Knots and dynamics
Étienne Ghys

Abstract. The trajectories of a vector field in 3-space can be very entangled; the flow can swirl, spiral, create vortices etc. Periodic orbits define knots whose topology can sometimes be very complicated. In this talk, I will survey some advances in the qualitative and quantitative description of this kind of phenomenon. The first part will be devoted to vorticity, helicity, and asymptotic cycles for flows. The second part will deal with various notions of rotation and spin for surface diffeomorphisms. Finally, I will describe the important example of the geodesic flow on the modular surface, where the linking between geodesics turns out to be related to well-known arithmetical functions.

Mathematics Subject Classification (2000). Primary 37–02; Secondary 57–02.

Keywords. Low-dimensional dynamical systems, periodic orbits, knots and links.

1. Flows

1.1. Vorticity. Let us start with some historical motivation. Consider a perfect incompressible fluid moving inside some bounded domain \( M \) in 3-space, with no external forces. At time \( t \), the velocity is described by a divergence free vector field \( v_t \), tangent to the boundary of \( M \), which evolves in time according to the classical Euler equation:

\[
\frac{D}{Dt} v_t = \nabla \cdot (\nabla p - v_t \cdot \nabla v_t) + \nabla \times \omega_t
\]

where \( \omega_t = \nabla \times v_t \) is known as the vorticity vector field.

One of the earliest results in fluid dynamics is due to H. Helmholtz and W. Kelvin:

*The vorticity \( \omega_t \) is merely transported by the flow, i.e. at any time \( t \), one has \( \omega_t = d\phi^t(\omega_0) \).*

This is not difficult to prove: take a closed loop \( c \) in \( M \), and compute the time derivative of the circulation of \( v_t \) along the loop \( c_t = \phi^t(c) \).

\[
\frac{D}{Dt} \left( \oint_{c_t} v_t \cdot dc_t \right) = \oint_{c_t} \left( \frac{Dv_t}{Dt} \cdot dc_t + v_t \cdot \frac{Dc_t}{Dt} \right) = \oint_{c_t} \left( -\nabla p \cdot dc_t + v_t \cdot dv_t \right) = \oint_{c_t} d(-p + \frac{|v_t|^2}{2}) = 0.
\]
When \( \epsilon \) reduces to an infinitesimal loop, Stokes’ formula shows that \( d\phi^{-1}(\omega_t) \) is indeed constant in time.

A much more conceptual proof is due to V. Arnold who realized that Euler’s equation can be seen as the geodesic flow on the infinite dimensional Lie group of volume preserving diffeomorphisms of \( M \), equipped with a natural right invariant metric [3], [5], [55]. This right invariance implies a symmetry group for the equation, which yields Helmholtz–Kelvin’s result as a special case of Noether’s general principle that symmetries imply conservation laws.

If one can define quantities associated to divergence free vector fields, which are invariant under conjugacies by volume preserving diffeomorphisms, these quantities evaluated on the vorticity \( \omega_t \) will therefore be constants of motion. In this talk, we will discuss some of these invariants, of topological origin.

One consequence seemed remarkable to W. Kelvin. Suppose that at time 0, the vector field \( v_0 \) possesses a vortex ring: a solid torus \( \mathbb{S}^1 \times \mathbb{D} \) embedded in \( M \) in such a way that \( \omega_0 \) is tangent to its boundary. Then, this ring will survive as a vortex ring under time evolution, preserving the same topology. This stability of vortices was the starting point of the (now forgotten) theory of “vortex atoms”, trying to explain elementary “atoms” as vortex rings in ether. Even though this turned out to be physically incorrect, it represents one of the first attempts to use topology in physics. In any case, it motivated P. Tait to start a systematic study of knots, therefore creating knot theory. See [26] for a fascinating historical survey of this great moment of interaction between physics and mathematics.

A similar phenomenon appeared much more recently in magneto-hydrodynamics: the dynamics of electrically conducting fluids (like a plasma). If one assumes that the fluid is perfect and has no resistance (ideal MHD), the magnetic (divergence free) vector field is merely transported by the flow of the fluid [21]. For instance, if two periodic orbits of the magnetic field are linked at time \( t = 0 \), these orbits will survive for ever and remain linked. Again, an invariant of divergence free vector fields yields conservation laws. See for instance [5], [15].

There are many wonderful examples of vector fields in 3-space whose phase portraits exhibit a rich topology and which obviously deserve a topological study. As a typical example, the Lorenz equation also originated from fluid dynamics:

\[
\frac{dx}{dt} = 10(y - x); \quad \frac{dy}{dt} = 28x - y - xz; \quad \frac{dz}{dt} = xy - 2.67z. 
\]

It has been extensively analyzed since the 1980s, and is now a paradigm of a “robust” dynamical system (see in particular the papers of J. Guckenheimer and R. Williams [48], [90], and the book [83]). Note that this vector field is not volume preserving, but admits many invariant measures.

Measure preserving flows do not only arise from physical considerations. Consider for instance a discrete subgroup \( \Gamma \) of \( PSL(2, \mathbb{R}) \). The 3-manifold \( M = PSL(2, \mathbb{R}) / \Gamma \) can be endowed with a (Haar)-volume preserving flow \( \phi^t \) given by left translations.
by diagonal matrices

\[
\begin{pmatrix}
\exp(t) & 0 \\
0 & \exp(-t)
\end{pmatrix}.
\]

The dynamics of this kind of flow has been widely investigated in particular because of its strong links with number theory (see for instance [84], [68]). We will come back to this key example in Section 3.

Finally, a huge source of examples of volume preserving vector fields comes from the suspension procedure: any area preserving diffeomorphism \( f \) of a surface \( S \) yields a volume preserving vector field on the 3-manifold obtained by gluing the two boundary components of \( S \times [0, 1] \) using \( f \). We will discuss these examples in Section 2.

1.2. Knots and periodic orbits. If a vector field in the 3-sphere or in a domain of \( \mathbb{R}^3 \) has a periodic orbit, this defines a knot whose topology can be used to describe the dynamics. Starting from H. Poincaré one century ago, the quest for periodic orbits has been rewarding\(^1\). Here is a sample of results.

As for the existence question, after a long search around Seifert’s conjecture, K. Kuperberg constructed a jewel. There exists a nonsingular real analytic vector field in the 3-sphere with no periodic orbit [59] (see also [44]). Note however that such a vector field is highly nongeneric.

H. Hofer showed that the Reeb vector field of any contact form in the 3-sphere has at least one periodic orbit [50]. H. Hofer, K. Wysocki and E. Zehnder even showed that at least one of these orbits is unknotted [51].

In between these two cases, the volume preserving case seems difficult:

Does there exist a volume preserving real analytic nonsingular vector field in the 3-sphere with no periodic orbit?

G. Kuperberg constructed examples of \( C^1 \) nonsingular aperiodic volume preserving vector fields in the 3-sphere, but they are not \( C^2 \) [58]! K. Kuperberg’s examples are analytic, but not volume preserving!

\(^1\)“Elles se sont montrées la seule brèche par où nous puissions pénétrer une place jusqu’ici réputée inabordable” (H. Poincaré).
In the opposite direction, there are vector fields in the 3-sphere with plenty of periodic orbits. R. Ghrist constructed another jewel: an explicit real analytic vector field in the 3-sphere whose periodic orbits represent all (isotopy classes of) knots and links! [41]. More recently, J. Etnyre and R. Ghrist even constructed an analytic contact form whose Reeb vector field has the same property [28].

Some vector fields have many periodic orbits representing many knots, but not all. J. Birman and R. Williams pioneered the subject and studied in great detail the case of the Lorenz equation. The main tool is Birman–Williams’ template theory. In Figure 3 (extracted from the original paper [17]), one sees a template: an embedding of a branched surface \(\Sigma\) in \(\mathbb{R}^3\), equipped with a semi-flow \((\psi^t)_{t \geq 0}\). The inverse limit \(\hat{\Sigma}\) of this semi-flow is the space of full orbits, i.e. curves \(c: \mathbb{R} \to \Sigma\) such that \(\psi^t(c(s)) = c(s + t)\) for all \(s \in \mathbb{R}\) and \(t \geq 0\). This is a compact space equipped with a flow \((\hat{\psi}^t)_{t \in \mathbb{R}}\), and an equivariant projection \(\pi: \hat{\Sigma} \to \Sigma\) (i.e. \(\pi \circ \hat{\psi}^t = \psi^t \circ \pi\)). One can embed the abstract space \(\hat{\Sigma}\) in a small neighborhood of \(\Sigma\) in \(\mathbb{R}^3\) in such a way that \(\pi^{-1}(x)\) lies in a small neighborhood of \(x\) in \(\mathbb{R}^3\) and that \(\hat{\psi}^t\) is induced by some smooth vector field in \(\mathbb{R}^3\) preserving \(\hat{\Sigma}\). Any orbit of \(\hat{\psi}^t\) stays close to a full orbit of the original semi-flow \(\psi^t\). This is the geometric Lorenz attractor which has been shown recently to be conjugate to the original Lorenz attractor by W. Tucker [86].

