finite) tensor categories.

Let $\mathcal{C}, \mathcal{C}''$ be tensor categories. Recall that a tensor functor $F : \mathcal{C} \to \mathcal{C}''$ is called *dominant* if any object Y of $\mathcal{C}''$ is a subobject of $F(X)$ for some X in $\mathcal{C}$. F is called *normal* if for any object X of $\mathcal{C}$, there exists a subobject $X_0 \subset X$ such that $F(X_0)$ is the largest trivial subobject of $F(X)$.

Let $\mathfrak{Ker}_F$ denote the full subcategory of $\mathcal{C}$ whose objects are those X such that $F(X)$ is trivial, that is, isomorphic to $\mathbf{1}^n$, $n \geq 1$. When $\mathcal{C}, \mathcal{C}''$ are fusion categories, the functor F is normal if and only if any simple object X of $\mathcal{C}$ such that $\mathrm{Hom}_{\mathcal{C}''}(\mathbf{1}, F(X)) \neq 0$ belongs to $\mathfrak{Ker}_F$.

Definition 5.1. [BrN, Definition 2.7]. Let $\mathcal{C}'$, $\mathcal{C}$, $\mathcal{C}''$ be tensor categories over k. A sequence of tensor functors $\mathcal{C}' \xrightarrow{f} \mathcal{C} \xrightarrow{F} \mathcal{C}''$ is called an exact sequence of tensor categories if the following conditions hold:

(i) F is dominant and normal;
(ii) f is a full embedding;
(iii) The essential image of f is $\mathfrak{Ker}_F$.

This definition leads to the related notions of normal fusion subcategory and simple fusion category. A fusion subcategory $\mathcal{C}' \subset \mathcal{C}$ is *normal* if $\mathcal{C}$ fits into an exact sequence of fusion categories $\mathcal{C}' \to \mathcal{C} \to \mathcal{C}''$. $\mathcal{C}$ is *simple* if it has no non-trivial normal fusion subcategory. This notion of simplicity differs from the one introduced in [ENO2]. For instance, when G is a finite group, then the simplicity of $\mathrm{Rep}\, G$ is equivalent to the simplicity of G and also to the simplicity of the fusion category $\mathcal{C}(G)$ of G-graded vector spaces.

It follows from [BrN, Proposition 3.9] that every exact sequence of finite dimensional (semisimple) Hopf algebras $k \to K \xrightarrow{i} H \xrightarrow{\pi} \overline{H} \to k$ gives rise to an exact sequence of tensor (fusion) categories

$$\mathrm{Rep}\, \overline{H} \xrightarrow{\pi^*} \mathrm{Rep}\, H \xrightarrow{i^*} \mathrm{Rep}\, K. \tag{5.1}$$

Here, the functors π^* and i^* correspond to restriction of representations along the Hopf algebra maps π and i, respectively.

Conversely, suppose that

$$\mathcal{C}' \xrightarrow{f} \mathcal{C} \xrightarrow{F} \mathcal{C}'' \tag{5.2}$$

is an exact sequence of tensor categories, and $\mathcal{C}'' = \operatorname{Rep} K$ for some finite dimensional Hopf algebra K. Denote by $\omega : \operatorname{Rep} K \to \operatorname{Vec}_k$ the forgetful functor, so that $K = \operatorname{End}(\omega)$. Then, letting $H = \operatorname{End}(\omega F)$ and $\overline{H} = \operatorname{End}(\omega F f)$, we have $\mathcal{C} \simeq \operatorname{Rep} H$, $\mathcal{C}' \simeq \operatorname{Rep} \overline{H}$, and there is an exact sequence of Hopf algebras

$$k \to K \xrightarrow{i} H \xrightarrow{\pi} \overline{H} \to k,$$

where $i : K \to H$ and $\pi : H \to \overline{H}$ are Hopf algebra maps induced by f and F, respectively, in such a way that the induced exact sequence (5.1) of tensor categories is isomorphic to (5.2). See [BrN, Remark 2.13].

In view of the above, we can state the following:

Corollary 5.2. *Let H be a finite dimensional Hopf algebra and let $\mathcal{C} = \operatorname{Rep} H$. Then $\mathcal{C}$ fits into a nontrivial exact sequence* (5.2) *if and only if H is twist equivalent to a Hopf algebra extension.* □

There exist examples of semisimple Hopf algebras which are twist equivalent to Hopf algebra extensions (actually to solvable groups), but which are simple as Hopf algebras. See [GN].

Combining Corollary 5.2 with the main result of [N4], we get:

Corollary 5.3. *Let H be a semisimple Hopf algebra of dimension < 60. Then* $\operatorname{Rep} H$ *is not simple in the sense of* [BrN]. □

The result discussed in Remark 3.2 was extended to general fusion categories in the case $p = 2$. The the *Frobenius-Perron index* of a dominant tensor functor $F : \mathcal{C} \to \mathcal{D}$ between fusion categories is the ratio $\operatorname{FPdim}\mathcal{C}/\operatorname{FPdim}\mathcal{C}''$, which is an algebraic integer [ENO].

Proposition 5.4. *Let $F : \mathcal{C} \to \mathcal{D}$ be a dominant tensor functor between fusion categories of Frobenius-Perron index 2. Then F is* *normal, and we have an exact sequence of fusion categories*

$$\operatorname{Rep}\mathbb{Z}_2 \to \mathcal{C} \xrightarrow{F} \mathcal{D}.$$

5.1. **Extensions and Hopf monads.** Exact sequences of tensor categories can be classified in terms of normal faithful Hopf monads.

A Hopf monad on a rigid category $\mathcal{D}$ is an algebra T in the monoidal category $\operatorname{End}(\mathcal{D})$ of endofunctors of $\mathcal{D}$, which is also a comonoidal functor in a compatible way, and possesses left and right antipodes. In this case the category $\mathcal{C}^T$ of T-modules in $\mathcal{C}$ is a tensor category (and, moreover, a fusion category if T is semisimple and so is $\mathcal{C}$). See [BV].

A k-linear right exact Hopf monad T on a tensor category $\mathcal{C}$ is called *normal* if $T(\mathbf{1})$ is a trivial object.

If T is a normal faithful Hopf monad on $\mathcal{C}$, then there is an exact sequence of tensor categories

$$\operatorname{Corep} H \to \mathcal{C}^T \to \mathcal{C},$$

where H is a certain Hopf algebra induced by T.

The following is one of the main results of [BrN].

Theorem 5.5. *Exact sequences of finite tensor categories $\mathcal{C}' \xrightarrow{f} \mathcal{C} \xrightarrow{F} \mathcal{C}''$ are classified by k-linear normal faithful Hopf monads T on category $\mathcal{C}''$, in such a way that $\mathcal{C} \simeq \mathcal{C}^T$.*

Examples of exact sequences of tensor categories arise from equivariantization under group actions. If a finite group G acts on a tensor category $\mathcal{C}$ by tensor autoequivalences, then we have an exact sequence

$$\operatorname{Rep} G \to \mathcal{C}^G \to \mathcal{C}.$$

Indeed, the action of G on $\mathcal{C}$ can be seen as a Hopf monad T on $\mathcal{C}$, defined by $T = \bigoplus_{g\in G} T_g$. In this case we have $\mathcal{C}^G = \mathcal{C}^T$. Hopf monads on $\mathcal{C}$ corresponding to a group action are characterized in [BrN, Theorem 4.24].

Consider for instance an exact sequence of finite groups $1 \to G'' \xrightarrow{\iota} G \xrightarrow{\pi} G' \to 1$, and the associated exact sequence of fusion categories $\operatorname{Rep} G' \to \operatorname{Rep} G \to \operatorname{Rep} G''$. The normal Hopf monad T on $\operatorname{Rep} G''$ associated with this exact sequence is described as follows. The induction functor $\operatorname{Ind}_{G''}^{G} : \operatorname{Rep} G'' \to \operatorname{Rep} G$ is left adjoint to the restriction functor $i^* = \operatorname{Res}_{G''}^{G}$. Let Y be a kG''-module. As a consequence of Mackey's Subgroup Theorem, there is a natural isomorphism

$$\operatorname{Res}_{G''}^{G} \operatorname{Ind}_{G''}^{G}(Y) \simeq \oplus_{\gamma \in G/G''} {}^{\gamma}Y,$$

where ${}^{\gamma}Y$ denotes the kG''-module conjugated to Y under the action of an element $g \in G$ representing the class γ. Then the Hopf monad T is given, as an endofunctor of $\operatorname{Rep} G'$, by:

$$T(Y) = \oplus_{\gamma \in G'} {}^{\gamma}Y.$$

This comes in fact from the action by tensor autoequivalences of G' on $\operatorname{Rep} G''$ by conjugation.

5.2. **Extensions and commutative central algebras.** We next discuss another characterization of exact sequences of fusion categories from [BrN], in terms of commutative central algebras. This relies on results of [BLV].

Let $\mathcal{C}$ be a fusion category. A *central algebra* of $\mathcal{C}$ is a pair (A, σ), where A is an algebra in $\mathcal{C}$ endowed with natural isomorphisms (half-braiding) $\sigma_X : A \otimes X \to X \otimes A$, $X \in \mathcal{C}$, such that the pair (A, σ) is an algebra in the center $\mathcal{Z}(\mathcal{C})$ of $\mathcal{C}$.

A central algebra (A, σ) is called *commutative* if $m\sigma_A = m$, where $m : A \otimes A \to A$ denotes the multiplication in A.

Let (A, σ) be a commutative central algebra of $\mathcal{C}$. Assume A is semisimple. The the category $\mod_{\mathcal{C}} A = \mod_{\mathcal{C}}(A, \sigma)$ of right A-modules in $\mathcal{C}$ is a fusion category with tensor product $\otimes_A$ and unit object $\mathbf{1}$ [BrN, Proposition 5.5].

There is a free module functor $F_A : \mathcal{C} \to \mod_{\mathcal{C}} A$, $X \mapsto X \otimes A$, which is a tensor functor. The central algebra (A, σ) is called the *induced central algebra of $F = F_A$*.

The algebra A is called *self-trivializing* if $F_A(A)$ is a trivial object of $\operatorname{mod}_{\mathcal{C}} A$. The following proposition is contained in [BrN, Proposition 5.7].

Proposition 5.6. *Suppose $F : \mathcal{C} \to \mathcal{D}$ is an exact tensor functor between fusion categories. Let (A, σ) be its induced central algebra. Then F is normal if and only if the algebra A is self-trivializing. In that case, $\mathfrak{Ker}_F = \langle A \rangle \subset \mathcal{C}$ and we have an exact sequence of tensor categories $\langle A \rangle \to \mathcal{C} \to \mathcal{D}$.*

Here, $\langle A \rangle$ denotes the smallest abelian subcategory of $\mathcal{C}$ containing A and stable by direct sums, subobjects and quotients.

The following characterization is contained in [BrN, Corollary 5.8].

Theorem 5.7. *An exact sequence of fusion categories $\mathcal{C}' \to \mathcal{C} \xrightarrow{F} \mathcal{C}''$ is equivalent to the exact sequence $\langle A \rangle \to \mathcal{C} \xrightarrow{F_A} \operatorname{mod}_{\mathcal{C}}(A, \sigma)$, where (A, σ) denotes the induced central algebra of F.*

As a consequence, further necessary and sufficient conditions for an exact sequence to come from an equivariantization were given in terms of the induced central algebra A.

Let H be a semisimple Hopf algebra over k. Consider the forgetful functor $U : \operatorname{Corep} H \to \operatorname{Vec}_k$. The induced central algebra (A, σ) is a commutative algebra in $\mathcal{Z}(\operatorname{Rep} H)$, that is, a commutative algebra in $\operatorname{Rep} D(H)$, where $D(H)$ is the Drinfeld double of H.

As an algebra in $\operatorname{Corep} H$, $A = H$ with right coaction given by the comultiplication. The half-braiding $\sigma_V : A \otimes V \to V \otimes A$ is defined by

$$\sigma_V(h \otimes v) = v_{(0)} \otimes S(v_{(1)})\, h\, v_{(2)}.$$

for any right H-comodule V. We have $\operatorname{mod}_{\operatorname{Corep} H}(A, \sigma) \simeq \operatorname{Vec}_k$ as tensor categories.

Let $f : H \to H'$ be a surjective Hopf algebra map. Let $F = f_* : \operatorname{Corep} H \to \operatorname{Corep} H'$ be the dominant tensor functor defined by f. The functor $R = - \Box_{H'} H$ is a right adjoint of F. The induced central algebra (B, σ') of F is a commutative algebra in $\mathcal{Z}(\operatorname{Corep} H)$.

The algebra B can be described as $B = R(\mathbf{1}) = k\Box_{H'}H = H^{\operatorname{co} H'} \subset H$, where H is seen as a commutative central algebra of $\operatorname{Corep} H$ as before.

We get thus the tensor equivalence (see [Sc1, Theorem II]):

$$\operatorname{mod}_{\operatorname{Corep} H}(B, \sigma') \simeq \operatorname{Corep} H'.$$

Note that the k-linear category $\operatorname{mod}_{\operatorname{Corep} H}(B, \sigma')$ is in this case the category of (B, H)-Hopf modules.

By Proposition 5.6, the functor F is normal (and thus we have an exact sequence of fusion categories) if and only if $F_B(B) = B \otimes B \simeq B^n$, $n = \dim B$, as (B, H)-Hopf modules.

5.3. **Some examples of** $\operatorname{Rep} G$**-extensions.** Examples of $\operatorname{Rep} G$-extensions of fusion categories which do not come from equivariantizations were given in [BrN, Example 4.26]. We discuss some further examples in this subsection.

Lemma 5.8. *Suppose $k \to k^\Gamma \to H \to kG \to k$ is a cocentral exact sequence of Hopf algebras. If Γ is cyclic, then H is cocommutative.*

Proof. The coalgebra structure of H is that of a crossed product $H \simeq k^{\Gamma}{}^{\tau}\#kG$, determined by a certain weak coaction $\rho : kG \to kG \otimes k^\Gamma$ coming, in this case, from an action $\triangleright : \Gamma \times G \to G$, and an invertible 2-cocycle $\tau : \Gamma \times \Gamma \to k^G$, in the form:

$$\Delta(e_h\#g) = \sum_{st=h} \tau(s,t)(g)\, e_s\#(t \triangleright g) \otimes e_t\#g, \tag{5.3}$$

for all $h \in \Gamma$, $g \in G$. Here, $e_h \in k^\Gamma$ is the basic idempotent such that $e_h(s) = \delta_{h,s}$, $s \in \Gamma$. By [N7, Lemma 3.3] ρ, hence also $\triangleright$, are trivial. Also τ is trivial, because Γ is cyclic. Therefore H is cocommutative. $\square$

Let p be a prime number and let K be one of the nontrivial self-dual semisimple Hopf algebras of dimension p^3 as in [M].

The Hopf algebra K fits into an exact sequence of Hopf algebras $k \to k^{\mathbb{Z}_p} \to K \to k(\mathbb{Z}_p \times \mathbb{Z}_p) \to k$, that gives rise to an exact sequence of fusion categories

$$\operatorname{Rep}(\mathbb{Z}_p \times \mathbb{Z}_p) \to \operatorname{Rep} K \to \operatorname{Rep} \mathbb{Z}_p. \tag{5.4}$$

We claim that this exact sequence does not come from an equivariantization. Indeed, if p is odd, then by [M1] K admits no quasitriangular structure, and thus it is not twist equivalent to any cocommutative Hopf algebra. If $p = 2$, then K is isomorphic to the Kac-Paljutkin Hopf algebra H_8 of dimension 8, which is not twist equivalent to a cocommutative Hopf algebra neither, by [TY]. It follows from Proposition 4.2 and Lemma 5.8 that $\operatorname{Rep} K$ is *not* equivalent to an equivariantization $(\operatorname{Rep} \mathbb{Z}_p)^{\mathbb{Z}_p \times \mathbb{Z}_p}$.

5.4. **Exact sequences in the braided context.** A special case of an extension of fusion categories is provided by the modularization process of a premodular (braided) category $\mathcal{C}$ [Br, Mg].

Let $\mathcal{C}$ be a premodular category with braiding $c_{X,Y} : X \otimes Y \to Y \otimes X$. Consider the category $\mathcal{T} \subseteq \mathcal{C}$ of *transparent* objects of $\mathcal{C}$. That is, objects of $\mathcal{T}$ are those X such that $c_{Y,X} c_{X,Y} = \mathrm{id}_{X \otimes Y}$, for all objects Y of $\mathcal{C}$.

Assume $\mathcal{C}$ is modularizable, and let $F : \mathcal{C} \to \widetilde{\mathcal{C}}$ be its modularization (see [Br]). The functor F is dominant and normal, and we have $\mathfrak{Ker}_F = \mathcal{T}$, by [Br, Propositions 2.3 and 3.2]. Therefore, we get an exact sequence of fusion categories:

$$\mathcal{T} \to \mathcal{C} \xrightarrow{F} \widetilde{\mathcal{C}}.$$

Moreover, this exact sequence comes from an equivariantization. In fact, $\mathcal{T}$ is a tannakian category, so that $\mathcal{T} \simeq \operatorname{Rep} G$ as symmetric tensor categories, where G is a finite group that acts on $\tilde{\mathcal{C}}$ and such that $\mathcal{C} = \tilde{\mathcal{C}}^G$.

As an application of the notion of exact sequence of fusion categories, the following classification result was proved in [BrN].

Theorem 5.9. *Let $\mathcal{C}$ be a braided fusion category over k such that* $\operatorname{FPdim} \mathcal{C}$ *is odd and square-free. Then $\mathcal{C}$ is equivalent to* $\operatorname{Rep} \Gamma$ *as a fusion category, for some finite group* Γ.

The proof relies on the concept of modularization and on the fact that a quasitriangular Hopf algebra whose dimension is odd and square-free is in fact a group algebra [N6].

In [N6], a construction of certain canonical quotients of a finite dimensional quasitriangular Hopf algebra, related to modularization, was given. This construction is based on properties of the 'transmutation' studied by S. Majid and on a correspondence between Hopf algebra quotients and coideal subalgebras, due to Takeuchi. The first notion concerns a natural map $\Phi_R : H^* \to H$ (which is not a Hopf algebra map) associated by Drinfeld to an R-matrix $R \in H \otimes H$.

Two extreme classes of quasitriangular Hopf algebras appear in relation with the map Φ_R: the class of *triangular* Hopf algebras, which are those such that Φ_R is trivial, and the class of *factorizable* Hopf algebras, which are those such that Φ_R is an isomorphism.

Finite dimensional triangular Hopf algebras were classified in a series of papers by Andruskiewitsch, Etingof and Gelaki, based on results of Deligne on symmetric categories. On the other hand, several important results have been established for factorizable Hopf algebras. The last ones are related to the so called *modular* categories, which give rise to invariants of 3-manifolds. The Drinfeld double of a finite dimensional Hopf algebra is always a factorizable Hopf algebra.

The main result of [N6] says that, in a certain sense, every quasitriangular Hopf algebra of finite dimension is a kind of extension of the image of the transmutation map Φ_R by a canonical triangular quotient. In particular, it was shown that every quasitriangular Hopf algebra whose dimension is odd and square free is necessarily semisimple and cocommutative. This gave a partial answer to Question 6.5 in [A] (still open in the general case, even in the situation when the dimension is the product of two distinct primes).

6. Some invariants of a fusion category

In this section, we discuss briefly some known results on certain specific invariants of a fusion category $\mathcal{C}$.

Regarding representations of semisimple (quasi-)Hopf algebras, since H and K have isomorphic representation categories if and only if H and K are twist equivalent, invariants of their fusion categories of representations are therefore *gauge invariants*, that is, invariants under twisting deformations.

6.1. **Grothendieck ring.** The tensor product of $\mathcal{C}$ endows its Grothendieck group $G(\mathcal{C})$ with a ring structure; this is a finitely generated ring with an involution induced by duality in $\mathcal{C}$. The product in $G(\mathcal{C})$ gives rise to a notion of dimension, called the Frobenius-Perron dimension, of objects in $\mathcal{C}$. For the category of modules or comodules over a (quasi-)Hopf algebra, the Frobenius-Perron dimension coincides with the dimension of the underlying vector space.

One of the most longstanding open problems about semisimple Hopf algebras is the following:

Conjecture 6.1. *(Kaplansky, 1975.) Let H be a semisimple Hopf algebra and let V be a simple H-module. Then* $\dim V$ *divides* $\dim H$.

A semisimple Hopf algebra satisfying this conjecture is called of *Frobenius type*. This notion can be extended to fusion categories using Frobenius-Perron dimensions; see [ENO2, Definition 1.4].

The conjecture is known to be true in either of the following cases:

• H is quasitriangular. This was proved by Etingof and Gelaki, using the modular structure of the representation category of the Drinfeld double of H.

• H is (upper or lower) semisolvable [MW].

• $\dim V = 2$. This was proved by Nichols and Richmond as a consequence of structural result [NR, Theorem 11] on simple comodules of dimension 2.

More generally, it is shown in [KSZ] that a semisimple Hopf algebra that has a nontrivial self-dual simple module must be even dimensional, generalizing a result of Burnside for finite groups. As a consequence, it turns out that *a semisimple Hopf algebra having a simple module of even dimension must have even dimension.*

• $\dim V = 3$ and H is odd dimensional [Bu, KSZ2]. The proof in [Bu] uses a result on the Grothendieck ring, similar to the one in [NR].

• H (and more generally, a fusion category $\mathcal{C}$) is weakly group-theoretical [ENO2].

For the category $\mathcal{C}(G, \omega, F, \alpha)$, the ring $G(\mathcal{C})$ is known in some cases: when $\omega = 1$, it is determined in [KMY]; in [Nk] it is shown to be a semidirect product for certain particular classes of abelian extensions; Y.-C. Zhu establishes in the paper [Z2] an interesting relation of the Grothendieck ring of $\mathcal{C}$ with the Hecke algebra of the pair G, F (in the case $\omega = 1$, $\alpha = 1$).

The Grothendieck ring of certain bimodule categories over a modular tensor category is described in [FRS].

An important result concerning the structure of the Grothendieck ring of a semisimple Hopf algebra is the following *Class Equation*, due to G. I. Kac and Y.-C. Zhu. Here, $R(H)$ denotes the subalgebra $G(\operatorname{Rep} H) \otimes_{\mathbb{Z}} k \subseteq H^*$. This is a semisimple algebra that coincides with the character algebra of H, that is, the subalgebra generated by the characters of (irreducible) representations: if V is an H-module affording the representation $\rho : H \to \operatorname{End}(V)$, the corresponding *character* is the element $\chi = \chi_V \in H^*$, where $\chi(h) = \operatorname{Tr}(\rho(h))$. Every representation of H is determined by its character, up to isomorphisms.

Theorem 6.2. [Ka2, Z]. *Let $e \in R(H)$ be a primitive idempotent. Then $\dim eH^*$ divides $\dim H$.*

A generalization of this result to spherical fusion categories appears in [ENO, Proposition 5.7].

The *degree* of the character χ is defined as $\deg \chi = \chi(1) = \dim V$, if $\chi = \chi_V$. Let $\operatorname{Irr}(H)$ denote the set of irreducible characters of H. Following [I, Chapter 12], let us consider the set

$$\operatorname{cd}(H) = \{\deg \chi |\, \chi \in \operatorname{Irr}(H)\}.$$

For a finite group, the knowledge of the set $\operatorname{cd}(G) = \operatorname{cd}(kG)$ gives in some cases substantial information about the structure of G. For instance, if $\operatorname{cd}(G) = \{1, m\}$, $m \geq 1$, then either G has an abelian normal subgroup of index m or m is a power of a prime p and G is the direct product of a p-group and an abelian group [I, Theorem 12.5].

For semisimple Hopf algebras, a result in this direction is the following one:

Theorem 6.3. [BN, Corollary 6.6]. *Suppose that* $\mathrm{cd}(H^*) = \{1, 2\}$. *Then H is lower semisolvable.*

The proof of this theorem relies on a refinement of the above mentioned result of Nichols and Richmond given in [BN].

6.2. **Module categories.** A generalized fiber functor is an exact faithful tensor functor $\mathcal{C} \to R - \mathrm{Bimod}$, where R is a separable algebra. This functors play the rôle of 'representations' of $\mathcal{C}$. They correspond to so called *module categories* over $\mathcal{C}$, that is, semisimple k-linear categories $\mathcal{M}$ over k endowed with an exact functor $\mathcal{C} \times \mathcal{M} \to \mathcal{M}$, $(X, M) \mapsto X \otimes M$, satisfying appropriate associativity and unit axioms. See [O1] and references therein. In this way, module categories can be seen as an analogue of the notion of modules over rings.

When $\mathcal{C}$ is the representation category of a semisimple Hopf algebra H, this functors are also in correspondence with H-Galois extensions $R \subseteq A$ [S2, Theorem 2.5.3]. Some references for the theory of Hopf Galois extensions of a Hopf algebra and their most important features are [B, Mo2, S2, SS].

Indecomposable module categories over $\mathrm{Rep}\, G$, where G is a finite group, are classified in [O1, Theorem 3.2]. They are in one-to-one correspondence with conjugacy classes of pairs (Γ, α), where $\Gamma \subseteq G$ is a subgroup and $\alpha \in H^2(\Gamma, k^\times)$. This contains the classification of Galois objects of Movshev and Davydov, described in Example 2.3.

For a group-theoretical category $\mathcal{C}(G, \omega, F, \alpha)$, module categories have been classified by Ostrik: according to [O2, Theorem 3.1], they are in one-to-one correspondence with classes of pairs (Γ, β) where Γ is a subgroup of G such that $\omega|_\Gamma$ is trivial, and $\beta \in H^2(\Gamma, k^\times)$.

The corresponding module category is the category of $(k_\alpha F, k_\beta \Gamma)$-bimodules in $\mathcal{C}(G, \omega)$. This module category is of rank one (that is, it corresponds to a fiber functor on $\mathcal{C}$, and thus to a semisimple Hopf algebra H with $\mathrm{Rep}\, H \simeq \mathcal{C}$), if and only if $F\Gamma = G$ and the cocycle $\alpha\beta^{-1}$ is non-degenerate on $F \cap \Gamma$.

Module categories over an equivariantized category $\mathcal{C}^G$ are classified in [ENO2, Proposition 5.4], generalizing results of [Nk2]. More recently, the classification has also been obtained in [MS] for any G-extension of a fusion category.

6.3. **Frobenius-Schur indicators.** These invariants are defined for a semisimple pivotal tensor category (that is, categories which admit a tensor isomorphism between the identity functor and the functor $V \to V^{**}$) in [FGSV]. They were studied by Mason and Ng [MN] in the context of semisimple quasi-Hopf algebras. They generalize the indicators for finite groups as well as those for semisimple Hopf algebras introduced by Linchenko and Montgomery [LM]. The indicators satisfy a Frobenius-Schur theorem, and they are related with the trace of the antipode in the case of a semisimple Hopf algebra.

Higher Frobenius-Schur indicators were studied by Kashina, Sommerhausser and Zhu [KSZ2] in the context of semisimple Hopf algebras, and later generalized to quasi-Hopf algebras by Ng and Schauenburg [NS1, NS2] (more generally, to pivotal categories). In the paper [NS3] a new invariant is defined, that turns out to generalize the exponent of finite groups and semisimple Hopf algebras (discussed below), that the authors call Frobenius-Schur (FS for short) exponent.

The computation of the second Frobenius-Schur indicators for group-theoretical quasi-Hopf algebras, in group-theoretical terms, was done in [N3]. This result generalized previous computations in [KMM] for abelian extensions.

In the paper [GM] the second indicators are computed for simple modules over the Drinfeld double (which, as a Hopf algebra, is an abelian extension) of a finite real reflection group, showing that they always equal $+1$, generalizing a classical result for the group itself.

Formulas for the higher indicators of simple objects of certain abelian extensions appear in [KSZ2]. For Tambara-Yamagami categories [TY], they were given in the recent paper [Sh], where some arithmetical properties were established.

6.4. **Exponent.** The exponent of a fusion category was introduced by [E], generalizing the notion of exponent of a semisimple Hopf algebra studied in [K, EG3]. By the results of [E] the exponent of $\mathcal{C}$ coincides with the exponent of its Drinfeld's center $\mathcal{Z}(\mathcal{C})$, and moreover, it is an invariant under Morita equivalence.

For a semisimple Hopf algebra H over k, the exponent $\exp H$ of H is, by definition, the smallest positive integer n such that $h_{(1)} \dots h_{(n)} = \epsilon(h)1$, for all $h \in H$.