![Figure 3. The Lorenz template.](image)

![Figure 4. The Ghrist template.](image)

In their seminal paper [17], J. Birman and R. Williams were able to reduce the topological study of the knots and links which are present in the Lorenz vector field to a combinatorial study on the template. For instance, all Lorenz knots are prime [91], are fibered knots, and have non negative signature. Hence, Lorenz knots are numerous, but very peculiar. See also [34], [52].

Amazingly, Ghrist’s original example of a vector field exhibiting all knots and links in the 3-sphere is “almost” the same as the Lorenz template (Figure 4)! See the beautiful book [42] for more information.

### 1.3. Asymptotic cycles.

Consider a vector field \(v\) on a compact manifold \(M\), possibly with boundary, preserving some probability measure \(\mu\), and generating a flow \(\phi^t\). Although there might be no periodic orbit, \(\mu\)-almost every point \(x\) is recurrent (Poincaré’s recurrence theorem): there is a sequence \(t_n \to \infty\) such that \(\phi^{t_n}(x)\) converges to \(x\); the long arc of trajectory from \(x\) to \(\phi^{t_n}(x)\) is therefore “almost closed”. Choose some auxiliary generic Riemannian metric on \(M\) and, for any point \(x\) and time \(T\), consider the
closed loop \( k(T, x) \) obtained by concatenation of the arc of trajectory from \( x \) to \( \phi^T(x) \) and some shortest geodesic from \( \phi^T(x) \) to \( x \). Denote by \( [k(T, x)] \in H_1(M, \mathbb{R}) \) the homology class of this loop. In the late 1950s, S. Schwartzman observed (in essence) that the limit \( S(\phi; x) = \lim_{T \to \infty} [k(T, x)]/T \) exists in the first homology group \( H_1(M, \mathbb{R}) \) for \( \mu \)-almost every point \( x \), and that this limit is independent of the auxiliary metric used to close the arcs [82]. The average value \( S(\phi) = \int_M S(\phi; x) d\mu \) is the Schwartzman asymptotic cycle of the flow. Proofs are variations around Birkhoff’s ergodic theorem.

As in the classical ergodic theorem, the actual value of \( S(\phi) \) can be computed as a space average. For each point \( x \), consider the trajectory \( \gamma_x \) from \( x \) to \( \phi^1(x) \) as a de Rham 1-current on \( M \), whose boundary is the difference between a Dirac mass at \( \phi^1(x) \) and a Dirac mass at \( x \). The integral \( \int_M \gamma_x d\mu(x) \) is a 1-cycle since the integral of boundaries vanishes (thanks to the invariance of \( \mu \)). The homology class of this Schwartzman cycle is indeed equal to the above limit \( S(\phi) \).

In other words, a measure preserving flow defines a canonical homology class which can be considered as an “infinitely long knot”. Schwartzman’s point of view has been greatly generalized by D. Sullivan and W. Thurston among others [85].

### 1.4. Helicity

A typical application of this kind of ideas has been carried out by V. Arnold [4]. Suppose for simplicity that \( M \) is the 3-sphere, and that the measure of periodic orbits is zero. Consider two distinct points \( x_1, x_2 \) in \( M \), and two times \( T_1, T_2 > 0 \). The two closed loops \( k(T_1, x_1) \) and \( k(T_2, x_2) \) are disjoint for almost every choice of \( x_1, x_2, T_1, T_2 \) (at least if the metric is generic), and one can consider the asymptotic behavior of their linking number \( \text{link}(k(T_1, x_1), k(T_2, x_2)) \) as \( T_1 \) and \( T_2 \) tend to infinity. Again, as a consequence of Birkhoff’s ergodic theorem, V. Arnold proved that for \( \mu \)-almost every choice of \( x_1, x_2 \), the limit

\[
\text{link}(x_1, x_2) = \lim_{T_1, T_2 \to \infty} \frac{1}{T_1T_2} \text{link}(k(T_1, x_1), k(T_2, x_2))
\]

exists (see also [23], [36], [88]).

*If \( \mu \) is a volume form, V. Arnold identified the integral*

\[
\iint_{M \times M} \text{link}(x_1, x_2) d\mu(x_1) d\mu(x_2)
\]

*that he called the asymptotic Hopf invariant as the helicity, which had been introduced previously by J.-J. Moreau [67] and K. Moffatt [62], [63], [64], [65], [66] and that we now recall. Since \( \phi^t \) preserves a volume form \( \mu \), the inner product \( i_{\alpha} \mu \) is a closed 2-form, hence can be written \( da \) for some 1-form \( \alpha \). The helicity \( \text{Hel}(v) \) is equal to the integral of \( \alpha \wedge da \) over \( M \) (which is easily seen to be independent of the choice of the primitive \( \alpha \)). Note the analogy with the usual definition of Hopf’s invariant for maps from the 3-sphere to the 2-sphere. See also [87] for an interesting definition of helicity in the spirit of Witten’s approach to Jones’ polynomial.*
The helicity \( \text{Hel}(v) \) defines a quadratic form on the Lie algebra of divergence free vector fields, which is invariant under the adjoint action of smooth volume preserving diffeomorphisms. V. Arnold suggests that \( \text{Hel}(v) \) is some “Killing form” for this Lie algebra.

The main open question concerning helicity has been raised by V. Arnold [4]:

*Suppose two smooth volume preserving flows are conjugate by some volume preserving homeomorphism (which is orientation preserving). Does it follow that the two flows have the same helicity?*

The qualitative description of helicity as a limit of linking numbers suggests a positive answer, but one should be cautious that a homeomorphism might entangle the small geodesic arcs that were used to close the trajectories. However, we will see in Section 2 that helicity is indeed a topological invariant for flows with a cross section.

Similarly, V. Arnold asked for a *definition of helicity for volume preserving topological flows*: this problem seems to be wide open.

### 1.5. Digression: the Gordian space.

For almost every point \( x \), the curve \( k(T, x) \) is a knot, *i.e.* has no double point. However, since we are using some auxiliary metric to close the trajectory arc, this knot does depend on the metric. The idea behind the previous constructions is that these knots are “approximately well defined” when \( T \) tends to infinity. This suggests looking at the *space of knots, as a rough metric space, à la Gromov*.

Denote by \( \mathcal{K} \) the (countable) set of (isotopy classes of) knots in 3-space. There is a natural *Gordian distance* \( d_{\text{Gordian}} \) on \( \mathcal{K} \) that we now define. Given two knots \( k_0, k_1 : S^1 \hookrightarrow \mathbb{R}^3 \), one considers homotopies \( (k_t)_{t \in [0, 1]} : S^1 \hookrightarrow \mathbb{R}^3 \) which connect the two knots and are such that for each \( t \in [0, 1] \), the curve \( k_t \) is an immersion with at most one double point, this double point being generic (the two local arcs that intersect have distinct tangents at the intersection). Denote by \( D((k_t)_{t \in [0, 1]}) \) the total number of double points of this family of curves. The Gordian distance between the two knots \( k_0 \) and \( k_1 \) is the minimum of \( D((k_t)_{t \in [0, 1]}) \) for all such homotopies connecting the knots.

The global geometry of this (discrete) metric space is quite intriguing and probably very intricate. Note for instance that this space is not locally finite (an infinite number of knots can be made trivial by allowing one crossing). We propose two kinds of “dual” questions.

One could try to prove (or disprove) that a given metric space \( (E, d) \) can be embedded quasi-isometrically in \( (\mathcal{K}, d_{\text{Gordian}}) \). Recall that a map \( u : E \rightarrow \mathcal{K} \) is a quasi-isometric embedding if there are constants \( C, C' > 0 \) such that

\[
C^{-1}d(x, y) - C' \leq d_{\text{Gordian}}(u(x), u(y)) \leq Cd(x, y) + C'
\]

for all \( x, y \). For instance, we proved in [39] that *every Euclidean space can be embedded quasi-isometrically in \( (\mathcal{K}, d_{\text{Gordian}}) \) and J. Marché showed that a countable
tree such that every vertex has countable valency can also be quasi-isometrically embedded [61].

Can one embed quasi-isometrically the Poincaré disk (or some higher rank symmetric space) in the Gordian space?

In a second approach, one could try to find maps $I : (\mathcal{K}, d_{\text{Gordian}}) \rightarrow (E, d)$ which are quasi-Lipschitz: $d(I(x), I(y)) \leq C d_{\text{Gordian}}(x, y) + C'$ for some suitable metric space $(E, d)$. Any such invariant $I$ would be a candidate for an adaptation to vector fields since $I(k(T, x))$ would not be very sensitive to the choice of the auxiliary Riemannian metric, and the ambiguity could disappear in the asymptotic behavior of $I(k(T, x))$ as $T$ tends to infinity. Very few examples of such invariants $I$ seem to be known. The most trivial one is of course the unknotting number, Gordian distance to the unknot, but this invariant is hard to compute. Equally hard to compute is the genus, i.e. the smallest genus of a Seifert surface. A very interesting (and easy to compute) classical invariant is the signature of knots $\text{sign} : \mathcal{K} \rightarrow \mathbb{Z}$ which is 2-Lipschitz for elementary reasons, as well as its twisted versions $\text{sign}_\omega$, associated to complex numbers of modulus 1 (see [37], [38], [39], [53]).