The following conjecture is still an open problem:

Conjecture 6.4. [K]. *Let H be a semisimple Hopf algebra over k. Then $\exp H$ divides $\dim H$.*

It was shown by Etingof and Gelaki that $\exp H$ divides $(\dim H)^3$ [EG3]. On the other hand, the conjecture was proved to be true for certain Hopf algebra extensions in [K].

In the paper [NS3] the notion of FS-exponent of a semisimple quasi-Hopf algebra was introduced. The FS-exponent does not coincide, in general, with the exponent defined by Etingof. Both exponents differ at most by a 2 factor (namely, FS-$\exp \mathcal{C} = \exp \mathcal{C}$ or $2 \exp \mathcal{C}$), and they do coincide, for instance, when $\dim H$ is odd.

Ng and Schauenburg proved also the following version of the Cauchy's Theorem for semisimple quasi-Hopf algebras; in turn this result was obtained previously in [KSZ2] for semisimple Hopf algebras, allowing to give an affirmative answer to another conjecture formulated by Etingof and Gelaki [EG3]:

Theorem 6.5. [NS3, Theorem 8.4]. *The exponent, the FS-exponent and the dimension of a quasi-Hopf algebra H have the same prime factors.*

A description of the exponent of group-theoretical category was given in [N5] in terms of group cohomology. It turns out that the exponent of $\mathcal{C}(G, \omega, F, \alpha)$ divides the modified exponent of G, defined by $\exp_\omega G := \operatorname{mcm}(e(\omega_g)|g| : g \in G)$; where $e(\omega_g)$ denotes the order of the cohomology class of the restriction of ω to the subgroup generated by $g \in G$. Moreover, $\exp \mathcal{C} = \exp_\omega G$ in certain cases. As a consequence, the exponent of a group-theoretical quasi-Hopf algebra divides the square of its dimension and, in addition, this bound is optimal.

In the paper [LMS] the authors studied the properties of the so called *Hopf order* of an element $h \in H$: this is the least n such that $h_{(1)} \dots h_{(n)} = \epsilon(h)1$. Hopf orders are investigated for some split abelian extensions, including Drinfeld doubles of certain groups (in particular, a semisimple Hopf algebra H may have elements of prime Hopf order p, even when p does not divide $\dim H$). The spaces of elements with trivial n-th Hopf powers are discussed, showing, however, that they do *not* give a twist invariant of H.

7. Some further questions

As already explained, there exist group-theoretical Hopf algebras (specifically, twistings of group algebras), which are not semisolvable. We believe it would be interesting to describe those semisimple Hopf algebras that can be obtained as extensions from (weakly) group-theoretical Hopf algebras. In particular, we do not know the answer to the following question:

Question 7.1. Let $k \to A \to H \to B \to k$ be an extension of Hopf algebras. Suppose A and B are weakly group-theoretical. Is it true that H is weakly group-theoretical?

It is known [N2] that if the extension is abelian then the answer is affirmative. In any case, if the answer were affirmative in general, this would imply that the class of semisimple Hopf algebras which are semisolvable would be contained in the class of weakly group-theoretical Hopf algebras.

In relation with the classification of semisimple Hopf algebras from its character degrees, we do not know the answers to the following questions:

Question 7.2. Let p be a prime number. Let H be a semisimple Hopf algebra such that $\mathrm{cd}(H) = \{1, p\}$. Is it true that H is semisolvable?

It is known [I, Theorem (12.11)] that a finite group G whose irreducible character degrees are either 1 or p must be an extension of an abelian group by $\mathbb{Z}_p$ or else $|G : Z(G)| = p^3$. So these groups are solvable.

Also, the result in [IK, Theorem IX.8 (iii)] implies that answer to Question 7.2 is 'yes' for Kac algebras H when $|G(H^*)| = p$.

Question 7.2 also makes sense in the context of fusion categories, considering solvability instead of semisolvability.

In the context of the exact sequences of tensor categories introduced in [BrN], we think it would be interesting to extend the semisolvability results in low dimension of [N4] to fusion categories. We know that the notion of simplicity of fusion categories considered in [BrN] extends that of finite groups. In particular, the category of representations of the alternating group $\mathbb{A}_5$ is a simple fusion category.

Question 7.3. Does there exist a fusion category of dimension < 60 which is simple in the sense of [BrN]?

As pointed out in Subsection 4.4, the answer is 'no' if one considers instead the notion of simplicity studied in [ENO2]. In view of the main result of [N4] the answer is also 'no' if one considers fusion categories that admit a fiber functor, that is, categories of representations of semisimple Hopf algebras.

In the same spirit, the following questions are natural:

Question 7.4. Does there exist a fusion category of dimension $p^a q^b$, where p and q are distinct prime numbers, which is simple in the sense of [BrN]?

Question 7.5. Does there exist a fusion category of prime power dimension p^n, $n > 1$, which is simple in the sense of [BrN]?

It is clear that fusion categories of prime dimension are simple (according to both definitions of simplicity). On the other hand, fusion categories of dimensions $p^a q^b$ are always solvable. In particular, they are not simple in the sense of [ENO2] if $a + b > 1$.

In relation with the invariants of fusion categories described in Section 6, we believe it would be of interest to compute them for the category $\mathcal{C}^T$, where T is a semisimple faithful (normal) Hopf monad on a fusion category $\mathcal{C}$. In particular, an answer to the following question would give a generalization of the description of module categories for equivariantized categories given in [ENO2]:

Question 7.6. What are module categories for the category $\mathcal{C}^T$?

Concerning extensions of fusion categories, as explained in Section 5, we have the following natural question:

Question 7.7. Let $\mathcal{C}' \to \mathcal{C} \to \mathcal{C}''$ be an exact sequence of fusion categories, and suppose that $\mathcal{C}'$ and $\mathcal{C}''$ are of Frobenius type. Is it true that $\mathcal{C}$ is of Frobenius type?

References

[A] N. Andruskiewitsch, *About finite dimensional Hopf algebras*, Contemp. Math. **294**, 1-57 (2002).

[AG] S. Arkhipov and D. Gaitsgory, *Another realization of the category of modules over the small quantum group*, Adv. Math. **173**, 114–143 (2003).

[BGM] E. Beggs, J. Gould, S. Majid, *Finite group factorizations and braiding*, J. Algebra **181**, 112-151 (1996).

[B] J. Bichon, *Hopf-Galois objects and cogroupoids*, preprint `arXiv:1006.3014`.

[BN] J. Bichon and S. Natale, *Hopf algebra deformations of binary polyhedral groups*, to appear in Transform. Groups, preprint `arXiv:0907.1879v1`.

[Br] A. Bruguières, *Catégories prémodulaires, modularisations et invariants des variétés de dimension* 3, Math. Ann. **316**, 215–236 (2000).

[BLV] A. Bruguières, S. Lack and A. Virelizier, *Hopf monads on monoidal categories*, preprint `arXiv:1003.1920`.

[BrN] A. Bruguières and S. Natale, *Exact sequences of tensor categories*, to appear in Int. Math. Res. Not., preprint `arXiv:1006.0569` (2010).

[BV] A. Bruguières and A. Virelizier, *Hopf monads*, Adv. Math. **215**, 679–733 (2007).

[Bu] S. Burciu, *Representations of degree three for semisimple Hopf algebras*, J. Pure Appl. Algebra **194**, 85-93 (2004).

[Bu2] S. Burciu, *On the classification of semisimple Hopf algebras: structure and applications*, Noncommutative structures in mathematics and physics, Contactforum, Brussel, Koninklijke Vlaamse Academie van Belgie voor Wetenschappen en Kunsten, 47-62 (2010).

[Da] A. A. Davydov, *Galois Algebras and Monoidal Functors between Categories of Representations of Finite Groups*, J. Algebra **244**, 273–301 (2001).

[DPR] R. Dijkgraaf, V. Pasquier, P. Roche, *Quasi-quantum groups related to orbifold models*, Proc. Modern Quantum Field Theory, Tata Institute, Bombay, 1990, 375383.

[D] Y. Doi, *Braided bialgebras and quadratic bialgebras*, Commun. Algebra **21**, 1731–1749 (1993).

[Dr] V. Drinfeld, *Quantum groups*, Proc. Int. Congr. Math., Berkeley, 798–820 (1987).

[Dr2] V. Drinfeld, *Quasi-Hopf algebras*, Leningrad Math. J. **1**, 14191457 (1990).

[DGNO] V. Drinfeld, S. Gelaki, D. Nikshych and V. Ostrik, *Group-theoretical properties of nilpotent modular categories*, preprint `arXiv:math.QA/0704.0195v2`.

[EV] M. Enock and L. Vainerman, *Deformation of a Kac algebra by an abelian subgroup*, Commun. Math. Phys. **178**, 571-596 (1996).

[E] P. Etingof, *On Vafa's Theorem for tensor categories*, Math. Res. Lett. **9**, 651-657 (2002).

[EG2] P. Etingof and S. Gelaki, *The classification of triangular semisimple and cosemisimple Hopf algebras over an algebraically closed field*, Int. Math. Res. Not. **5**, 223-234 (2000).

[EG3] P. Etingof and S. Gelaki, *On the exponent of finite-dimensional Hopf algebras*, Math. Res. Lett. **6**, 131-140 (1999).

[EG4] P. Etingof and S. Gelaki, *The classification of finite dimensional triangular Hopf algebras over an algebraically closed field of characteristic 0*, Mosc. Math. J. **3**, 37–43 (2003).

[ENO] P. Etingof, D. Nikshych and V. Ostrik, *On fusion categories*, Ann. Math. (2) **162**, 581-642 (2005).

[ENO2] P. Etingof, D. Nikshych and V. Ostrik, *Weakly group-theoretical and solvable fusion categories*, Adv. Math. **226**, 176–205 (2011).

[FW] E. Frenkel, and E. Witten, *Geometric endoscopy and mirror symmetry*, Commun. Number Theory Phys. **2**, 113–283 (2008).

[FGSV] J. Fuchs, C. Ganchev, K. Szlachányi and P. Vescernyés, *S_4-symmetries of 6j-symbols and Frobenius-Schur indicators in rigid monoidal C^*-categories*, J. Math. Phys. **40**, 408–426 (1999).

[FRS] J. Fuchs, I. Runkel, C. Schweigert, *The fusion algebra of a bimodule category*, Appl. Categ. Struct. **16**, 123-140 (2008).

[GM] R. Guralnick and S. Montgomery, *Frobenius-Schur indicators for subgroups and the Drinfel'd double of Weyl groups*, Trans. Am. Math. Soc. **361**, 3611–3632 (2009).

[GN] C. Galindo and S. Natale, *Simple Hopf algebras and deformations of finite groups*, Math. Res. Lett. **14** (5-6), 943–954 (2007).

[GN2] C. Galindo and S. Natale, *Normal Hopf subalgebras in cocycle deformations of finite groups*, Manuscrip ta Math. **125**, 501–514 (2008).

[GNa] S. Gelaki and D. Naidu, *Some properties of group-theoretical categories*, J. Algebra **322**, 2631-2641 (2009).

[GNi] S. Gelaki and D. Nikshych, *Nilpotent fusion categories*, Adv. Math. **217**, 1053-1071 (2008).

[I] I. Isaacs, *Character theory of finite groups*, Academic Press, 1976.

[IK] M. Izumi and H. Kosaki, *Kac algebras arising from composition of subfactors: general theory and classification*, Mem. Amer. Math. Soc. **158** (2002).

[Ka] G. Kac, *Extensions of groups to ring groups*, Math. USSR. Sb. **5**, 451–474 (1968).

[Ka2] G. Kac, *Certain arithmetic properties of ring groups*, Functional Anal. Appl. **6**, 158-160 (1972).

[K] Y. Kashina, *On the order of the antipode of Hopf algebras in* ${}^H_HYD$, Commun. Algebra **27**, 1261-1273 (1999).

[KMM] Y. Kashina, G. Mason and S. Montgomery, *Computing the Frobenius-Schur indicator for Abelian extensions of Hopf algebras*, J. Algebra **251**, 888-913 (2002).

[KSZ] Y. Kashina, Y. Sommerhausser and Y. Zhu, *Self-dual modules of semisimple Hopf algebras*, J. Algebra **257**, 88-96 (2002).

[KSZ2] Y. Kashina, Y. Sommerhausser and Y. Zhu, *On higher Frobenius-Schur indicators*, Mem. Am. Math. Soc. **855** (2006).

[Ks] C. Kassel, *Quantum groups*, Graduate Texts in Mathematics **155**, Springer-Verlag, New York, 1995.

[KMY] H. Kosaki, A. Munemasa and S. Yamagami, *On fusion algebras associated to finite group actions*, Pacific J. Math. **177**, 269–290 (1997).

[LMS] R. Landers, S. Montgomery and P. Schauenburg, *Hopf powers and orders for some bismash products*, J. Pure Appl. Algebra **205**, 156–188 (2006).

[LR] R. Larson and D. Radford, *Semisimple cosemisimple Hopf algebras*, Amer. J. Math. **110**, 187–195 (1988).

[LR2] R. Larson and D. Radford, *Finite dimensional cosemisimple Hopf algebras in characteristic* 0 *are semisimple*, J. Algebra **117**, 267–289 (1988).

[LM] V. Linchenko and S. Montgomery, *A Frobenius-Schur theorem for Hopf algebras*, Algebr. Represent. Theory **3** 347-355 (2000).

[Mj] S. Majid, *Foundations of quantum group theory*, Cambridge Univ. Press, 1995.

[MN] G. Mason and S.-H. Ng, *Central invariants and Frobenius-Schur indicators for semisimple quasi-Hopf algebras*, Adv. Math. **190** 161-195 (2005).

[M] A. Masuoka, *Self dual Hopf algebras of dimension p^3 obtained by extension*, J. Algebra **178**, 791–806 (1995).

[M1] A. Masuoka, *Example of almost commutative Hopf algebras which are not coquasitriangular*, Lect. Notes Pure Appl. Math. **237**, 185–191 (2004).

[M2] A. Masuoka, *The p^n Theorem for semisimple Hopf algebras*, Proc. Amer. Math. Soc. **124**, 187-195 (1996).

[M3] A. Masuoka, *Extensions of Hopf algebras*, Trabajos de Matemática **41/99**, Universidad Nacional de Córdoba, 1999.

[M4] A. Masuoka, *Hopf algebra extensions and cohomology*, in: New Directions in Hopf Algebras, MSRI Publ. **43**, 167–209 (2002).

[M5] A. Masuoka, *Classification of semisimple Hopf algebras*, in: Handbook of Algebra **5**, Amsterdam, Elsevier/Noth-Holland (2008).

[MS] E. Meir and E. Semikantov, *Module categories over graded fusion categories*, preprint `arXiv:1010.4333`.

[Mo] S. Montgomery, *Hopf Algebras and their Actions on Rings*, CMBS Reg. Conf. Ser. in Math. **82**, Amer. Math. Soc., 1993.

[Mo2] S. Montgomery, *Hopf Galois theory: A survey*, Geometry & Topology Monographs **16**, 367-400 (2009).

[Mo3] S. Montgomery, *Classifying finite dimensional semisimple Hopf algebras*, Contemp. Math. **229**, 265-279 (1998).

[Mo4] S. Montgomery, *Representation theory of semisimple Hopf algebras*, Math. Phys. Chem. **28**, 189-218 (2001).

[MW] S. Montgomery and S. Whiterspoon, *Irreducible representations of crossed products*, J. Pure Appl. Algebra **129**, 315–326 (1998).

[Mv] M. Movshev, *Twisting in group algebras of finite groups*, Func. Anal. Appl. **27**, 240–244 (1994).

[Mg] M. Müger, *Galois theory for braided tensor categories and the modular closure*, Adv. Math. **150**, 151–201 (2000).

[Mg2] M. Müger, *From subfactors to categories and topology, I. Frobenius algebras in and Morita equivalence of tensor categories*, J. Pure Appl. Algebra **180**, 81157 (2003).

[Mg3] M. Müger, *Tensor categories: A selective guided tour*, to appear in Rev. Unión Mat. Argentina, preprint `arXiv:0804.3587`.

[N2] S. Natale, *On group-theoretical Hopf algebras and exact factorizations of finite groups*, J. Algebra **270**, 199-211 (2003).

[N3] S. Natale, *Frobenius-Schur indicators for a class of fusion categories*, Pacific J. Math. **221**, 353–378 (2005).

[N4] S. Natale, *Semisolvability of semisimple Hopf algebras of low dimension*, Mem. Amer. Math. Soc. **186**, 123 pp. (2007).

[N5] S. Natale, *On the exponent of tensor categories coming from finite groups*, Israel J. Math. **162**, 253–273 (2007).

[N6] S. Natale, *R-matrices and Hopf algebra quotients*, Int. Math. Res. Not. **2006**, 1–18 (2006).

[N7] S. Natale, *Hopf algebra extensions of group algebras and Tambara-Yamagami categories*, Algebr. and Repr. Theory **13**, 673-691 (2010).

[N8] S. Natale, *Semisimple Hopf algebras of dimension* 60, J. Algebra **324**, 3017-3034 (2010).

[N9] S. Natale, *On semisimple Hopf algebras of dimension* pq^2, J. Algebra **221**, 242–278 (1999).

[N10] S. Natale, *Semisimple Hopf algebras of dimension* pq^2, II, Alg. and Repr. Theory **5**, 277–291 (2001).

[NR] W. Nichols and M. Richmond, *The Grothendieck group of a Hopf algebra*, J. Pure Appl. Algebra **106**, 297–306 (1996).

[Nk] D. Nikshych, K_0*-rings and twisting of finite-dimensional semisimple Hopf algebras*, Commun. Algebra **26**, 321–342 (1998). (Corrigendum: Commun. Algebra **26**, 2019 (1998).)

[Nk2] D. Nikshych, *Non group-theoretical semisimple Hopf algebras from group actions on fusion categories*, Selecta Math. (N.S.) **14**, 145–161 (2008).

[Ng1] S.-H. Ng, *Non-semisimple Hopf algebras of dimension* p^2, J. Algebra **255**, 182-197 (2002).

[Ng2] S.-H. Ng, *Hopf algebras of dimension* pq, J. Algebra **276**, 399-406 (2004).

[Ng3] S.-H. Ng, *Hopf algebras of dimension* $2p$, Proc. Am. Math. Soc. **133**, 2237-2242 (2005).

[NS1] S. Ng and P. Schauenburg, *Higher Frobenius-Schur indicators for pivotal categories*, Contemporary Mathematics **441**, 63–90 (2007).

[NS2] S.-H. Ng and P. Schauenburg, *Central invariants and Higher Frobenius-Schur indicators for semisimple quasi-Hopf algebras*, Trans. Amer. Math. Soc. **360**, 1839-1860 (2008).

[NS3] S.-H. Ng and P. Schauenburg, *Frobenius-Schur indicators and exponents of spherical categories*, Adv. Math. **211**, 34-71 (2007).

[O1] V. Ostrik, *Module categories, weak Hopf algebras and modular invariants*, Transform. Groups **8**, 177-206 (2003).

[O2] V. Ostrik, *Module categories over the Drinfeld double of a finite group*, Int. Math. Res. Not. **2003**, 1507–1520 (2003).

[S] P. Schauenburg, *Hopf bigalois extensions*, Commun. Algebra **24**, 3797-3825 (1996).

[S2] P. Schauenburg, *Hopf-Galois and Bi-Galois Extensions*, Fields Institute Communications **43**, 469–515 (2004).

[S3] P. Schauenburg, *Hopf bimodules, coquasibialgebras, and an exact sequence of Kac*, Adv. Math. **165**, 194–263 (2002).

[SS] P. Schauenburg and H.-J. Schneider, *Galois type extensions and Hopf algebras*, preprint (2004).

[Sc1] H.-J. Schneider, *Principal homogeneous spaces for arbitrary Hopf algebras*, Israel J. Math. **72**, 167–195 (1990).

[Sc2] H.-J. Schneider, *Some properties of factorizable Hopf algebras*, Proc. Amer. Math. Soc. **29**, 1891–1898 (2001).

[Sc3] H.-J. Schneider, *Lectures on Hopf algebras*, Trab. Matemática **31/95**, FaMAF, 1995.

[Sh] K. Shimizu, *Frobenius-Schur indicators in Tambara-Yamagami categories*, preprint `arXiv:1005.4500`.

[St] D. Stefan, *Hopf algebras of low dimension*, J. Algebra **111**, 343-361 (1999).

[T2] D. Tambara, *Invariants and semi-direct products for finite group actions on tensor categories*, J. Math. Soc. Japan **53**, 429–456 (2001).

[TY] D. Tambara and S. Yamagami, *Tensor categories with fusion rules of self-duality for finite abelian groups*, J. Algebra **209**, 692–707 (1998).

[V] L. Vainerman, *2-cocycles and twisting of Kac algebras*, Commun. Math. Phys. **191**, 697–721 (1998).

[Z] Y.-C. Zhu, *Hopf algebras of prime dimension*, Int. Math. Res. Not. **1**, 53–59 (1994).

[Z2] Y.-C. Zhu, *Hecke algebras and representation ring of Hopf algebras*, Adv. Math. **20**, 219-227 (2001).

Facultad de Matemática, Astronomía y Física, Universidad Nacional de Córdoba. CIEM – CONICET. (5000) Ciudad Universitaria, Córdoba, Argentina

E-mail address: `natale@famaf.unc.edu.ar,`

URL: `http://www.famaf.unc.edu.ar/∽natale`

AN EXAMPLE CONCERNING THE THEORY OF LEVELS FOR CODIMENSION-ONE FOLIATIONS

ANDRÉS NAVAS

An important aspect of foliations concerns the existence of local minimal sets. Recall that a foliated manifold has the LMS property if, for every open, saturated set W and every leaf $L \subset W$, the relative closure $\bar{L} \cap W$ contains a minimal set of $F|_W$. A fundamental result (due to Cantwell-Conlon [2] and Duminy-Hector [5]) establishes the LMS property for codimension-one foliations that are transversely of class $C^{1+\text{Lipschitz}}$. This is the basic tool of the so-called *Theory of Levels.*

A classical example due to Hector (which corresponds to the suspension of a group action on the interval) shows that the LMS property is no longer true for codimension-one foliations which transversely are only continuous (see [1, Example 8.1.13]). Despite of this, in recent years, the possibility of extending some of the results of the Theory of Levels to smoothness smaller than $C^{1+\text{Lipschitz}}$ has been naturally addressed [3, 4]. In this Note we will show that, however, analogues of Hector's example appear in class C^1 (and actually in class $C^{1+\alpha}$ for small values of α).

1. A General Construction

Let $(a_n)_{n\in\mathbb{Z}}$ be a sequence such that $a_{n+1} < a_n$ for all $n \in \mathbb{Z}$, $a_n \to 0$ as $n \to \infty$, and $a_n \to 1$ as $n \to -\infty$. Let (n_k) be a strictly increasing sequence of positive integers, and let $f\colon [0,1] \to [0,1]$ be a homeomorphism such that $f(a_{n+1}) = a_n$ for all $n \in \mathbb{Z}$. For each $k > 0$, we let u_k, v_k, b_k, c_k be such that $a_{n_k+1} < b_k < u_k < v_k < c_k < a_{n_k}$. For each $i \in \{0, \ldots, n_{k+1} - n_k\}$, we set $u_k^i := f^i(u_k)$ and $v_k^i := f^i(v_k)$.

Partially funded by the Math-AMSUD Project DySET..

Notice that

$$f^i([u^0_{k+1}, v^0_{k+1}]) = [u^i_{k+1}, v^i_{k+1}] \subset f^i([a_{1+n_{k+1}}, a_{n_{k+1}}]) = [a_{n_{k+1}-i+1}, a_{n_{k+1}-i}].$$

Now, we let $g\colon [0,1] \to [0,1]$ be a homeomorphism such that:

– $g = Id$ on $[a_{n+1}, a_n]$ for each $n < 0$, as well as each $n > 0$ such that $n \neq n_k$ for every k;
– $g = Id$ on $[a_{1+n_k}, b_k] \cup [c_k, a_{n_k}]$, $g(u^0_k) = v^0_k$, and g has no fixed point on $]b_k, c_k[$.

Main assumption: In order that f, g generate a group of homeomorphisms of $[0,1]$ whose associated suspension does not have the LMS property, we assume that (see Figure 1)

$$u^{n_{k+1}-n_k}_{k+1} = b_k \qquad \text{and} \qquad v^{n_{k+1}-n_k}_{k+1} = c_k.$$

With these general notations, Hector's example corresponds to the choice $n_k = k$. We will show that, by taking $n_k = 2^k$, one may perform this construction in such a way the resulting maps f and g are diffeomorphisms of class C^1 (actually, of class $C^{1+\alpha}$ for any $\alpha < (\sqrt{5}-1)/2$). It is quite possible that slightly improving our method, one can smooth the action up to the class $C^{2-\delta}$ for any $\delta > 0$. (Compare [7], where for a similar construction, T. Tsuboi deals with the $C^{3/2-\delta}$ case before the $C^{2-\delta}$ case due to technical difficulties.)

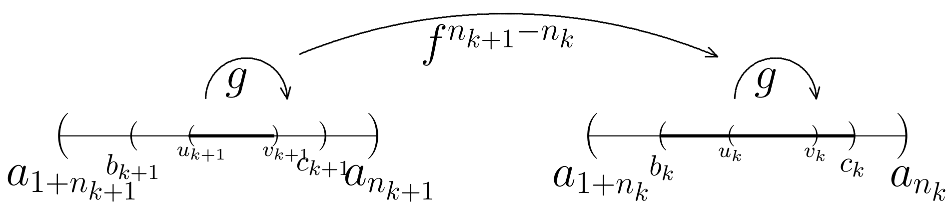


Figure 1

2. The length of the intervals and some estimates

We let $|[u_{k+1}^i, v_{k+1}^i]| := \lambda_k^i |[u_{k+1}, v_{k+1}]|$, where the constant $\lambda_k > 1$ satisfies the compatibility relation

$$\lambda_k^{2^k} = \frac{|[u_{k+1}^{2^k}, v_{k+1}^{2^k}]|}{|[u_{k+1}, v_{k+1}]|} = \frac{|[b_k, c_k]|}{|[u_{k+1}, v_{k+1}]|}. \tag{1}$$

Let $\varepsilon > 0$ be very small (to be fixed in a while). We set:

– $|[a_{n+1}, a_n]| := \frac{c_\varepsilon}{(1+|n|)^{1+\varepsilon}}$, where c_ε is chosen so that $\sum_{n \in \mathbb{Z}} |[a_{n+1}, a_n]| = 1$;

– $|[b_k, c_k]| := \frac{1}{2}|[a_{2^k+1}, a_{2^k}]| = \frac{c_\varepsilon}{2(1+2^k)^{1+\varepsilon}}$, where $k > 0$;

– $|[u_k, v_k]| := |[b_k, c_k]|^{1+\theta}$.

We assume that the center of $[a_{2^k+1}, a_{2^k}]$ coincides with the center of $[b_k, c_k]$ and with that of $[u_k, v_k]$. Furthermore, we assume that for each $i \in \{0, \ldots, 2^k\}$, the centers of $[u_{k+1}^i, v_{k+1}^i]$ and $[a_{2^{k+1}-i+1}, a_{2^{k+1}-i}]$ coincide.

For the estimates concerning regularity, we will strongly use the following lemma from [6].

Technical Lemma. *Let $\omega\colon [0, \eta] \to [0, \omega(\eta)]$ be a function (modulus of continuity) such that $s \mapsto s/\omega(s)$ is non-increasing. If I, J are closed non-degenerate intervals such that $1/2 \le |I|/|J| \le 2$ and*

$$\left| \frac{|J|}{|I|} - 1 \right| \frac{1}{\omega(|I|)} \le M,$$

then there exists a $\mathrm{C}^{1+\omega}$ diffeomorphism $f : I \to J$ that is tangent to the identity at the endpoints and whose derivative has ω-norm bounded from above by $6\pi M$.

Actually, for $I := [a,b]$ and $J := [a',b']$, one may take $f = \varphi_{a',b'}^{-1} \circ \varphi_{a,b}$, where $\varphi_{a,b}$ is defined by (a similar definition stands for $\varphi_{a',b'}$)

$$\varphi_{a,b}(x) = -\frac{1}{(b-a)}\mathrm{ctg}\left(\pi\Big(\frac{x-a}{b-a}\Big)\right).$$

The condition on the derivative at the endpoints allows us to fit together the maps in order to create a diffeomorphism of a larger interval. Actually, if all of the involved sub-intervals of type I, J satisfy the hypothesis of the lemma above for the same constant M, then the ω-norm of the derivative of the induced diffeomorphism is bounded from above by $12\pi M$.