In [37], we consider a measure preserving vector field $v$ in a bounded domain $M$ of $\mathbb{R}^3$, and we prove that the limit $\text{sign}(v; x) = \lim_{T \rightarrow \infty} \text{sign}(k(T, x))/T^2$ exists for almost every point $x$. Its average $\langle \text{sign}(v; x) \rangle_M = \int_M \text{sign}(v; x) d\mu(x)$ is the signature of the vector field. When $v$ is ergodic with respect to the invariant measure, this signature coincides (surprisingly?) with (one half of) the helicity.

Some other “new” invariants have this Lipschitz property, like the $\tau$ invariant of P. Ozsváth and Z. Szabó, and the $s$ invariant of J. Rasmussen. Do they lead to new dynamical invariants for flows?

In a similar vein, it would be interesting to get some information on the rough geometry of the space of (homeomorphism types of) closed 3-manifolds where the distance between two manifolds is defined as the minimum number of Morse surgeries which are necessary to transform one into the other.

2. Diffeomorphisms of surfaces

Braids are useful to study knots and links mainly because they form a group. In the same way, surface diffeomorphisms are useful to study flows, and also form a group, so that we can use algebraic tools.

If $f$ is a diffeomorphism of a surface $S$, its suspension is obtained by identifying $(x, 0)$ and $(f(x), 1)$ in the cylinder $S \times [0, 1]$. The corresponding manifold $S_f$ is equipped with a flow and a cross section on which the first return map is precisely $f$. If $f$ preserves a measure or an area form, the suspension preserves a natural measure or volume form. In this section, we describe many invariants measuring some kind of twisting in surface diffeomorphisms.
2.1. The Calabi homomorphism. Denote by $G = \text{Diff}(\overline{D}, \partial \overline{D}, \text{area})$ the group of area preserving diffeomorphisms (say of class $C^\infty$) of the closed disk, which are the identity near the boundary. E. Calabi defined a homomorphism

$$C : \text{Diff}(\overline{D}, \partial \overline{D}, \text{area}) \to \mathbb{R}$$

in the following way [19]. Choose a primitive $\alpha$ of the area form in the disk. For each element $f$ of $G$, the form $f^*\alpha - \alpha$ is closed and is therefore the differential $dH$ of a unique function $H$ on the disk which is zero near the boundary. Then $C(f)$ is defined as the integral of $H$.

There is an intuitive description of Calabi’s homomorphism which is due to A. Fathi (unpublished), expressing it as an “average amount of rotation”. The group $G$ is contractible. Choose some isotopy $(f_t)_{t \in [0,1]}$ connecting $f_0 = \text{id}$ and $f_1 = f$. If $x_1, x_2$ are distinct points in the disk, the argument of the nonzero vector $f_t(x_1) - f_t(x_2)$ in $\mathbb{R}^2 \setminus \{(0,0)\}$ rotates by some angle $\text{Angle}(f; x_1, x_2)$ when $t$ goes from 0 to 1 (as a unit for angles, we use the full turn). It is easy to see that this definition is independent of the chosen isotopy. It turns out that

$$C(f) = \iint_{\overline{D} \times \overline{D}} \text{Angle}(f; x_1, x_2) \, dx_1 \, dx_2.$$ 

This interpretation enables a proof of topological invariance for Calabi’s invariant [36]:

*If $f$ and $g$ are two elements of $G$ which are conjugate by some area preserving homeomorphism $h$ of the disk, which is the identity near the boundary, then $C(f) = C(g)$.*

Indeed, even though $h$ is not assumed to be smooth, one can define the number $\text{Angle}(h; x_1, x_2)$, and it is obvious that

$$\text{Angle}(f; x_1, x_2) - \text{Angle}(g; h(x_1), h(x_2)) = \text{Angle}(h; x_1, x_2) - \text{Angle}(h; f(x_1), f(x_2)).$$

Note that $\text{Angle}(h; -, -)$ is a continuous function on the complement of the diagonal in $\overline{D} \times \overline{D}$, and could be nonintegrable if $h$ is not smooth (there could be an unbounded local twist). However, the left hand side of the previous equality is bounded since $f$ and $g$ are assumed to be smooth. As for the right hand side, it is easy to see that its integral, which is defined, has to vanish (for instance approximating $\text{Angle}(h; -, -)$ by a sequence of bounded functions). Hence $C(f) = C(g)$.

Observe that Calabi’s definition extends to more general symplectic manifolds on which the symplectic form is exact. However, no analogous interpretation as an average rotation is known.

The suspension of a diffeomorphism $f$ in $G$ defines a flow $\hat{f}$ on a solid torus $\overline{D} \times S^1$. If one embeds this solid torus in $\mathbb{R}^3$ in a standard way, one can compute the helicity of the suspended flow. In [36], we proved that this helicity is equal to (an explicit multiple of) Calabi’s invariant of $f$. (One has to be slightly careful with...
definitions in nonsimply connected manifolds, see [36]). This follows rather easily from Fathi’s interpretation of Calabi’s invariant and Arnold’s interpretation of helicity.

As a consequence of the topological invariance of Calabi’s number, we get the topological invariance of helicity for flows which are suspensions of area preserving diffeomorphisms of the disk. This is a positive answer to a special case of V. Arnold’s question mentioned above.

2.2. Some algebraic properties of diffeomorphism groups. The kernel of Calabi’s homomorphism $C$ is a simple group [8], [9]. However, the following fundamental question remains open:

Is the group $\text{Homeo}(\bar{D}, \partial \bar{D}, \text{area})$ of area preserving homeomorphisms of the disk which are the identity near the boundary a simple group?

One could try to extend Calabi’s homomorphism to this group of homeomorphisms, but the obvious idea of using the integral of Angle$(h; -, -)$ does not work! If one assumes some rather low regularity for the homeomorphisms, one can nevertheless use this idea, as in the quasi-conformal case [49].

Consider now a closed surface $S$, equipped with some area form $\omega$ (say of total area 1), and let $Diff_0(S, \omega)$ denote the identity component of the group of smooth (say of class $C^\infty$) diffeomorphisms preserving $\omega$. The question of the simplicity of these groups has been settled in the early 1980s (see [7], [9]).

• The group $Diff_0(S^2, \text{area})$ of area (and orientation) preserving diffeomorphisms of the 2-sphere is a simple group. As above, the question of the simplicity of the group of area preserving homeomorphisms of the sphere is open.

• If $S$ is a compact oriented surface of genus at least 2, there is a flux homomorphism $F : Diff_0(S, \text{area}) \to H_1(S, \mathbb{R}) \cong \mathbb{R}^{2g}$ whose kernel is simple, as proved by A. Banyaga. The definition of $F$, due to E. Calabi, is in the spirit of Schwartzman [19]. Let $f \in Diff_0(S, \text{area})$, and choose some isotopy $(f_t)_{t \in [0, 1]}$ connecting the identity to $f$. For each point $x \in S$, the curve $\gamma_x : t \in [0, 1] \mapsto f_t(x) \in S$ can be considered as a 1-current, and the integral $\int_S \gamma_*, d\text{area}(x)$ is a 1-cycle whose homology class $F(f)$ is independent of the choice of the isotopy (this follows from the contractibility of $Diff_0(S, \text{area})$). The kernel of $F$ consists of Hamiltonian diffeomorphisms of $S$.

• $Diff_0(T^2, \text{area})$ is not contractible, but retracts to the subgroup of translations, isomorphic to $T^2$. It follows that the flux is well defined on the universal cover or, better, is defined as a homomorphism $F : Diff_0(T^2, \text{area}) \to H_1(T^2, \mathbb{R})/H_1(T^2, \mathbb{Z}) \cong \mathbb{R}^2/\mathbb{Z}^2$. Again, the kernel of the flux is the simple group of Hamiltonian diffeomorphisms of the torus.

Note that these flux homomorphisms can easily be extended to the groups of area preserving homeomorphisms which are homotopic to the identity. In particular, these fluxes are invariant under topological area preserving conjugacy.

2.3. Dynamical quasi-morphisms. A map $F$ from a group $G$ to $\mathbb{R}$ is a quasi-morphism if $|F(g_1g_2) - F(g_1) - F(g_2)|$ is uniformly bounded (see for instance [57]). One says that $F$ is homogeneous if $F(g^n) = nF(g)$ for every $n \in \mathbb{Z}$ and $g \in G$. For
every quasi-morphism, the limit \( \bar{F}(g) = \lim_{n \to \infty} F(g^n)/n \) exists, and this homogenization defines a quasi-morphism such that \( |\bar{F} - F| \) is bounded.

Some quasi-morphisms have a dynamical flavor. Let \( \text{Homeo}(\mathbb{S}^1) \) be the universal cover of the group of orientation preserving homeomorphisms of the circle, seen as the group of homeomorphisms of \( \mathbb{R} \) commuting with integral translations. The map \( f \in \text{Homeo}(\mathbb{S}^1) \mapsto f(0) \in \mathbb{R} \) is a quasi-morphism whose homogenization is precisely Poincaré’s rotation number (see for instance [46]).