In what follows, we will deal with the modulus of continuity $\omega(s) = s^\alpha$ for the derivative, where $\alpha > 0$. A constant depending on the three parameters $\alpha, \theta, \varepsilon$, and whose value is irrelevant for our purposes, will be generically denoted by M.

Estimates for f: The diffeomorphism f is constructed by fitting together the maps provided by the Technical Lemma sending (see Figure 2):

(i) $[u_{k+1}^i, v_{k+1}^i]$ into $[u_{k+1}^{i+1}, v_{k+1}^{i+1}]$,

(ii) $[a_{2^{k+1}-i}, u_{k+1}^i]$ into $[a_{2^{k+1}-i-1}, u_{k+1}^{i+1}]$,

(iii) $[v_{k+1}^i, a_{2^{k+1}-i-1}]$ into $[v_{k+1}^{i+1}, a_{2^{k+1}-i-2}]$.

For (i), we have

$$\begin{aligned}\left|\frac{|[u_{k+1}^{i+1}, v_{k+1}^{i+1}]|}{|[u_{k+1}^i, v_{k+1}^i]|} - 1\right| \frac{1}{|[u_{k+1}^i, v_{k+1}^i]|^\alpha} &= |\lambda_k - 1| \frac{1}{(\lambda_k^i |[u_{k+1}^0, v_{k+1}^0]|)^\alpha} \\ &\leq |\lambda_k - 1| \frac{1}{|[b_{k+1}, c_{k+1}]|^{(1+\theta)\alpha}}.\end{aligned}$$

Now from (1) one obtains

$$\lambda_k^{2^k} = \frac{\frac{c_\varepsilon}{2(1+2^k)^{1+\varepsilon}}}{\left(\frac{c_\varepsilon}{2(1+2^{k+1})^{1+\varepsilon}}\right)^{1+\theta}} \leq M\Big(\frac{(1+2^{k+1})^{1+\theta}}{1+2^k}\Big)^{1+\varepsilon} \leq M 2^{k\theta(1+\varepsilon)}.$$

From the inequality $|2^\alpha - 1| \leq \alpha$ (which holds for α positive and small) one concludes that

$$|\lambda_k - 1| \leq M\frac{k}{2^k}.$$

On the other hand,

$$\frac{1}{|[b_{k+1}, c_{k+1}]|} \leq M(1 + 2^{k+1})^{1+\varepsilon} \leq M2^{k(1+\varepsilon)}.$$

Therefore,

$$\left|\frac{|[u_{k+1}^{i+1}, v_{k+1}^{i+1}]|}{|[u_{k+1}^{i}, v_{k+1}^{i}]|} - 1\right| \frac{1}{|[u_{k+1}^{i}, v_{k+1}^{i}]|^\alpha} \leq M\frac{k}{2^k} 2^{k(1+\varepsilon)(1+\theta)\alpha}. \tag{2}$$

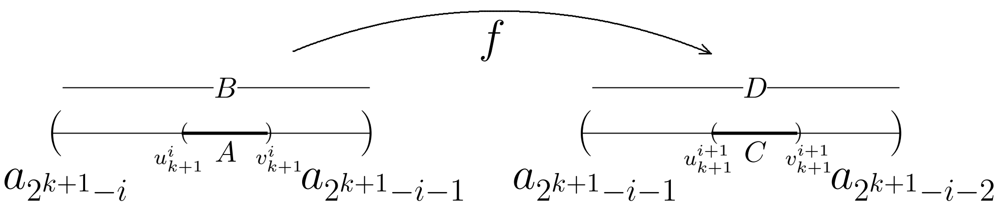


Figure 2

Now, for (ii), set $A := |[u_{k+1}^{i}, v_{k+1}^{i}]|$, $B := |[a_{2^{k+1}-i}, a_{2^{k+1}-i-1}]|$, $C := |[u_{k+1}^{i+1}, v_{k+1}^{i+1}]|$, and $D := |[a_{2^{k+1}-i-1}, a_{2^{k+1}-i-2}]|$. Then

$$\left|\frac{|[a_{2^{k+1}-i-1}, u_{k+1}^{i+1}]|}{|[a_{2^{k+1}-i}, u_{k+1}^{i}]|} - 1\right| \frac{1}{|[a_{2^{k+1}-i}, u_{k+1}^{i}]|^\alpha} = \left|\frac{D-C}{B-A} - 1\right| \frac{2^\alpha}{(B-A)^\alpha}.$$

Moreover, since $A \leq B/2$ and $C = \lambda_k A$,

$$\begin{aligned}
\left|\frac{D-C}{B-A} - 1\right| &\leq \left|\frac{D-B}{B-A}\right| + \left|\frac{C-A}{B-A}\right| \leq 2\left|\frac{D-B}{B}\right| + |\lambda_k - 1| \\
&= \frac{M}{B}\left[\frac{1}{(2^{k+1}-i-2)^{1+\varepsilon}} - \frac{1}{(2^{k+1}-i-1)^{1+\varepsilon}}\right] + M\frac{k}{2^k} \\
&\leq MB\left[(2^{k+1}-i-1)^{1+\varepsilon} - (2^{k+1}-i-2)^{1+\varepsilon}\right] + M\frac{k}{2^k} \\
&\leq \frac{M}{2^{k(1+\varepsilon)}}2^{k\varepsilon} + M\frac{k}{2^k} \\
&\leq M\frac{k}{2^k}.
\end{aligned}$$

Therefore,

$$\left|\frac{D-C}{B-A} - 1\right| \frac{2^\alpha}{(B-A)^\alpha} \leq M\frac{k}{2^k}2^{k(1+\varepsilon)\alpha},$$

hence

$$\left|\frac{|[a_{2^{k+1}-i-1}, u_{k+1}^{i+1}]|}{|[a_{2^{k+1}-i}, u_{k+1}^{i}]|} - 1\right| \frac{1}{|[a_{2^{k+1}-i}, u_{k+1}^{i}]|^\alpha} \leq M\frac{k}{2^{k(1-(1+\varepsilon)\alpha)}}. \tag{3}$$

Finally, notice that by construction, the estimates for (iii) are the same as those for (ii).

Estimates for g: The diffeomorphism g is obtained by fitting together the maps provided by the Technical Lemma sending:

(i) $[b_k, u_k^0]$ into $[b_k, v_k^0]$,

(ii) $[u_k^0, c_k]$ into $[v_k^0, c_k]$,

(iii) $[a_{2^k+1}, b_k]$ and $[c_k, a_{2^k}]$ into themselves as the identity.

For (i), notice that

$$\begin{aligned}
\left|\frac{|[b_k, v_k^0]|}{[b_k, u_k^0]} - 1\right| \frac{1}{|[b_k, u_k^0]|^\alpha} = \frac{|[u_k^0, v_k^0]|}{|[b_k, u_k^0]|^{1+\alpha}} &\leq \frac{2^{1+\alpha}|[u_k^0, v_k^0]|}{(|[b_k, c_k]| - |[u_k^0, v_k^0]|)^{1+\alpha}} \\
&= \frac{2^{1+\alpha}|[b_k, c_k]|^{1+\theta}}{(|[b_k, c_k]| - |[b_k, c_k]|^{1+\theta})^{1+\alpha}},
\end{aligned}$$

thus

$$\left| \frac{|[b_k, v_k^0]|}{[b_k, u_k^0]} - 1 \right| \frac{1}{|[b_k, u_k^0]|^\alpha} \leq M|[b_k, c_k]|^{\theta - \alpha}. \tag{4}$$

The estimates for (ii) are similar to those for (i) and we leave them to the reader.

The choice of the parameters: According to our Technical Lemma, and due to (2), (3), and (4), sufficient conditions for the $C^{1+\alpha}$ smoothness of f, g are:

– $(1+\varepsilon)(1+\theta)\alpha < 1$,

– $\frac{1}{1+\varepsilon} > \alpha$

– $\theta > \alpha$.

Now, for $0 < \alpha < (\sqrt{5}-1)/2$, one easily checks that these conditions are satisfied for $\theta := \alpha + \varepsilon$, where $\varepsilon > 0$ is small enough so that $(1+\varepsilon)(1+\alpha+\varepsilon)\alpha < 1$.

Acknowledgments. I would like to thank J. Cantwell and L. Conlon for motivating me to work on and write out the example of this Note.

References

[1] Candel, A. & Conlon, L. *Foliations, I.* American Mathematical Society, Providence, Rhode Island (1999).

[2] Cantwell, J. & Conlon, L. Poincaré-Bendixson theory for leaves of codimension-one. *Trans. Amer. Math. Soc.* **265** (1981), 181-209.

[3] Cantwell, J. & Conlon, L. A new look at the Theory of Levels. In: A celebration of the mathematical legacy of Raoul Bott. CRM Proc. Lecture Notes **50**, Amer. Math. Soc., Providence, RI (2010), 125-137.

[4] Deroin, B., Kleptsyn, V. & Navas, A. Sur la dynamique unidimensionnelle en régularité intermédiaire. *Acta Math.* **199** (2007), 199-262.

[5] Hector, G. Architecture des feuilletages de classe C^2. *Astérisque* **107-108** (1983), 243-258.

[6] NAVAS, A. Growth of groups and diffeomorphisms of the interval. *Geom. and Functional Analysis* **18** (2008), 988-1028.

[7] TSUBOI, T. Homological and dynamical study on certain groups of Lipschitz homeomorphisms of the circle. *J. Math. Soc. Japan* **47** (1995), 1-30.

UNIV. DE SANTIAGO DE SANTIAGO, ALAMEDA 3363, SANTIAGO, CHILE
E-mail address: andres.navas@usach.cl

ACCESSIBILITY AND ABUNDANCE OF ERGODICITY IN DIMENSION THREE: A SURVEY.

FEDERICO RODRIGUEZ HERTZ, JANA RODRIGUEZ HERTZ, AND RAÚL URES

Abstract. In [18] the authors proved the Pugh-Shub conjecture for partially hyperbolic diffeomorphisms with 1-dimensional center, i.e. stably ergodic diffeomorphisms are dense among the partially hyperbolic ones and, in subsequent results [20, 21], they obtained a more accurate description of this abundance of ergodicity in dimension three. This work is a survey type paper of this subject.

1. Introduction

The purpose of this survey is to present the state of the art in the study of the ergodicity of conservative partially hyperbolic diffeomorphisms on three dimensional manifolds. In fact, we shall mainly describe the results contained in [20, 21]. The study of partial hyperbolicity has been one of the most active topics on dynamics over the last years and we do not pretend to describe all the related results, even for 3-manifolds. Some of the important themes excluded in this survey are entropy maximizing measures, absolute continuity of center foliations, co-cycles over partially hyperbolic systems, SRB-measures, dynamical coherence, classification, etc.

A diffeomorphism $f : M \to M$ of a closed smooth manifold M is partially hyperbolic if TM splits into three invariant bundles such that one of them is contracting, the other is expanding, and the

Date: August 24, 2011.

2000 *Mathematics Subject Classification.* Primary: 37D30, Secondary: 37A25.

Key words and phrases. partial hyperbolicity; accessibility property; ergodicity; laminations.

third, called the center bundle, has an intermediate behavior, that is, not as contracting as the first, nor as expanding as the second (see Subsection 2.3 for a precise definition). The first and second bundles are called strong bundles.

A central point in dynamics is to find conditions that guarantee ergodicity. In 1994, the pioneer work of Grayson, Pugh and Shub [9] suggested that partial hyperbolicity could be "essentially" a sufficient condition for ergodicity. Indeed, soon afterwards, Pugh and Shub conjectured that stable ergodicity (open sets of ergodic diffeomorphisms) is dense among partially hyperbolic systems. They proposed as an important tool the accessibility property (see also the previous work by Brin and Pesin [2]): f is accessible if any two points of M can be joined by a curve that is a finite union of arcs tangent to the strong bundles. Essential accessibility is the weaker property that any two measurable sets of positive measure can be joined by such a curve. In fact, accessibility will play a key role in this survey.

Pugh and Shub split their Conjecture into two sub-conjectures: (1) essential accessibility implies ergodicity, (2) the set of partially hyperbolic diffeomorphisms contains an open and dense set of accessible diffeomorphisms.

Many advances have been made since then in the ergodic theory of partially hyperbolic diffeomorphisms. In particular, there is a result by Burns and Wilkinson [4] proving that essential accessibility plus a bunching condition (trivially satisfied if the center bundle is one dimensional) implies ergodicity. There is also a result by the authors [18] obtaining the complete Pugh-Shub conjecture for one-dimensional center bundle. See [19] for a survey on the subject.

We have therefore that almost all partially hyperbolic diffeomorphisms with one dimensional bundle are ergodic. This means that the *non-ergodic* partially hyperbolic systems are very few. Can we describe them? Concretely,

QUESTION 1.1. *Which manifolds support a non-ergodic partially hyperbolic diffeomorphism? How do they look like?*

In this survey we give a description of what is known about this question for three dimensional manifolds. We study the sets of points that can be joined by paths everywhere tangent to the strong bundles (accessibility classes), and arrive, using tools of geometry of laminations and topology of 3-manifolds, to the somewhat surprising conclusion that there are strong obstructions to the non-ergodicity of a partially hyperbolic diffeomorphism. See Theorems 1.4, 1.6 and 1.7.

This gave us enough evidence to conjecture the following:

CONJECTURE 1.2 ([20]). *The only orientable manifolds supporting non-ergodic partially hyperbolic diffeomorphisms in dimension 3 are the mapping tori of diffeomorphisms of surfaces which commute with Anosov diffeomorphisms.*

Specifically, they are (1) the mapping tori of Anosov diffeomorphisms of $\mathbb{T}^2$*, (2)* $\mathbb{T}^3$*, and (3) the mapping torus of* $-id$ *where* $id : \mathbb{T}^2 \to \mathbb{T}^2$ *is the identity map on the 2-torus.*

Indeed, we believe that for 3-manifolds, all partially hyperbolic diffeomorphisms are ergodic, unless the manifold is one of the listed above.

In the case that $M = \mathbb{T}^3$ we can be more specific and we also conjecture that:

CONJECTURE 1.3. *Let* $f : \mathbb{T}^3 \to \mathbb{T}^3$ *be a conservative partially hyperbolic diffeomorphism homotopic to a hyperbolic automorphism. Then,* f *is ergodic.*

In [20] we proved Conjecture 1.2 when the fundamental group of the manifold is nilpotent:

THEOREM 1.4. *All the conservative* C^2 *partially hyperbolic diffeomorphisms of a compact orientable 3-manifold with nilpotent fundamental group are ergodic, unless the manifold is* $\mathbb{T}^3$.