Given a group \( G \), the existence of quasi-morphisms which are nontrivial (i.e., not at a bounded distance from a homomorphism) is related to the second bounded cohomology group of \( G \) (see [47]). In turn, this is related to the commutator length. If an element \( g \) belongs to the first derived group \([G, G]\), it can be written, by definition, as a product of commutators. Let us denote by \( \text{comm}(g) \) the smallest length of such a product, and set \( \text{comm}(g) = \lim_{n \to \infty} \text{comm}(g^n)/n \). It turns out that nontrivial quasi-morphisms exist if and only if \( \text{comm} \) does not vanish identically on \([G, G]\) [12]. For instance, if \( \Gamma \) is a nonelementary Gromov hyperbolic group, the space of homogeneous quasi-morphisms is infinite dimensional [27]. In the opposite direction, if \( \Gamma \) is a uniform lattice in a simple Lie group of real rank at least 2, then every homogeneous quasi-morphism is trivial: a strong improvement of the now classical vanishing of the first Betti number of such lattices [18].

Since we know all homomorphisms from \( \text{Diff}_0(S, \text{area}) \) to \( \mathbb{R} \) (and they are not so numerous), it is tempting to try to understand nontrivial quasi-morphisms, in the spirit of Poincaré’s rotation number, as an attempt to measure some amount of “twisting”, or “rotation”, or “braiding”, contained in some area preserving diffeomorphism. In the next subsections, we will sketch several constructions showing that:

For every closed oriented surface \( S \), the space of homogeneous quasi-morphisms from \( \text{Diff}_0(S, \text{area}) \) to \( \mathbb{R} \) is infinite dimensional [38].

Hopefully, such invariants could be numerous enough to provide a precise description of the topological dynamics, as in the case of circle homeomorphisms (see for instance [46]). As a motivation, let us recall a (generalization of a) conjecture of R. Zimmer which attracted a lot of attention [93]:

Suppose that a lattice in a simple Lie group of real rank \( r \) acts faithfully by homeomorphisms on some compact manifold \( M \) of dimension \( d \). Does it follow that \( d \geq r \)?

Some very special cases of this conjecture are known to be true:

• In dimension \( d = 1 \), the conjecture is settled for smooth actions of general lattices [18], [29], [45], and even for groups with Kazhdan’s property T [69]. It is open for topological actions of general lattices. It has been proved for topological actions for some specific lattices (typically lattices commensurable to \( \text{SL}(n, \mathbb{Z}) \)) [60], [92].

• In dimension \( 2 \), the conjecture is open in full generality. However, it has been proved by very different techniques for specific lattices (for instance lattices commensurable to \( \text{SL}(n, \mathbb{Z}) \) with \( n \geq 3 \)) under some additional assumptions: in [32], [33]
for smooth area preserving actions; in [76] for smooth area preserving actions on a closed oriented surface of genus at least 1; in [43] for real analytic actions on closed surfaces different from the torus, and in [79] for the torus case.

- In higher dimension, not much is known, unless one adds strong conditions on the action, like assuming that the action preserves a connection [31], or is holomorphic on a Kaehler manifold [20], or for specific lattices acting analytically on 4-manifolds with non zero Euler–Poincaré characteristic [30] etc.

One of the first nontrivial open cases of this conjecture is the following.

Can a uniform lattice in a simple Lie group of real rank at least 2 act faithfully on a compact surface by area preserving diffeomorphisms?

Suppose a group $\Gamma$ embeds in $\text{Diff}_0(S, \text{area})$, and let $F$ be a quasi-morphism on $\text{Diff}_0(S, \text{area})$. This produces a quasi-morphism on $\Gamma$. For instance, if $\Gamma$ is a uniform lattice in $\text{SL}(n, \mathbb{R})$ ($n \geq 3$), we already mentioned that such a quasi-morphism has to be trivial. If one could construct a wealth of quasi-morphisms on $\text{Diff}_0(S, \text{area})$ with strong dynamical content, this vanishing result could lead to strong dynamical restrictions on possible actions of lattices on surfaces, by area preserving diffeomorphisms.

The previous comments on quasi-morphisms imply that it might be relevant to search for quasi-morphisms on $\text{Diff}_0(S, \text{area})$. The next few sections will survey some recent progress in this direction.

2.4. Ruelle’s rotation numbers. The following construction is due to D. Ruelle (in a higher dimensional symplectic situation [80]), and was placed in the setting of bounded cohomology in [11].

Let $f$ be an element of $\text{Diff}(\mathbb{D}, \partial \mathbb{D}, \text{area})$, and choose an isotopy $(f_t)_{t \in [0,1]}$ between $f_0 = \text{id}$ and $f_1 = f$. For each point $x$ in the disk, consider the differential $df_t(x)$. Using the natural trivialization of the tangent bundle of the disk, we can consider this differential as a $2 \times 2$ matrix, element of $\text{SL}(2, \mathbb{R})$. The first column $v_t(x)$ of $df_t(x)$ is a non zero vector in $\mathbb{R}^2$. Denote by $\text{Angle}(f; x) \in \mathbb{R}$ the variation of the angle of this curve $v_t(x)$ of nonzero vectors when $t$ runs from 0 to 1. This number does not depend on the choice of the isotopy $f_t$ since $\text{Diff}(\mathbb{D}, \partial \mathbb{D}, \text{area})$ is contractible. Let us define

$$r(f) = \int_\mathbb{D} \text{Angle}(f; x) \, d\text{area}(x).$$

Consider now two elements $f$ and $g$ of $\text{Diff}(\mathbb{D}, \partial \mathbb{D}, \text{area})$, and choose two isotopies $f_t$ and $g_t$ as above. Using the concatenation of these isotopies, one sees that

$$|\text{Angle}(fg; x) - \text{Angle}(g; x) - \text{Angle}(f; g(x))| < 1/2.$$

It follows that $r$ is a quasi-morphism. After homogenization, we get Ruelle’s homogeneous quasi-morphism

$$\mathcal{R}_\mathbb{D}(f) = \lim_{n \to +\infty} \frac{1}{n} r(f^n).$$
It is not difficult to check on simple examples that \( \mathcal{R}_{\mathbb{D}} \) is indeed nontrivial. For instance, let \( H : \overline{\mathbb{D}} \to \mathbb{R} \) be a (Hamiltonian) function on the disk which vanishes near the boundary, and suppose for simplicity that the critical points of \( H \) consist of a finite number of nondegenerate critical points \( x_i \), together with some annular neighborhood of the boundary (on which \( H = 0 \)). Denote by \( X_H \) the symplectic gradient of \( H \), and by \( H^{(1)} \) the time 1 diffeomorphism defined by \( X_H \). Then

\[
\mathcal{R}_{\mathbb{D}}(H^{(1)}) = \sum_i \varepsilon_i H(x_i),
\]

where \( \varepsilon_i = +1 \) if \( x_i \) is a local extremum and \(-1\) if it is a saddle point (up to some irrelevant multiplicative constant, compare [36]).

This construction can readily be extended to \( \text{Diff}_0(\mathbb{T}^2, \text{area}) \). Indeed, since the tangent bundle of \( \mathbb{T}^2 \) is trivial, the differential \( d_f \) can still be considered as a matrix. One has to be careful since the isotopy is not unique up to homotopy, but any loop in \( \text{Diff}_0(\mathbb{T}^2, \text{area}) \) is homotopic to a loop in the translation subgroup and this guarantees that \( \text{Angle}(g; x) \) is indeed well defined. Hence, we get a Ruelle quasi-morphism \( \mathcal{R}_{\mathbb{T}^2} \) on \( \text{Diff}_0(\mathbb{T}^2, \text{area}) \).

The case of closed surfaces of higher genus \( S \) is more subtle since the tangent bundle is nontrivial! However, one can proceed in the following way [38]. Choose a hyperbolic Riemannian metric on \( S \), and let \( f \in \text{Diff}_0(S, \text{area}) \). Choose as usual some isotopy \( f_t \) from the identity to \( f \), a point \( x \) in \( S \), and a unit vector \( u \) tangent at \( x \). Consider the curve \( d_f(u) \) in the tangent bundle of \( S \), and lift it as a curve \( \tilde{d}_f(u) \) to the tangent bundle of the Poincaré disk. Every nonzero tangent vector in the Poincaré disk defines a geodesic ray which converges to some point at infinity, so that one gets a curve in the circle at infinity. We denote by \( \text{Angle}(f; u, x) \) the number of full turns made by this curve at infinity. This is independent of the choices of the hyperbolic metric, of the isotopy \( f_t \), and of the lift. Moreover, \( \text{Angle}(f; u, x) \) changes by at most 1 when one changes \( u \), keeping \( f \) and \( x \) fixed, so that one can now define \( \text{Angle}(f; x) \) to be the minimum value of \( \text{Angle}(f; u, x) \). As before, we now define \( r(f) = \int_{\mathbb{S}^2} \text{Angle}(f; x) \, d\text{area}(x) \), and finally \( \mathcal{R}_S(f) \) by homogenization of \( r \).