A paradigmatic example is the following. Let M be the mapping torus of $A_k : \mathbb{T}^2 \to \mathbb{T}^2$, where A_k is the automorphism given by the matrix $\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$, k a non-zero integer. That is, M is the quotient of $\mathbb{T}^2 \times [0, 1]$ by the relation $\sim$, where $(x, 1) \sim (A_k x, 0)$. The manifold M has nilpotent fundamental group; in fact, it is a nilmanifold.

Theorem 1.4 then implies that all conservative partially hyperbolic diffeomorphisms of M are ergodic.

In fact, the above case, namely the case of nilmanifolds, is the only one where Theorem 1.4 is non-void (see [20]). In fact, the other cases of Theorem 1.4 are ruled out by a remarkable result by Burago and Ivanov [3]:

THEOREM 1.5 ([3]). *There are no partially hyperbolic diffeomorphisms in* $\mathbb{S}^3$ *or* $\mathbb{S}^2 \times \mathbb{S}^1$.

The proofs of most of the theorems of this survey involve deep results of the geometry of codimension one foliations and the topology of 3-manifolds. In Subsection 2.1 we shall include, for completeness, the basic facts and definitions that we shall be using in this work. However, the interested reader is strongly encouraged to consult [5], [6], [12] and [13] for a well organized and complete introduction to the subject.

Let us explain a little bit our strategy. In the first place, it follows from the results in [4, 18] that accessibility implies ergodicity. So, our strategy will be to prove that all partially hyperbolic diffeomorphisms of compact 3-manifolds except the ones of the manifolds listed in Conjecture 1.2 satisfy the (essential) accessibility property.

In dimension 3, and in fact, whenever the center bundle is 1-dimensional, the non-open accessibility classes are codimension one immersed manifolds [18]; the union of all non-open accessibility classes is a compact set *laminated* by the accessibility classes (see Subsection 2.1 for definitions). So, either f has the accessibility property or else there is a non-trivial lamination formed by non-open accessibility classes.

Let us first assume that the lamination is not a foliation (i.e. does not cover the whole manifold). Then in [20] it is showed that it either extends to a true foliation without compact leaves, or else it contains a leaf that is a periodic 2-torus with Anosov dynamics. In the first case, we have that the boundary leaves of the lamination contain a dense set of periodic points [18]. Moreover, the fundamental group of any boundary leaf injects in the fundamental group of the manifold. In the second case, let us call any embedded 2-torus admitting an Anosov dynamics extendable to the whole manifold, an *Anosov torus*. That is, $T \subset M$ is an Anosov torus if there exists a homeomorphism $h : M \to M$ such that $h|_T$ is homotopic to an Anosov diffeomorphism. In [21] we obtained that the manifold must be again one of the manifolds of Conjecture 1.2 if it has an Anosov torus.

Theorem 1.6. *A closed oriented irreducible 3-manifold admits an Anosov torus if and only if it is one of the following:*

(1) *the 3-torus*
(2) *the mapping torus of $-id$*
(3) *the mapping torus of a hyperbolic automorphism*

Let us recall that a 3-manifold is irreducible if any embedded 2-sphere bounds a ball. After the proof of the Poincaré conjecture this is the same of having trivial second fundamental group. Three dimensional manifolds supporting a partially hyperbolic diffeomorphism are always irreducible thanks to Burago and Ivanov results in [3]. Indeed, the existence of a Reebless foliation implies that the manifold is irreducible or it is $\mathbb{S}^2 \times \mathbb{S}^1$.

Secondly suppose that there are no open accessibility classes. Then, accessibility classes must foliate the whole manifold. Let us see that this foliation can not have compact leaves. Observe that any such compact leaf must be a 2-torus. So, we have three possibilities: (1) there is an Anosov torus, (2) the set of compact leaves forms a strict non-trivial lamination, (3) the manifold is foliated by 2-tori. The first case has just been ruled out. In the second case, we would have that the boundary leaves contain a dense set of periodic points, as stated above, and hence they would be Anosov tori again, which is impossible. Finally, in the third case, we conclude that the manifold is a fibration of tori over $\mathbb{S}^1$. This can only occur, in our setting, only if the manifold is the mapping torus of a diffeomorphism which commutes with an Anosov diffeomorphism as in Conjecture 1.2.

The following theorem is the first step in proving Conjectures 1.2 and 1.3. See definitions in Subsection 2.1:

THEOREM 1.7. *Let $f : M \to M$ be a conservative partially hyperbolic diffeomorphism of an orientable 3-manifold M. Suppose that the bundles E^σ are also orientable, $\sigma = s, c, u$, and that f is not accessible. Then one of the following possibilities holds:*

1. *M is the mapping torus of a diffeomorphism which commutes with an Anosov diffeomorphism as in Conjecture 1.2.*
2. *there is an f-invariant lamination $\emptyset \neq \Gamma(f) \neq M$ tangent to $E^s \oplus E^u$ that trivially extends to a (not necessarily invariant) foliation without compact leaves of M. Moreover, the boundary leaves of $\Gamma(f)$ are periodic, have Anosov dynamics and dense periodic points.*
3. *there is a minimal invariant foliation tangent to $E^s \oplus E^u$.*

The assumption on the orientability of the bundles and M is not essential, in fact, it can be achieved by a finite covering. The proof of Theorem 1.7 appears at the end of Section 5.

We do not know of any example satisfying (2) in the theorem above. We have the following question.

QUESTION 1.8. *Let $f : N \to N$ be an Anosov diffeomorphism on a complete Riemannian manifold N. Is it true that if $\Omega(f) = N$ then N is compact?*

2. Preliminaries

2.1. **Geometric preliminaries.** In this section we state several definitions and concepts that will be useful in the rest of this paper. From now on, M will be a compact connected Riemannian 3-manifold.

A *lamination* is a compact set $\Lambda \subset M$ that can be covered by open charts $U \subset \Lambda$ with a local product structure $\phi : U \to \mathbb{R}^n \times T$, where T is a locally compact subset of $\mathbb{R}^k$. On the overlaps $U_\alpha \cap U_\beta$, the transition functions $\phi_\beta \circ \phi_\alpha^{-1} : \mathbb{R}^n \times T \to \mathbb{R}^n \times T$ are homeomorphisms and take the form:

$$\phi_\beta \circ \phi_\alpha^{-1}(u, v) = (l_{\alpha\beta}(u, v), t_{\alpha\beta}(v)),$$

where $l_{\alpha\beta}$ are C^1 with respect to the u variable. No differentiability is required in the transverse direction T. The sets $\phi^{-1}(\mathbb{R}^n \times \{t\})$ are called *plaques*. Each point x of a lamination belongs to a maximal connected injectively immersed n-submanifold, called the *leaf* of x in L. The leaves are union of plaques. Observe that the leaves are C^1, but vary only continuously. The number n is the *dimension* of the lamination. If $n = \dim M - 1$, we say Λ is a *codimension-one* lamination. The set L is an *f-invariant lamination* if it is a lamination such that f takes leaves into leaves.

We call a lamination a *foliation* if $\Lambda = M$. In this case, we shall denote by $\mathcal{F}$ the set of leaves. In principle, we shall not assume any transverse differentiability. However, in case $l_{\alpha\beta}$ is C^r with respect to the v variable, we shall say that the foliation is C^r. Note that even purely C^0 codimension-one foliations admit a transverse 1-dimensional foliation (see Siebenmann [25],). In our case the existence of this 1-dimensional foliation is trivial thanks to the existence of the 1-dimensional center bundle E^c. These allows us to translate many local deformation arguments, usually given in the C^2 category, into the C^0 category as observed, for instance, by Solodov [26]. In particular, Theorems 2.1 and 2.3, which were originally formulated for C^2 foliations hold in the C^0 case. We shall say that a codimension-one foliation $\mathcal{F}$, is *transversely orientable* if the transverse 1-dimensional foliation mentioned above is orientable. An *invariant foliation* is a foliation that is an invariant lamination.

Let Λ be a codimension-one lamination that is not a foliation. A *complementary region* V is a component of $M \setminus \Lambda$. A *closed complementary region* $\hat{V}$ is the metric completion of a complementary region V with the path metric induced by the Riemannian metric, the distance between two points being the infimum of the lengths of paths in V connecting them. A closed complementary region is independent of the metric. Note that they are not necessarily compact. If Λ does not have compact leaves, then every closed complementary region decomposes into a compact *gut* piece and non-compact *interstitial regions* which are I-bundles over non-compact surfaces, and get thinner and thinner as they go away from the gut (see [13] or [8]). The interstitial regions meet the gut along annuli. The decomposition into interstitial regions and guts is unique up to isotopy. Moreover, one can take the interstitial regions as thin as one wishes.

A *boundary leaf* is a leaf corresponding to a component of ∂V, for V a closed complementary region. That is, a leaf is a non-boundary leaf if it is not contained in a closed complementary region.

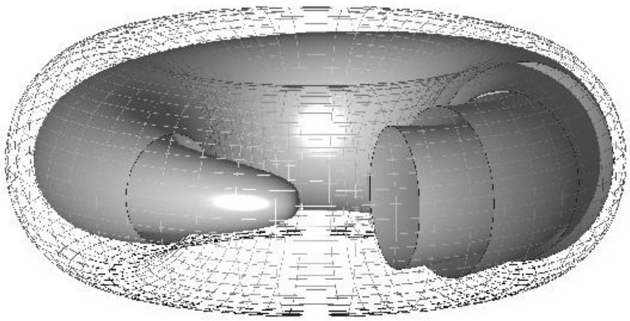


FIGURE 1. A Reeb component

The geometry of codimension-one foliations is deeply related to the topology of the manifold that supports them. The following subset of a foliation is important in their description. A *Reeb component* is a solid torus whose interior is foliated by planes transverse to the of core of the solid torus, such that each leaf limits on the boundary torus, which is also a leaf (see Figure 1). A foliation that has no Reeb components is called *Reebless*.

The following theorems show better the above mentioned relation:

THEOREM 2.1 (Novikov). *Let M be a compact orientable 3-manifold and $\mathcal{F}$ a transversely orientable codimension-one foliation. Then each of the following implies that $\mathcal{F}$ has a Reeb component:*

(1) *There is a closed, nullhomotopic transversal to $\mathcal{F}$*
(2) *There is a leaf L in $\mathcal{F}$ such that $\pi_1(L)$ does not inject in $\pi_1(M)$*

The statement of this theorem can be found, for instance, in [6, Theorems 9.1.3 & 9.1.4., p.288]. We shall also use the following theorem

THEOREM 2.2 (Haefliger). *Let Λ be a codimension one lamination in M. Then the set of points belonging to compact leaves is compact.*

This theorem was originally formulated for foliations [10]. However, it also holds for laminations, see for instance [13].

We have the following consequence of Novikov's Theorem about Reebless foliations. This theorem is stated in [24] as Corollary 2 on page 44.

THEOREM 2.3. *If M is a compact 3-manifold and $\mathcal{F}$ is a transversely orientable codimension-one Reebless foliation, then either $\mathcal{F}$ is the product foliation of $\mathbb{S}^2 \times \mathbb{S}^1$, or $\tilde{\mathcal{F}}$, the foliation induced by $\mathcal{F}$ on the universal cover $\tilde{M}$ of M, is a foliation by planes $\mathbb{R}^2$. In particular, if $M \neq \mathbb{S}^2 \times \mathbb{S}^1$ then M is irreducible.*

This theorem was originally stated for C^2 foliations, but it also holds for C^0 foliations, due to Siebenmann's theorem mentioned above.

2.2. **Topologic preliminaries.** Let M be a 3-dimensional manifold. A manifold M is *irreducible* if every 2-sphere $\mathbb{S}^2$ embedded in the manifold bounds a 3-ball. Recall that a 2-torus T embedded in M is an *Anosov torus* if there exists a diffeomorphism $f : M \to M$ such that $f(T) = T$ and the action induced by f on $\pi_1(T)$, that is, $f_\#|_T : \pi_1(T) \to \pi_1(T)$, is a hyperbolic automorphism. Equivalently, f restricted to T is isotopic to a hyperbolic automorphism.

We shall assume from now on, that M is an irreducible 3-manifold since this is the case for 3-manifolds supporting partially hyperbolic diffeomorphisms. In this subsection, we will focus on what is called the JSJ-decomposition of M (see below). That is, we will cut M along certain kind of tori, called incompressible, and will obtain certain 3-manifolds with boundary that are easier to handle, which are, respectively, Seifert manifolds, and atoroidal and acylindrical manifolds. Let us introduce these definitions first.

An orientable surface S embedded in M is *incompressible* if the homomorphism induced by the inclusion map $i_\# : \pi_1(S) \hookrightarrow \pi_1(M)$ is injective; or, equivalently, if there is no embedded disc $D^2 \subset M$ such that $D \cap S = \partial D$ and $\partial D \not\sim 0$ in S (see, for instance, [12, Page 10]). We also require that $S \neq \mathbb{S}^2$.

A manifold with or without boundary is *Seifert*, if it admits a one dimensional foliation by closed curves, called a Seifert fibration. The boundary of an orientable Seifert manifold with boundary consists of finite union of tori. There are many examples of Seifert manifolds, for instance $\mathbb{S}^3$, T_1S where S is a surface, etc.

The other type of manifold obtained in the JSJ-decomposition is atoroidal and acylindrical manifolds. A 3-manifold with boundary N is *atoroidal* if every incompressible torus is *∂-parallel*, that is, isotopic to a subsurface of ∂N. A 3-manifold with boundary N is *acylindrical* if every incompressible annulus A that is *properly embedded*, i.e. $\partial A \subset \partial N$, is ∂-parallel, by an isotopy fixing ∂A.

As we mentioned before, a closed irreducible 3-manifold admits a natural decomposition into Seifert pieces on one side, and atoroidal and acylindrical components on the other:

Theorem 2.4 (JSJ-decomposition [14], [15]). *If M is an irreducible closed orientable 3-manifold, then there exists a collection of disjoint incompressible tori $\mathcal{T}$ such that each component of $M \setminus \mathcal{T}$ is either Seifert, or atoroidal and acylindrical. Any minimal such collection is unique up to isotopy. This means that if $\mathcal{T}$ is a collection as described above, it contains a minimal sub-collection $m(\mathcal{T})$ satisfying the same claim. All collections $m(\mathcal{T})$ are isotopic.*

Any minimal family of incompressible tori as described above is called a *JSJ-decomposition* of M. When it is clear from the context we shall also call JSJ-decomposition the set of pieces obtained by cutting the manifold along these tori. Note that if M is either atoroidal or Seifert, then $\mathcal{T} = \emptyset$.