The definitions of Ruelle’s quasi-morphisms on the disk and the torus use the triviality of the tangent bundle of these surfaces, and the definition on higher genus surfaces uses some kind of “quasi” triviality of the tangent bundle given by the circle at infinity. We now give a definition in the case of the sphere [38]. Instead of using the action on tangent vectors, we use the action on pairs of tangent vectors. Denote by \( T_2(\mathbb{S}^2) \) the space of pairs of nonzero tangent vectors \((\delta x_1, \delta x_2)\) at distinct points \( x_1, x_2 \) of the sphere. Observe that the fundamental group of \( T_2(\mathbb{S}^2) \) is isomorphic to \( \mathbb{Z} \times (\mathbb{Z}/2\mathbb{Z}) \), so that it does make sense to say that a curve in \( T_2(\mathbb{S}^2) \) turns. In order to give a quantitative statement, we identify the sphere with the Riemann sphere \( \mathbb{C} \cup \{\infty\} \). The complex differential form

\[
\theta = \frac{dx_1 dx_2}{(x_1 - x_2)^2}
\]
can be seen as a holomorphic form on the space of pairs of distinct points on \( \mathbb{CP}^1 \), or as a function on \( T_2(S^2) \). Note that this form is invariant under the projective action of \( \text{PGL}(2, \mathbb{C}) \), and in particular \( \theta \) is well defined and nonsingular when \( x_1 \) or \( x_2 \) is at infinity. As for the geometrical meaning of \( \theta \), observe that \( \theta \) is the cross ratio of the four points \( "x_1, x_1 + \delta x_1, x_1, x_2 + \delta x_2" \). Given a curve \( c : [0, 1] \to T_2(S^2) \), we define \( \text{Angle}(c) \in \mathbb{R} \) as the variation of the argument of the complex number \( \theta(c) \). This is invariant under homotopies fixing the endpoints.

We can proceed as in the case of the disk. Start with a diffeomorphism \( f \) in \( \text{Diff}_0(S^2, \text{area}) \). Choose an isotopy \((f_t)_{t\in[0,1]} \) and an element \( v = (x_1, \delta x_1; x_2, \delta x_2) \) of \( T_2(S^2) \). We can consider the image \( v_t \) of \( v \) by the differential of \( f_t \). This gives a curve in \( T_2(S^2) \) and therefore defines some \( \text{Angle}(f; x_1, x_2, \delta x_1, \delta x_2) \). Fixing \( x_1 \) and \( x_2 \) and changing the tangent vectors \( \delta x_1, \delta x_2 \) changes this rotation angle by at most 2 full turns. We can therefore define \( \text{Angle}(f; x_1, x_2) \) as the minimum of \( \text{Angle}(f; x_1, \delta x_1; x_2, \delta x_2) \) over all choices of \( \delta x_1, \delta x_2 \). We now define \( r(f) \) as the double integral of \( \text{Angle}(f; x_1, x_2) \) and \( \mathcal{R}_{S^2} \) as the homogenization of \( r \). Clearly this defines a homogeneous quasi-morphism on \( \text{Diff}_0(S^2, \text{area}) \) that we call Ruelle’s quasi-morphism on the sphere.

All these Ruelle quasi-morphisms turn out to be topological invariants:

Two elements of \( \text{Diff}_0(S, \text{area}) \) which are conjugate by some area preserving homeomorphism, respecting orientation, have the same Ruelle invariants [38]. Can one extend their definitions to homeomorphisms?

Note that one can also define a Ruelle invariant for a nonsingular flow on a 3-manifold with trivialized normal bundle (for example on the 3-sphere). One looks at the rotation action of the differential of the flow on a plane field containing the flow. Similar methods imply its topological invariance.

### 2.5. Quasi-fluxes, quasi-translation numbers

Let \( S \) be a closed surface equipped with a metric with curvature \(-1\). Let \( f \in \text{Diff}_0(S, \text{area}) \), and choose some isotopy \( f_t \) from the identity to \( f \). For each point \( x \) in \( S \), consider the unique geodesic arc \( \gamma(f; x) \) connecting \( x \) and \( f(x) \) which is homotopic to the curve \( t \mapsto f_t(x) \). If one considers \( \gamma(f; x) \) as a 1-current, the integral \( t(f) = \int_S \gamma(f; x) \, d\text{area}(x) \) is a 1-cycle, and the homogenization \( \mathcal{T}(f) = \lim t(f^n)/n \) exists in the space of 1-cycles with the weak topology. Indeed, let \( f, g \) denote two elements of \( \text{Diff}_0(S, \text{area}) \) and, for \( x \) in \( S \), denote by \( \Delta(x, g(x), fg(x)) \) the (immersed) geodesic triangle whose boundary consists of \( \gamma(g; x), \gamma(f; g(x)) \) and \( \gamma(g^{-1}f^{-1}; fg(x)) \). For any 1-form \( \alpha \) on \( S \), one can compute

\[
(t(fg) - t(f) - t(g))(\alpha) = \int_S \left( \int_{\Delta(x,g(x),fg(x))} d\alpha \right) d\text{area}(x)
\]

which is bounded by \( \pi \times \sup \text{norm of } d\alpha \), since the areas of triangles in the Poincaré disk are bounded by \( \pi \). Hence, for every 1-form \( \alpha \), the evaluation \( t(f)(\alpha) \) is a quasi-morphism, so that the homogenization is indeed well
defined. In other words, we defined a quasi-flux with values in the space $Z_1(S)$ of 1-cycles:

$$\mathcal{T}_S : \text{Diff}_0(S, \text{area}) \rightarrow Z_1(S).$$

Of course, the homology class of $\mathcal{T}_S$ reduces to Calabi’s flux homomorphism. In [38], we proved that the image of $\mathcal{T}_S$ does not lie in a finite dimensional subspace. It is not difficult, using methods from [10], to show that the image of $\mathcal{T}_S$ actually spans a dense subspace of the space of cycles.

Note that this construction obviously extends to area preserving homeomorphisms.

2.6. Braiding. We have seen that Calabi’s invariant of a diffeomorphism of the disk measures the average rotation on pairs of points. It is natural to look at the action on $n$-tuples of points [40]. Recall that the braid group $B_n$ is the fundamental group of the space $X_n(\overline{D})$ of unordered $n$-tuples of distinct points in a disk. We choose $n$ distinct base points $(x_1^0, \ldots, x_n^0)$ in the disk so that a braid can be visualized as a union of $n$ disjoint arcs in $\overline{D} \times [0, 1]$ transversal to each disk $\overline{D} \times \{\ast\}$ and connecting $(x_1^0, \ldots, x_n^0) \times \{0\}$ to $(x_1^0, \ldots, x_n^0) \times \{1\}$. By closing these arcs in $\mathbb{R}^3$ in a canonical way outside $\overline{D} \times [0, 1] \subset \mathbb{R}^3$, we see that every braid $\beta$ defines a link $\hat{\beta}$ in $\mathbb{R}^3$.

Suppose $f$ is an element of $\text{Diff}(\overline{D}, \partial \overline{D}, \text{area})$, and choose some isotopy $(f_t)_{t \in [0, 1]}$ connecting $f_0 = \text{id}$ to $f_1 = f$. For every $n$-tuple of distinct points $(x_1, \ldots, x_n)$ in the disk, we get a curve $(f_t(x_1), \ldots, f_t(x_n))$ in $X_n(\overline{D})$. Of course, this curve does not define a braid since it is not a loop, but one can easily construct a braid as we did when we closed trajectories of flows by short geodesics. More precisely, for each $i$ we concatenate three curves; the first (resp. third) connects $x_i^0$ to $x_i$ (resp. $f_1(x_i)$ to $x_i^0$) in an affine way, and the second is the curve $f_t(x_i)$. These curves now define a closed loop in the space of $n$-tuples, which is contained in the space of $n$-tuples of distinct points for almost every $(x_1, \ldots, x_n)$. In other words, we get a (pure) braid $\beta(f; x_1, \ldots, x_n)$ in $B_n$. As before this is independent of the choice of the isotopy, and this provides a cocycle, i.e. for $f, g$ in $\text{Diff}(\overline{D}, \partial \overline{D}, \text{area})$ and almost every $n$-tuple, one has

$$\beta(fg; x_1, \ldots, x_n) = \beta(g; x_1, \ldots, x_n)\beta(f; g(x_1), \ldots, g(x_n)).$$

Consider now some quasi-morphism $F : B_n \rightarrow \mathbb{R}$. One can average the value of $F(\beta(f; x_1, \ldots, x_n))$ over the space of $n$-tuples of distinct points if this is integrable. This strategy is valid for the signature quasi-morphism. Indeed, the map which associates to each braid $\beta$ the signature of its closure $\hat{\beta}$ is a quasi-morphism. This follows from the Lipschitz property of the signature in the Gordian space, that we mentioned earlier (see also [39] for a description of the coboundary of this quasi-morphism). As in the case of Calabi’s homomorphism, it is not difficult to check that $\text{sign}(\beta(f; x_1, \ldots, x_n))$ is indeed an integrable function. After integration over the space of $n$-tuples and homogenization, we get for each $n \geq 2$ a quasi-morphism:

$$\text{Sign}_{n, \overline{B}} : \text{Diff}(\overline{D}, \partial \overline{D}, \text{area}) \rightarrow \mathbb{R}.$$
Although it is not defined for homeomorphisms, one can also prove its topological invariance.