2.3. **Dynamic preliminaries.** Throughout this paper we shall work with a *partially hyperbolic diffeomorphism* f, that is, a diffeomorphism admitting a non-trivial Tf-invariant splitting of the tangent bundle

$TM = E^s \oplus E^c \oplus E^u$, such that all unit vectors $v^\sigma \in E^\sigma_x$ ($\sigma = s, c, u$) with $x \in M$ verify:

$$\|T_x f v^s\| < \|T_x f v^c\| < \|T_x f v^u\|$$

for some suitable Riemannian metric. f also must satisfy that $\|Tf|_{E^s}\| < 1$ and $\|Tf^{-1}|_{E^u}\| < 1$. We shall say that a partially hyperbolic diffeomorphism f that satisfies

$$\|T_x f v^s\| < \|T_y f v^c\| < \|T_z f v^u\| \ \forall x, y, z \in M$$

is *absolutely partially hyperbolic.*

We shall also assume that f is *conservative*, i.e. it preserves Lebesgue measure associated to a smooth volume form.

It is a known fact that there are foliations $\mathcal{W}^\sigma$ tangent to the distributions E^σ for $\sigma = s, u$ (see for instance [2]). The leaf of $\mathcal{W}^\sigma$ containing x will be called $W^\sigma(x)$, for $\sigma = s, u$. The connected component of x in the intersection of $W^s(x)$ with a small ε-ball centered at x is the ε-*local stable manifold of* x, and is denoted by $W^s_\varepsilon(x)$.

In general it is not true that there is a foliation tangent to E^c. It is false even in case $\dim E^c = 1$ (see [22]). However, in Proposition 3.4 of [1] it is shown that if $\dim E^c = 1$, then f is *weakly dynamically coherent*. This means that for each $x \in M$ there are complete immersed C^1 manifolds which contain x and are everywhere tangent to E^c, E^{cs} and E^{cu}, respectively. We will call a *center curve* any curve which is everywhere tangent to E^c. Moreover, we will use the following fact:

PROPOSITION 2.5 ([1]). *If γ is a center curve through x, then*

$$W^s_\varepsilon(\gamma) = \bigcup_{y \in \gamma} W^s_\varepsilon(y) \qquad \text{and} \qquad W^u_\varepsilon(\gamma) = \bigcup_{y \in \gamma} W^u_\varepsilon(y)$$

are C^1 immersed manifolds everywhere tangent to $E^s \oplus E^c$ and $E^c \oplus E^u$ respectively.

We shall say that a set X is s-*saturated* or u-*saturated* if it is a union of leaves of the strong foliations $\mathcal{W}^s$ or $\mathcal{W}^u$ respectively. We also say that X is su-saturated if it is both s- and u-saturated. The *accessibility class* $AC(x)$ of the point $x \in M$ is the minimal su-saturated set containing x. Note that the accessibility classes form a partition of M. If there is some $x \in M$ whose accessibility class is M, then the diffeomorphism f is said to have the *accessibility property*. This is equivalent to say that any two points of M can be joined by a path which is piecewise tangent to E^s or to E^u. A diffeomorphism is said to be *essentially accessible* if any su-saturated set has full or null measure.

The theorem below relates accessibility with ergodicity. In fact it is proven in a more general setting, but we shall use the following formulation:

THEOREM 2.6 ([4],[18]). *If f is a C^2 conservative partially hyperbolic diffeomorphism with the (essential) accessibility property and $\dim E^c = 1$, then f is ergodic.*

In [20] it is proved that there are manifolds whose topology implies the accessibility property holds for all partially hyperbolic diffeomorphisms. In these manifolds, all partially hyperbolic diffeomorphisms are ergodic.

Sometimes we will focus on the openness of the accessibility classes. Note that the accessibility classes form a partition of M. If all of them are open then, in fact, f has the accessibility property. We will call $U(f) = \{x \in M; AC(x) \text{ is open}\}$ and $\Gamma(f) = M \setminus U(f)$. Note that f has the accessibility property if and only if $\Gamma(f) = \emptyset$. We have the following property of non-open accessibility classes:

PROPOSITION 2.7 ([18]). *The set $\Gamma(f)$ is a codimension-one lamination, having the accessibility classes as leaves.*

In fact, any compact su-saturated subset of $\Gamma(f)$ is a lamination.

The above proposition is Proposition A.3. of [18]. The fact that the leaves of $\Gamma(f)$ are C^1 may be found in [7]. The following proposition is Proposition A.5 of [18]:

PROPOSITION 2.8 ([18]). *If Λ is an invariant sub-lamination of $\Gamma(f)$, then each boundary leaf of Λ is periodic and the periodic points are dense in it (with the induced topology).*

Moreover, the stable and unstable manifolds of each periodic point are dense in each plaque of a boundary leaf of Λ.

Observe that the proof of Proposition A.5 of [18] shows in fact that periodic points are dense in the accessibility classes of the boundary leaves of V endowed with its intrinsic topology. In other words, periodic points are dense in each plaque of the boundary leaves of V.

We shall also use the following theorem by Brin, Burago and Ivanov, whose proof is in [1], after Proposition 2.1.

THEOREM 2.9 ([1]). *If $f : M^3 \to M^3$ is a partially hyperbolic diffeomorphism, and there is an open set V foliated by center-unstable leaves, then there cannot be a closed center-unstable leaf bounding a solid torus in V.*

3. ANOSOV TORI

In this section we will say a few words about the proof of Theorem 1.6. The idea in its proof is that, given an Anosov torus T, we can "place" T so that either T belongs to the family $\mathcal{T}$ given by the JSJ-decomposition (Theorem 2.4), or else T is in a Seifert component, and it is either transverse to all fibers, or it is union of fibers of this Seifert component. See Proposition 3.3.

It is important to note the following property of Anosov tori:

THEOREM 3.1 ([20]). *Anosov tori are incompressible.*

An Anosov torus in an atoroidal component will then be ∂-parallel to a component of its boundary. In this case, we can assume $T \in \mathcal{T}$. On the other hand, the Theorem of Waldhausen below, guarantees that we can always place an incompressible torus in a Seifert manifold in a "standard" form; namely, the following: a surface is *horizontal* in a Seifert manifold if it is transverse to all fibers, and *vertical* if it is union of fibers:

THEOREM 3.2 (Waldhausen [27]). *Let M be a compact connected Seifert manifold, with or without boundary. Then any incompressible surface can be isotoped to be horizontal or vertical.*

The architecture of the proof of Theorem 1.6 is contained in the following proposition.

Proposition 3.3. *Let T be an Anosov torus of a closed irreducible orientable manifold M. Then, there exists a diffeomorphism $f: M \to M$ and a JSJ-decomposition $\mathcal{T}$ such that*

(1) *$f|T$ is a hyperbolic toral automorphism,*
(2) *$f(\mathcal{T}) = \mathcal{T}$, and*
(3) *one of the following holds*
    (a) *$T \in \mathcal{T}$*
    (b) *T is a vertical torus in a Seifert component of $M \setminus \mathcal{T}$, and T is not ∂-parallel in this component.*
    (c) *M is a Seifert manifold ($\mathcal{T} = \emptyset$), and T is a horizontal torus,*

The proposition above allows us to split the proof of Theorem 1.6 into cases. Note that case (3b) includes the case in which M is a Seifert manifold and T is a vertical torus.

In the case that T is a vertical torus in a Seifert component we can cut this component along T. Then we can suppose that T is in the boundary. We take profit of the fact that in most manifolds the Seifert fibration is unique up to isotopy. Since the dynamics restricted to T is Anosov we have that the manifold has more than one Seifert fibration. This lead us to show that this Seifert component must be $\mathbb{T}^2 \times [0,1]$. This gives that the whole manifold must be one of the manifolds of Theorem 1.6.

If T is horizontal torus then the manifold M is Seifert and T intersects all the fibers. This is discarded in a case by case study thanks to the fact that the Seifert manifolds having horizontal torus a finite.

The last and more difficult case is when T is part of the JSJ-decomposition but it is not the boundary of a Seifert component. The proof in this case is complicated but a very rough idea is to take a properly embedded surface S with an essential circle of T in its boundary. Taking a large iterate $f^n(S)$ and considering $S \cap f^n(S)$, it is possible to construct a non-parallel incompressible cylinder as a union of a band in S and a band in $f^n(S)$. This leads to contradiction because the component is not Seifert and then, it is acylindrical.

4. The su-lamination $\Gamma(f)$

Let f be a partially hyperbolic diffeomorphism of a compact 3-manifold M. From Subsection 2.3 it follows that we have three possibilities: (1) f has the accessibility property, (2) the union of all non-open accessibility classes is a strict lamination, $\emptyset \subsetneq \Gamma(f) \subsetneq M$ or (3) the union of all non-open accessibility classes foliates M: $\Gamma(f) = M$.

Now, we shall distinguish two possible cases in situations (2) and (3):

(a) the lamination $\Gamma(f)$ does not contain compact leaves
(b) the lamination $\Gamma(f)$ contains compact leaves

In this section we deal with the case (2a). In fact, for our purposes it will be sufficient to assume that there exists an f-invariant sub-lamination Λ of $\Gamma(f)$ without compact leaves. Section 5 treats the cases (2b) and (3b). Section 6 treats the case (3a).

In this section, we will prove that the complement of Λ consists of I-bundles. To this end, we shall assume that the bundles E^σ ($\sigma = s, c, u$) and the manifold M are orientable (we can achieve this by considering a finite covering).

Theorem 4.1 ([20], Theorem 4.1). *If $\emptyset \subsetneq \Lambda \subset \Gamma(f)$ is an orientable and transversely orientable f-invariant sub-lamination without compact leaves such that $\Lambda \neq M$, then all closed complementary regions of Λ are I-bundles.*

Theorem 4.1 was proved by showing:

Proposition 4.2. *Let $\Lambda \subset \Gamma(f)$ be a nonempty f-invariant sub-lamination without compact leaves. Then E^c is uniquely integrable in the closed complementary regions of Λ.*

The proof of this proposition is rather technical. The interested reader may found a proof in [20].

Let us consider $\hat{V}$ a closed complementary region of Λ, and call $\mathcal{I}(V)$ the union of all interstitial regions of V and $\mathcal{G}(V)$ the gut of $\hat{V}$ (see Subsection 2.1), so that

$$\hat{V} = \mathcal{I}(V) \cup \mathcal{G}(V).$$

The following statement is rather standard:

LEMMA 4.3. *Let $f : M \to M$ be a partially hyperbolic diffeomorphism. If U is an open invariant set such that $U \subset \Omega(f)$, then the closure of U is su-saturated.*

Let us observe that if $\hat{V}$ is connected then there are only two boundary leaves of $\hat{V}$. Indeed, as we mentioned before periodic points are dense in boundary leaves. This fact jointly with the local product structure imply, using standard arguments, that the stable and unstable leaves of periodic points are dense too. Take a periodic point p in a boundary leaf and in the intersticial region. There are center curves joining the points in the local stable manifold of p with other boundary curve L_1 of $\hat{V}$ (the same property holds for the local unstable manifold). Invariance of the stable manifold of p and boundary leaves give that the center curve of any point of the stable manifold joins the boundary leaf L_0 containing p with L_1. Denseness of the stable and unstable manifolds of p implies that the complement of the set of points such that their center manifold join L_0 with L_1 is totally disconnected. Then, it is not difficult to see that L_0 and L_1 are the unique boundary leaves of $\hat{V}$.

Also, since periodic points are dense in the boundary leaves due to Proposition 2.8, there is an iterate of f that fixes all connected components of $\hat{V}$, so we will assume when proving Theorem 4.1 that $\hat{V}$ is connected and has two boundary leaves L_0 and L_1.

Proof of Theorem 4.1. We will present a sketch of a different approach to a proof than the one given in [20]. The strategy will be to show that all center leaves in $\hat{V}$ meet both L_0 and L_1. Let p be a periodic point in $L_0 \cap \mathcal{I}(V)$. As we mentioned before its center leaf meets L_1, and the same happens for all points in its stable and unstable manifolds. Now stable and unstable manifolds of a periodic point are dense in each plaque of L_0 (Proposition 2.8). So the set of points in L_0 whose center leaf does not reach L_1 is contained in a totally disconnected set.

Let us suppose that x_0 is a point in L_0 whose center leaf does not reach L_1. Then, since center curves of points of the intersticial region clearly reach the boundary, $W^c(x_0)$ is contained in $\mathcal{G}(V)$. Take a small rectangle R in L_0 around x_0 formed by arcs of stable and unstable manifolds of a periodic point. Moreover, we can assume that the center curves of the points of R_0 reach L_1. Of course, the image is another rectangle R_1 formed by stable and unstable arcs. Then, the center arcs of the points of R_0 and the interiors of R_0 and R_1 form a 2-sphere S. Since Rosenberg's theorem [23] remains valid in this setting and $\hat{V}$ is foliated by $\mathcal{W}^{cs}$ that is Reebless and transverse to the boundary, we have that $\hat{V}$ is irreducible. Then, S bounds a ball B. Now, since $W^c(x_0)$ does not reach L_1 and is contained in B, it accumulates in B but Novikov's Theorem implies the existence of a Reeb component, a contradiction. □

Theorem 4.1 implies that any non trivial invariant sub-lamination $\Lambda \subset \Gamma(f)$ without compact leaves can be extended to a foliation of M without compact leaves. Indeed, any complementary region V is an I-bundle, and hence it is diffeomorphic to the product of a boundary leaf times the open interval: $L_0 \times (0,1)$. The foliation $F_t = L_0 \times \{t\}$ induces a foliation of V.

This has the following consequence in case the fundamental group of M is nilpotent:

Proposition 4.4. *If M is a compact 3-manifold with nilpotent fundamental group, and $\emptyset \subsetneq \Lambda \subsetneq M$, is an invariant sub-lamination of $\Gamma(f)$, then there exists a leaf of Λ that is a periodic 2-torus with Anosov dynamics.*

Proof. If Λ has a compact leaf, let us consider the set Λ_c of all compact leaves of Λ. Λ_c is in fact an invariant sub-lamination, due to Theorem 2.2. Hence Proposition 2.8 implies that the boundary leaves of Λ_c are periodic 2-tori with Anosov dynamics, and we obtain the claim.