One can compute explicitly these invariants on simple examples. For instance, let \( h : [0, 1] \to \mathbb{R} \) be a smooth function, which is equal to 0 in a neighborhood of 1, define a Hamiltonian function \( H \) on the disk by \( H(x) = h(\|x\|^2) \), and consider the associated time 1 diffeomorphism \( H^{(1)} \). Then one has

\[
\text{Sign}_{n, \mathbb{D}}(H^{(1)}) = \int_0^1 h(u)(un^{-2} + 1) \, du
\]

(up to some explicit multiplicative constant \([38]\)). Of course, the case \( n = 2 \) reduces to (a constant multiple of) Calabi’s invariant \((B_2 \simeq \mathbb{Z})\). Note that these numbers determine all moments of \( h \) and therefore the function \( h \) itself. This is a (small) hint that these quasi-morphisms give a good description of the dynamics.

One can proceed in a similar way with diffeomorphisms of the sphere except that we now get a cocycle with values in the pure braid group of the sphere \( P_n(S^2) \) (fundamental group of the space of ordered \( n \)-tuples of distinct points on \( S^2 \)). It is not difficult to express \( P_n(S^2) \) as a central extension of the standard pure braid group \( P_{n-1}(\mathbb{D}) \) (i.e. the pure braid group of the disk):

\[
0 \to \mathbb{Z} \to P_{n-1}(\mathbb{D}) \to P_n(S^2) \to 1.
\]

In this exact sequence, the central \( \mathbb{Z} \) is generated by a “double full turn” in \( \text{SO}(2) \) (which is homotopically trivial in \( \text{SO}(3) \)). The projection from \( P_{n-1}(\mathbb{D}) \) to \( P_n(S^2) \) consists in “adding a strand at infinity”. On \( P_{n-1}(\mathbb{D}) \), we have a nontrivial homomorphism \( \text{lk}_{n-1} \) onto \( \mathbb{Z} \) given by the total linking number of the strands, and a quasi-morphism given by the signature. A suitable linear combination \( \text{sign}_{n-1} - c_n \text{lk}_{n-1} \) descends to a quasi-morphism on \( P_n(S^2) \) that we simply called the signature of a spherical braid in \([38]\). As before, we can use these spherical signatures to define quasi-morphisms \( \text{Sign}_{n, S^2} \) on \( \text{Diff}_0(S^2, \text{area}) \) which are again topological invariants.

If we think of \( S^2 \) as the unit sphere in \( \mathbb{R}^3 \), and we consider some Hamiltonian function only depending on the third coordinate \( z \) through a smooth function \( h : [-1, 1] \to \mathbb{R} \), the invariant of the associated time 1 diffeomorphism \( H^{(1)} \) is given by the following formula (up to some irrelevant multiplicative constant and for \( n \) even):

\[
\text{Sign}_{n, S^2}(H^{(1)}) = \int_{-1}^{+1} ((n-1)u^{n-2} - 1) h(u) \, du.
\]

The first interesting case is \( n = 4 \) and deserves special attention. Let us give some interpretation of \( \text{Sign}_{4, S^2} \) as an “amount of braiding”. Given four distinct points \( z_1, z_2, z_3, z_4 \) of the sphere, seen as the Riemann sphere, their crossratio \( [z_1, z_2, z_3, z_4] = \frac{(z_3-z_1)(z_4-z_2)}{(z_3-z_2)(z_4-z_1)} \) is in \( \overline{\mathbb{C}} \setminus \{0, 1, \infty\} \). The universal cover of a sphere minus three points can be identified with the Poincaré disk \( \mathbb{D} \). More precisely, there is a covering map from \( \mathbb{D} \) onto \( \overline{\mathbb{C}} \setminus \{0, 1, \infty\} \) and the inverse images of points of \( \mathbb{R} \setminus \{0, 1, \infty\} \) define a tesselation of \( \mathbb{D} \) by ideal triangles.
Let $f_t$ be some isotopy of the sphere from $f_0 = \text{id}$ to some area preserving diffeomorphism $f$. Choose four distinct points $z_1, z_2, z_3, z_4$ in the sphere, consider the path $\{f_t(z_1), f_t(z_2), f_t(z_3), f_t(z_4)\}$ in the sphere minus three points, lift it to the disk, and finally consider the geodesic arc connecting the end points of this lift. Each time this geodesic enters and exits one of the ideal triangles, the exit may be the left or the right side of the triangle, as seen from the entrance side. Counting the number of left exits minus the number of right exits, one gets an integer $t(f; z_1, z_2, z_3, z_4)$ that one can integrate on the space of 4-tuples. After homogenization, one produces a quasi-morphism on $\text{Diff}_0(S^2, \text{area})$ which turns out to be (a constant multiple of) $\text{Sign}_{4, S^2}$ [38], [39]. We will meet again this left-right exits count in the last section, in relation with the so-called Rademacher function. See also [13], [14].

2.7. Calabi quasi-morphisms on surfaces: Py’s construction. Consider a closed connected surface $S$ equipped with an area form, and denote by $\text{Ham}(S, \text{area})$ the group of Hamiltonian diffeomorphisms of $S$. Let $D \subset S$ be some open set diffeomorphic to a disk. One can consider the group $\text{Diff}_c(D, \text{area})$ of area preserving diffeomorphisms of $D$ with compact support as a subgroup of $\text{Ham}(S, \text{area})$, extending by the identity outside $D$. Note that $\text{Ham}(S, \text{area})$ is a simple group, but that $\text{Diff}_c(D, \text{area})$ is not simple since it surjects onto $\mathbb{R}$ by Calabi’s homomorphism.

M. Entov and L. Polterovich suggested looking for Calabi quasi-morphisms, i.e., homogeneous quasi-morphisms $F : \text{Ham}(S, \text{area}) \to \mathbb{R}$ which restrict to Calabi’s homomorphisms on subgroups of the form $\text{Diff}_c(D, \text{area})$ if $D$ is “small enough”. They proved the remarkable result that such a Calabi quasi-morphism does exist when $S$ is the sphere (and for many other higher dimensional symplectic manifolds) where “small enough” means “of area less than one half of the area of the sphere”.

P. Py succeeded with the same task when $S$ is of genus at least 2 and $D$ is any disk [77].


We begin with a description of Py’s invariant since it is more elementary and more in the spirit of the previous discussion.

Choose a Riemannian metric with curvature $-1$ on $S$, and denote by $p : T^1 S \to S$ its unit tangent bundle, seen as a principal $\text{SO}(2)$ bundle with a natural connection.
Denote by $\partial/\partial \theta$ the vector field generating the SO(2) action. Note that the map which associates to any vector field $X$ on $S$ its horizontal lift $\tilde{X}$ in $T^1S$ is not a Lie algebra homomorphism since the connection is not flat. However, if $H: S \to \mathbb{R}$ is a Hamiltonian with zero integral, $X_H$ its symplectic gradient, and $\tilde{X_H} = \tilde{X} + H \circ p \cdot \partial/\partial \theta$, the map $H \mapsto \tilde{X_H}$ is a Lie algebra homomorphism from the Poisson algebra to the Lie algebra of vector fields on $T^1S$ commuting with the SO(2) action. Integrating this homomorphism, we get a homomorphism $f \mapsto \tilde{f}$ from Ham($S$, area) (which is simply connected) to the group of diffeomorphisms of $T^1S$ commuting with SO(2). This construction is due to A. Banyaga [7].

Now, consider an element $f$ in Ham($S$, area), written as time 1 of a Hamiltonian isotopy $(f_t)_{t \in [0,1]}$. For each point $x$ in $S$, and each unit vector $v$ tangent at $x$, one gets a curve $\tilde{f_t}(v)$ in $T^1S$ which can be lifted as a curve in the unit tangent bundle of the disk. As we did before, we can now project this curve to the boundary of the Poincaré disk, so that we get a curve in a circle, giving a certain number of full turns, as $t$ goes from 0 to 1. Fixing $f$ and $x$, this integer changes by at most 2 when one changes $v$, so that we can consider its minimum $A(f; x)$. As usually, we can define $\pi(f) = \int_S A(f; x) \, d\text{area}(x)$, and homogenize to produce a homogeneous quasi-morphism $\Pi: \text{Ham}(S, \text{area}) \to \mathbb{R}$. When the support of $f$ lies in a disk $D \subset S$, the invariant $\Pi(f)$ coincides with the value of Calabi’s invariant $C(f|_D)$ of the restriction of $f$ to $D$. In other words, $\Pi$ is indeed a Calabi quasi-morphism.

The computation of this invariant is especially interesting for a diffeomorphism $H^{(1)}$ which is the time 1 of some autonomous Hamiltonian $H: S \to \mathbb{R}$ with zero integral. Denote by $\nu$ the genus of $S$ and assume for simplicity that $H$ is a Morse function with only $2\nu + 2$ critical points $x_1, x_2, \ldots, x_{2\nu+2}$, such that $H(x_1) < H(x_2) < \cdots < H(x_{2\nu+2})$. In this case, it turns out that

$$\Pi(H^{(1)}) = \sum_{i=3}^{2\nu} H(x_i).$$

(up to some irrelevant multiplicative constant). For a general Morse function $H$ with distinct critical values, the invariant $\Pi(H^{(1)})$ is the sum of the values of $H$ on the $2\nu - 2$ saddle points $x_i$ which are such that the fundamental group of the connected component of $H^{-1}(H(x_i))$ containing $x_i$ embeds in the fundamental group of the surface.

2.8. The Entov–Polterovich quasi-morphism. We briefly sketch the construction by M. Entov and L. Polterovich of a Calabi quasi-morphism on the sphere (and on many other symplectic manifolds) using elaborate tools from symplectic topology [25]. We will restrict our description to the 2 dimensional case, and refer to [16], [25], [72] for higher dimensional examples.3

The free loop space of the 2-sphere is not simply connected. Let us denote by $\Lambda$ its universal cover, that one can consider as the space of pairs $(\gamma, w)$ where $\gamma : S^1 \to S^2$ is a loop and $w : \mathbb{D} \to S^2$ is a disk with boundary $\gamma$, where one identifies $(\gamma, w)$ with $(\gamma, w')$ if $w$ and $w'$ are homotopic relative to their boundary.

Fix some time dependent Hamiltonian $H : S^2 \times S^1 \to \mathbb{R}$, normalized in such a way that for each time $t \in S^1$ the integral of $H(\gamma(\cdot), t)$ over the sphere is zero. Denote by $H^{(1)}$ the Hamiltonian diffeomorphism of the sphere which is the time 1 of the isotopy defined by $H$. The action is a functional defined on $\Lambda$ by

$$A_H : (\gamma, w) \in \Lambda \mapsto \int_0^1 H(\gamma(t), t) \, dt - \text{area}(w) \in \mathbb{R}.$$

The critical points of $A_H$ correspond to the fixed points of $H^{(1)}$. The Floer homology is a tool to analyze these critical points. One considers a differential complex freely generated by critical points, whose differential is defined using connecting orbits for the gradient flow of the action functional, which can be interpreted as pseudo-holomorphic cylinders (see [73], [71], [70] for many more “details”). The main point is that the corresponding Floer homology $HF(\Lambda)$ is independent of the choice of the Hamiltonian $H$. In our case, $HF(\Lambda)$ is some simple quantum deformation of the homology of the sphere.

However, the chain complex used to compute the Floer homology does depend on the choice of the Hamiltonian. One defines a spectral invariant for a Hamiltonian $H$: the infimum of the set of $z \in \mathbb{R}$ such that the sub-level $\{A_H < z\}$ contains a Floer cycle representing the fundamental class in $HF(\Lambda)$. It turns out that this infimum only depends on the Hamiltonian diffeomorphism $H^{(1)}$, and defines therefore a map $ep : \text{Ham}(S^2) \to \mathbb{R}$. M. Entov and L. Polterovich prove that $ep$ is a quasi-morphism and define their Calabi quasi-morphism $E_P$ by homogenization. They also prove that the restrictions of $E_P$ to the subgroups $\text{Diff}_c(D, \text{area})$, where $D$ is a disk with area less than one half of the sphere, coincide with Calabi’s homomorphisms. The key point is that such a disk $D$ is displaceable, which means that there is a Hamiltonian diffeomorphism $h$ such that $h(D)$ and $D$ are disjoint.

The computation of the Entov–Polterovich Calabi quasi-morphism on time 1 maps of autonomous Hamiltonians is very interesting. Assume for simplicity that $H$ is a Morse function on $S^2$. It is not difficult to see that there is a unique “median” value $z_H \in \mathbb{R}$ such that the complement of one of connected component of $H^{-1}(z_H)$ is a disjoint union of open disks with areas less 1/2. Then

$$E_P(H^{(1)}) = \int_{S^2} H \, d\text{area} - H(z_H).$$

(The total area of the sphere is normalized to 1). Hence, Entov–Polterovich’s invariant is the difference between the “average” and the “median” values of the Hamiltonian. The uniqueness of such a Calabi quasi-morphism on $\text{Ham}(S^2, \text{area})$ is an open question.
Remarkably, this Entov–Polterovich Calabi quasi-morphism provides a natural example of a quasi-measure on the sphere. A quasi-measure \( \mu \) on a compact space \( K \) is a map \( \mu : C(K) \to \mathbb{R} \) defined on the algebra of continuous functions, which is linear on subalgebras generated by one element, and monotonic (\( f \leq g \) implies \( \mu(f) \leq \mu(g) \)). Such a quasi-measure does not need to be a measure, i.e. does not need to be linear (see [1], [24], [56], [81]). If \( H \) is a Morse function on \( S^2 \), one may set
\[
\mu(H) = H(z_H).
\]
It is not difficult to check that \( \mu \) extends to continuous functions on the sphere as a quasi-measure.

3. An example: geodesics on the modular surface

3.1. The unit tangent bundle. The following is well known:

The quotient \( M = \text{PSL}(2, \mathbb{R}) / \text{PSL}(2, \mathbb{Z}) \) is homeomorphic to the complement of the trefoil knot in the 3-sphere.

An explicit homeomorphism is given by classical modular functions. Observe that \( M \) can be identified with the space of lattices \( \Lambda \subset \mathbb{C} \) such that the area of the quotient torus \( \mathbb{C}/\Lambda \) is 1. For any lattice \( \Lambda \), one defines
\[
g_2(\Lambda) = 60 \sum_{\omega \in \Lambda \setminus \{0\}} \omega^{-4}; \quad g_3(\Lambda) = 140 \sum_{\omega \in \Lambda \setminus \{0\}} \omega^{-6}
\]
(see for instance [2]). Conversely, a pair \((g_2, g_3)\) of \( \mathbb{C}^2 \) such that \( \Delta = g_2^3 - 27g_3^2 \neq 0 \) determines a unique lattice \( \Lambda \). Note that the unit sphere \( S^3 \subset \mathbb{C}^2 \) intersects the algebraic curve \( \{\Delta = 0\} \) along a trefoil knot \( \ell \subset S^3 \). Given \((g_2, g_3)\) in \( S^3 \setminus \ell \), the associated lattice is not necessarily of co-area 1, but has a unique “rescaling” of co-area 1. This provides a homeomorphism from the complement of the trefoil knot to the space \( M \).

We have already mentioned that \( M \) is equipped with a flow \( \phi^t \) which is given by left translations by diagonal matrices
\[
\delta(t) = \begin{pmatrix} \exp(t) & 0 \\ 0 & \exp(-t) \end{pmatrix}.
\]
If one thinks of \( M \) as a space of lattices in \( \mathbb{C} \cong \mathbb{R}^2 \), the action of \( \phi^t \) is simply induced by the action of \( \delta^t \) on \( \mathbb{R}^2 \).

The space \( M \) can also be seen as the unit tangent bundle of the modular orbifold \( \Sigma = \mathbb{D} / \text{PSL}(2, \mathbb{Z}) \). Indeed, the group of positive isometries of the Poincaré disk \( \mathbb{D} \) is isomorphic to \( \text{PSL}(2, \mathbb{R}) \) and acts freely and transitively on the unit tangent bundle of the disk. From this point of view, \( \phi^t \) appears as the geodesic flow of the modular orbifold (rescaled by a factor of 2).
Periodic orbits of this geodesic flow \( \phi^t \) have a long mathematical tradition. Note that an element \( P \in \PSL(2, \mathbb{R}) \) defines an element in \( \PSL(2, \mathbb{R})/\PSL(2, \mathbb{Z}) \) which is fixed by \( \phi^t \) if \( \delta(t)P = \pm PA \) for some \( A \) in \( \PSL(2, \mathbb{Z}) \), which means that \( PA P^{-1} \) is diagonal. One deduces that there is a natural bijection between periodic orbits of \( \phi^t \) and conjugacy classes of hyperbolic elements in \( \PSL(2, \mathbb{Z}) \).

These periodic orbits are also related to indefinite integral quadratic forms in \( \mathbb{Z}^2 \), or to the structure of ideals in real quadratic fields (Gauss, see for instance [22]). Of course, one could also say that periodic orbits correspond to closed geodesics on \( \Sigma = \mathbb{D}/\PSL(2, \mathbb{Z}) \), or to free homotopy classes of closed curves in \( \Sigma \) (with the exception of parabolic and elliptic elements).

Summing up, any hyperbolic matrix \( A \in \PSL(2, \mathbb{Z}) \) defines a periodic orbit of \( \phi^t \), hence a knot \( k_A \) in the complement of the trefoil knot.

In this section, we describe the topology of these knots that we call modular knots.

### 3.2. The Rademacher function

Our first task will be to relate the linking number between \( k_A \) and the trefoil knot \( \ell \) to a classical arithmetical invariant that we now recall.

The Dedekind \( \eta \) function defined for \( \Im \tau > 0 \) by

\[
\eta(\tau) = \exp(i\pi \tau/12) \prod_{n=1}^{\infty} (1 - \exp(2i\pi n \tau))
\]

"is one of the most famous and well-studied in mathematics" [6]. Its 24th power is a modular form of weight 12, which means that

\[
\eta^{24} \left( \frac{a\tau + b}{c\tau + d} \right) = \eta^{24}(\tau)(c\tau + d)^{12}
\]

for every matrix \( A = \pm \left( \begin{smallmatrix} a & b \\ c & d \end{smallmatrix} \right) \) in \( \PSL(2, \mathbb{Z}) \) (see for instance [2]). Since \( \eta \) does not vanish, there is a holomorphic determination of \( \log \eta \) defined on the upper half plane. Taking logarithms on both sides of the previous identity, we get

\[
24(\log \eta) \left( \frac{a\tau + b}{c\tau + d} \right) = 24(\log \eta)(\tau) + 6 \log(- (c\tau + d)^2) + 2i\pi \Re(A)
\]

for some function \( \Re : \PSL(2, \mathbb{Z}) \to \mathbb{Z} \) (the second log in the right hand side is chosen with imaginary part in \( (-\pi, \pi) \)). The numerical determination of \( \Re(A) \) has been a challenge, and turned out to be related to many different topics, in particular number theory, topology, and combinatorics. The inspiring paper by M. Atiyah [6] contains an “omnibus theorem” proving that seven definitions of \( \Re \) are equivalent! In [11], we proposed an approach to understand better these coincidences, based on the more or less obvious fact that \( \Re \) is a quasi-morphism. It is difficult to choose a name for this “ubiquitous” function: Arnold, Atiyah, Brooks, Dedekind, Dupont, Euler, Guichardet, Hirzebruch, Kashiwara, Leray, Lion, Maslov, Meyer, Rademacher, Souriau, Vergne, Wigner? For simplicity, we will call it the Rademacher function [78].
3.3. Linking with the trefoil. We now state a result relating modular knots with the Rademacher function.