If, on the contrary, Λ does not have compact leaves, then due to Theorem 4.1 above, we can extend Λ to a foliation $\mathcal{F}$ of M without compact leaves. In particular, $\mathcal{F}$ is a Reebless foliation. Item (2) of Theorem 2.1 implies that for all boundary leaves L of Λ, $\pi_1(L)$ injects in $\pi_1(M)$, and is therefore nilpotent.

Now, this implies that the boundary leaves can only be planes or cylinders. Theorem 2.8 implies that stable and unstable leaves of periodic points are dense in those leaves, which is impossible for the case of the plane or the cylinder. Therefore, Λ must contain a compact leaf, and due to what was shown above, it must contain a periodic 2-torus with Anosov dynamics. □

In fact, Theorem 1.6 implies that periodic 2-tori with Anosov dynamics are not possible in 3-manifolds with nilpotent fundamental group, unless the manifold is $\mathbb{T}^3$. Hence the hypotheses of Proposition 4.4 are not fulfilled, unless the manifold is $\mathbb{T}^3$. This will eliminate case (2) mentioned at the beginning of this section.

5. The trichotomy of Theorem 1.7

In this section we will prove Theorem 1.7. This theorem and the results in this section are valid for any 3-manifold M, and do not require that its fundamental group be nilpotent. Moreover, Theorem 3.1 does not even require the existence of a partially hyperbolic diffeomorphism.

Let T be an embedded 2-torus in M. We shall call T an ***Anosov torus*** if there exists a homeomorphism $g : M \to M$ such that T is g-invariant, and $g|_T$ is homotopic to an Anosov diffeomorphism.

Also, let S be a two-sided embedded closed surface of M^3 other than the sphere. S is ***incompressible*** if and only if the homomorphism induced by the inclusion map $i_{\#} : \pi_1(S) \hookrightarrow \pi_1(M)$ is injective; or, equivalently, after the Loop Theorem, if there is no embedded disc $D^2 \subset M$ such that $D \cap S = \partial D$ and $\partial D \nsim 0$ in S (see, for instance, [12]).

Recall that Theorem 3.1 says that Anosov tori are incompressible. We insist that this theorem is general, and does not depend on the existence of a partially hyperbolic dynamics in the manifold.

We also need the following fact about codimension one laminations.

Theorem 5.1. *Let $\mathcal{F}$ be a codimension one C^0-foliation without compact leaves of a three dimensional compact manifold M. Then, $\mathcal{F}$ has a finite number of minimal sets.*

We are now in position to prove Theorem 1.7 of Page 143:

Proof of Theorem 1.7. If $\Gamma(f) = M$ then there are no Reeb components. Indeed, since f is conservative, if there were a Reeb component, then its boundary torus should be periodic. We get a contradiction from Theorem 3.1. This gives case (3) except the minimality.

Let us assume that $\Gamma(f) \neq M$. If $\Gamma(f)$ contains a compact leaf then the set of compact leaves is a sub-lamination Λ of $\Gamma(f)$ by Theorem 2.2. Proposition 2.8 implies that the boundary leaves of Λ are Anosov tori, and we obtain case (1) as a consequence of Theorem 1.6.

If $\Gamma(f) \neq M$ and contains no compact leaves, then Theorem 4.1 and Proposition 2.8 give us case (2).

Finally we show minimality in case (3). On the one hand, if $\Gamma(f) = M$ and has a compact leaf we have two possibilities: either all leaves are compact or not. If not then, the previous argument implies the existence of an Anosov torus and we are in case (1). If all leaves are compact, as we mentioned before, the manifold is a torus bundle and the hyperbolic dynamics on fibers implies that we are again in case (1). On the other hand, if $\Gamma(f)$ has no compact leaves and has a minimal sub-lamination $\mathcal{L}$, we have that $\mathcal{L}$ is periodic (recall that minimal sub-laminations of a codimension one foliation are finite, Theorem 5.1). Then, we are again in case (2). □

6. Nilmanifolds

This section deals with the proof of Theorem 1.4. Let $f : M \to M$ be a conservative partially hyperbolic diffeomorphism of a compact orientable three dimensional nilmanifold $M \neq \mathbb{T}^3$. As consequence of Proposition 4.4 and Theorems 1.6 and 1.7 we have that $E^s \oplus E^u$ integrates to a minimal foliation $\mathcal{F}^{su}$ if f does not have the accessibility property. Indeed the only possibilities in the trichotomy of Theorem 1.7 are (2) and (3) and Proposition 4.4 says that there is an Anosov torus if we are in case (2). But this last case is impossible for a nilmanifold $M \neq \mathbb{T}^3$. In this section we shall give some arguments showing that the existence of a minimal foliation tangent to $E^s \oplus E^u$ leads us to a contradiction. In [20] the reader can find a different proof of the same fact. Without loss of generality we may assume, by taking a double covering if necessary, that $\mathcal{F}^{su}$ is transversely orientable. Observe that the double covering of a nilmanifold is again a nilmanifold.

The first step is that Parwani [16] proved (following Burago-Ivanov [3] arguments) that the action induced by f in the first homology group of M is hyperbolic. By duality the same is true for the first cohomology group.

The second step is given by Plante results in [17] (see also [13]). Since $\mathcal{F}^{su}$ is a minimal foliation of a manifold whose fundamental group has non-exponential growth there exists a transverse holonomy invariant measure μ of full support. This measure is unique up to multiplication by a constant and represents an element of the first cohomology group of M. The action of f leaves $\mathcal{F}^{su}$ invariant and induces a new transverse measure ν, an image of former one. The uniqueness implies that $\nu = \lambda\mu$ for some $\lambda > 0$. Since of the action of f on $H^1(M)$ is hyperbolic, then $\lambda \neq 1$. Suppose that $\lambda > 1$ (if the contrary is true take f^{-1}).

The third step is to observe that $\lambda > 1$ implies that f is expanding the μ measure of center curves. Since μ has full support and the $su-$bundle is hyperbolic we would obtain that f is conjugated to Anosov leading to contradiction with the fact that $M \neq \mathbb{T}^3$.

7. $M = \mathbb{T}^3$

In this section we present the results announced by Hammerlindl and Ures on Conjecture 1.3, that the nonexistence of nonergodic partially hyperbolic diffeomorphisms homotopic to Anosov in dimension 3. They are able to prove the following result.

THEOREM 7.1 ([11]). *Let $f : \mathbb{T}^3 \to \mathbb{T}^3$ be a $C^{1+\alpha}$ conservative partially hyperbolic diffeomorphism homotopic to a hyperbolic automorphism A. Suppose that f is not ergodic. Then,*

1. *$E^s \times E^u$ integrates to a minimal foliation.*
2. *f is topologically conjugated to A and the conjugacy sends strong leaves of f into the corresponding strong leaves of A.*
3. *The center Lyapunov exponent is 0 a.e.*

We remark that it is not known if there exists a diffeomorphism satisfying the conditions of the theorem above.

Now, in order to prove Conjecture 1.3 we have two possibilities: either we prove that a diffeomorphism satisfying the conditions of Theorem 7.1 is ergodic or we prove that such a diffeomorphism cannot exist. Hammerlindl and Ures announced that if f is C^2 and the center stable and center unstable leaves of a periodic point are C^2 then, f is ergodic.

References

[1] M. Brin, D. Burago, S. Ivanov, On partially hyperbolic diffeomorphisms of 3-manifolds with commutative fundamental group, Modern Dynamical Systems and Applications, B. Hasselblatt, M. Brin and Y. Pesin, eds, Cambridge Univ. Press, New York (2004), 307–312.

[2] M. Brin, Ya Pesin, Partially hyperbolic dynamical systems, *Math. USSR Izv.***8**, (1974) 177-218.

[3] D. Burago, S. Ivanov, Partially hyperbolic diffeomorphisms of 3-manifolds with abelian fundamental groups, *J. of Modern Dynamics* **4** (2008), 541–580.

[4] K. Burns, A. Wilkinson, On the ergodicity of partially hyperbolic systems, *Ann. Math.***171** (2010), 451–489.

[5] A. Candel, L. Conlon, Foliations I, Graduate Studies in Mathematics, vol. 23, American Mathematical Society, Providence, RI (2003).

[6] A. Candel, L. Conlon, Foliations II, Graduate Studies in Mathematics, vol. 60, American Mathematical Society, Providence, RI (2003).

[7] P. Didier, Stability of accessibility, *E*rg.Th & Dyn. Sys. **23**, no.6, (2003), 1717-1731.

[8] D. Gabai, W. Kazez, Group negative curvature for 3-manifolds with genuine laminations, *Geom. Topol.* **2** (1998), 65–77.

[9] M. Grayson, C. Pugh, M. Shub, Stably ergodic diffeomorphisms, *Ann. Math.* **140** (1994), 295–329.

[10] A. Haefliger, Variétés feuilletés, *Topologia Differenziale (Centro Int. Mat. Estivo, 1 deg Ciclo, Urbino (1962)). Lezione 2 Edizioni Cremonese, Rome 367-397 (1962)*

[11] A. Hammerlindl, R. Ures, Partial hyperbolicity and ergodicity in $\mathbb{T}^3$, in preparation.

[12] A. Hatcher, Notes on basic 3-manifold topology, Hatcher's webpage at Cornell Math. Dep. (http://www.math.cornell.edu/ hatcher/3M/3Mdownloads.html)

[13] G. Hector, U. Hirsch, Introduction to the geometry of foliations: Part B, Aspects of Mathematics, Vieweg, Braunschweig/Wiesbaden (1987).
[14] W. Jaco, P. Shalen, Seifert fibered spaces in 3-manifolds, *Memoirs AMS* **21**, 220 (1979).
[15] K. Johannson, *Homotopy equivalences of 3-manifolds with boundaries*, Lecture Notes in Math. **761**, Spriinger Verlag, 1979.
[16] K. Parwani, On 3-manifolds that support partially hyperbolic diffeomorphisms. Nonlinearity 23 (2010), 589–606.
[17] J.F. Plante, Foliations with measure preserving holonomy, *Annals of Math.* **102** (1975), 327–361.
[18] F. Rodriguez Hertz, M. Rodriguez Hertz, R. Ures, Accessibility and stable ergodicity for partially hyperbolic diffeomorphisms with 1d-center bundle, *Invent. Math.* 172 (2008), 353–381..
[19] F. Rodriguez Hertz, M. Rodriguez Hertz, R. Ures, A survey of partially hyperbolic dynamics, in “Partially Hyperbolic Dynamics, Laminations, and Teichmüller Flow,” (eds G. Forni, M. Lyubich, C. Pugh and Michael Shub), Fields Inst. Comm., AMS, (2007), 35-88.
[20] F. Rodriguez Hertz, M. Rodriguez Hertz, R. Ures, Partial hyperbolicity and ergodicity in dimension three, *J. of Modern Dynamics* **2** (2008), 187–208.
[21] F. Rodriguez Hertz, M. Rodriguez Hertz, R. Ures, Tori with hyperbolic dynamics in 3-manifolds, preprint, (2010).
[22] F. Rodriguez Hertz, M. Rodriguez Hertz, R. Ures, A non dynamically coherent example of $\mathbb{T}^3$, in preparation.
[23] H. Rosenberg, Foliations by planes. *Topology* **7** (1968), 131-138.
[24] R. Roussarie, Sur les feuilletages des variétès de dimension trois, *Ann. de l’Institut Fourier*, **21** (1971), 13–82.
[25] L. Siebenmann, Deformations of homeomorphisms on stratified sets, *Comment. Math. Helv.* **47** (1972) 123-163.
[26] V.V. Solodov, Components of topological foliations, *Mat. Sb.(N.S.)*, **119** (1982), 340–354.
[27] F. Waldhausen, Eine Klasse von 3-dimensionalen Mannigfaltigkeiten. I, II, *Invent. Math.*, **3**, (1967) 308–333; ibid., **4**, (1967) 87–117.

IMERL-Facultad de Ingeniería, Universidad de la República, CC 30 Montevideo, Uruguay

E-mail address: `frhertz@fing.edu.uy`
E-mail address: `jana@fing.edu.uy`
E-mail address: `ures@fing.edu.uy`

Las Publicaciones Matemáticas del Uruguay (PMU) tienen como objetivo reflejar parte de las actividades de investigación matemática que se lleva a cabo en Uruguay. Nuestro interés es publicar artículos de investigación, así como artículos de tipo survey, anuncios, y otros trabajos que el comité editorial considere adecuado.

Los volúmenes no necesariamente serán arbitrados. Esto se indicará cuidadosamente en cada volumen.

Todos los artículos de este volumen han sido arbitrados.

The goal of Publicaciones Matemáticas del Uruguay (PMU) is to reflect part of the mathematics research activities taking place in Uruguay. It is our interest to publish research articles, survey-type articles, research announcements and other papers considered suitable by the Editorial Board.

The editorial process may or may not involve a revision by referees. This will be carefully indicated in each volume.

All papers in this volume have been peer-reviewed.

Contents

Preface

A review of some recent results on Random Polynomials over $\mathbb{R}$ and over $\mathbb{C}$.
DIEGO ARMENTANO 1

Rice formulas and Gaussian waves II.
JEAN-MARC AZAÏS, JOSÉ R. LEÓN, and MARIO WSCHEBOR 15

On automorphism groups of fiber bundles
MICHEL BRION 39

On the focusing of Cramér - von Mises test
ALEJANDRA CABAÑA and ENRIQUE CABAÑA 67

Feuilletage de Hirsch, mesures harmoniques et g-mesures
BERTRAND DEROIN and CONSTANTIN VERNICOS 79

On existence of smooth critical subsolutions of the Hamilton-Jacobi Equation
ALBERT FATHI 87

Paths towards adaptive estimation for Instrumental Variable Regression
JEAN-MICHEL LOUBES and CLÉMENT MARTEAU 99

Semisimple Hopf algebras and their representations
SONIA NATALE 123

An example concerning the Theory of Levels for codimension-one foliations
ANDRÉS NAVAS 169

Accessibility and abundance of ergodicity in dimension three: a survey.
FEDERICO RODRIGUEZ HERTZ, JANA RODRIGUEZ HERTZ, and RAÚL URES 177