For every hyperbolic element $A$ in $\text{PSL}(2, \mathbb{Z})$, the linking number between the knot $k_A$ and the trefoil knot $\ell$ is equal to $\Re(A)$, where $\Re$ is the Rademacher function.

We will give three proofs, connecting link$(k_A, \ell)$ to three different aspects of this ubiquitous function $\Re$ (thus providing new proofs of the identifications of these various versions of $\Re$). The third proof will give an extra bonus, and will allow a precise description of the topology of modular knots.

Our first proof relies on the definition of $\Re$ based on the Dedekind $\eta$ function. The trefoil knot is a fibered knot. The map $\Delta/|\Delta|: S^3 \setminus \ell \to S^1 \subset \mathbb{C}$ is a locally trivial fibration whose fibers are punctured tori. Given a closed oriented curve $\gamma$ in the complement of the trefoil knot, the linking number $\text{link}(\gamma, \ell)$ is the topological degree of the restriction of $\Delta/|\Delta|$ to $\gamma$ (en passant, this defines an orientation for $\ell$).

Jacobi established a connection between $\Delta/|\Delta|$ and the Dedekind $\eta$ function (see [2]). If we denote by $\Delta(\omega_1, \omega_2)$ the $\Delta(\omega_1, \omega_2)$ invariant of the lattice $\mathbb{Z} \cdot \omega_1 + \mathbb{Z} \cdot \omega_2 \subset \mathbb{C}$ (with $\Im(\omega_2/\omega_1) > 0$), then

$$\Delta(\omega_1, \omega_2) = (2\pi)^{12} \omega_1^{-12} \eta \left( \frac{\omega_2}{\omega_1} \right)^{24}.$$  

Consider the periodic orbit of period $T > 0$ associated to a hyperbolic element $A = \pm (a \ b \ c \ d)$ in $\text{PSL}(2, \mathbb{Z})$. One can describe it as a closed curve of lattices $\mathbb{Z} \cdot \delta t \omega_1 + \mathbb{Z} \cdot \delta t \omega_2$ ($t \in [0, T]$) such that

$$\delta T (\omega_2) = a\omega_1 + b\omega_2; \quad \delta T (\omega_1) = c\omega_1 + d\omega_2.$$  

We wish to compute the variation $\text{VarArg} \Delta$ of the argument of $\Delta(\delta t \omega_1, \delta t \omega_2)$ as $t$ goes from 0 to $T$. To fix notation, given a curve $q(t)$ in $\mathbb{C}^*$ ($t \in [0, T]$), written as $\exp(2i\pi \tau(t))$ for some continuous $\tau(t)$, the variation of the argument $\text{VarArg} q$ is defined as $\Re(\tau(T) - \tau(0))$. By Jacobi's theorem, $\text{VarArg} \Delta(\delta t \omega_1, \delta t \omega_2)$ is equal to:

$$-12 \text{VarArg}(\delta t \omega_1) + 24 \frac{1}{2\pi} \Re \left( \log \eta \left( \frac{\delta T \omega_2}{\delta T \omega_1} \right) - \log \eta \left( \frac{\omega_2}{\omega_1} \right) \right).$$  

Note that $\delta T \omega_2/\delta T \omega_1 = (a\omega_1 + b)/(c\omega_1 + d)$, so that we can use the definition of the Rademacher function using the logarithm of $\eta$. We get:

$$-12 \text{VarArg}(\delta t \omega_1) + 6 \frac{1}{2\pi} \Re \log \left( - \left( \frac{\omega_2}{\omega_1} + d \right)^2 \right) + \Re(A).$$  

Observe that the curve $\delta t \omega_1$ is contained in a quadrant, so that $\text{VarArg}(\delta t \omega_1)$ belongs to the interval $(-1/4, 1/4)$ and is therefore equal to $\frac{1}{2\pi} \Re \log \left( - \left( \frac{\delta t \omega_1}{\omega_1} \right)^2 \right)$. Recall that $\log$ denotes the determination with imaginary part in $(-\pi, +\pi)$. Hence two terms cancel, and we get that link$(k_A, \ell)$ is indeed equal to $\Re(A)$, as claimed.
3.4. A topological approach. Let us sketch a purely topological computation of \( \text{link}(k_A, \ell) \), related to another approach to the Rademacher function.

Consider a compact oriented surface \( S \) with fundamental group \( \Gamma \) equipped with a hyperbolic metric. For each element \( \gamma \) in \( \Gamma \), denote by \( \overline{\gamma} \) the closed geodesic which is freely homotopic to \( \gamma \). This defines a periodic orbit \( k_\gamma \) of the geodesic flow in the unit tangent bundle \( T^1 S \) of \( S \). If \( \gamma_1, \gamma_2 \) are in \( \Gamma \), there is an obvious singular 2-chain \( c(\gamma_1, \gamma_2) \) in \( S \) whose boundary is \( \overline{\gamma_1 \gamma_2} - \overline{\gamma_1} - \overline{\gamma_2} \). The obstruction to lift \( c(\gamma_1, \gamma_2) \) to a 2-chain in \( T^1 S \) with boundary \( k_{\gamma_1 \gamma_2} - k_{\gamma_1} - k_{\gamma_2} \) is an integer \( \text{eu}(\gamma_1, \gamma_2) \in \mathbb{Z} \). In other words, one can find a 2-chain in \( T^1 S \) whose boundary is \( k_{\gamma_1 \gamma_2} - k_{\gamma_1} - k_{\gamma_2} + \text{eu}(\gamma_1, \gamma_2) \ell \) where \( \ell \) denotes one fiber of \( T^1 S \), and which projects on \( c(\gamma_1, \gamma_2) \). This defines a 2-cocycle on \( \Gamma \) whose cohomology class is the Euler class of the circle bundle. This construction generalizes to the noncompact modular orbifold \( \Sigma = \mathbb{D} / \text{PSL}(2, \mathbb{Z}) \) with a little care. One has to adapt the definition of \( k_A \) for elliptic and parabolic elements. Since the second rational cohomology of \( \text{PSL}(2, \mathbb{Z}) \) is trivial, there is a map \( \Phi : \text{PSL}(2, \mathbb{Z}) \to \mathbb{Q} \) such that \( \Phi(\gamma_1 \gamma_2) - \Phi(\gamma_1) - \Phi(\gamma_2) = \text{eu}(\gamma_1, \gamma_2) \). Note that this defines uniquely \( \Phi \) since there is no nontrivial homomorphism from \( \text{PSL}(2, \mathbb{Z}) \) to \( \mathbb{Q} \). It turns out that \( 6 \Phi \) and \( \Re \) agree on hyperbolic elements of \( \text{PSL}(2, \mathbb{Z}) \) (see [6], [11]): this is the topological aspect of \( \Re \).

Let us temporarily denote \( \text{link}(k_A, \ell) \) by \( \lambda(A) \). In order to show that \( \lambda(A) = 6 \Phi(A) \), it is enough to show that \( \lambda(AB) - \lambda(A) - \lambda(B) = 6 \text{eu}(A, B) \). Let \( D_A, D_B \) and \( D_{AB} \) be singular disks in \( \mathbb{S}^3 \) with boundaries \( k_A, k_B, k_{AB} \) respectively. By definition of the linking number, the intersection numbers of these disks with \( \ell \) are \( \lambda(A), \lambda(B), \lambda(AB) \). Choose a singular surface in \( T^1 \Sigma \cong \mathbb{S}^3 \setminus \ell \) with boundary \( k_{AB} - k_A - k_B + \text{eu}(A, B) \ell \). Glue this surface to \( D_A, D_B, D_{AB} \) along the boundaries and cap the result with a disk in \( \mathbb{S}^3 \) with boundary \( \text{eu}(A, B) \ell \), with intersection number \( 6 \text{eu}(A, B) \) with \( \ell \). Note that the linking number between \( \ell \) and \( \ell \) is 6. The resulting boundaryless (singular) surface in \( \mathbb{S}^3 \) has an intersection number 0 with \( \ell \) since the homology of the sphere is trivial. Putting things together, we get
\[
\lambda(AB) - \lambda(A) - \lambda(B) - 6 \text{eu}(A, B) = 0
\]
as required.

3.5. Lorenz and modular knots. We now turn to a dynamical proof which will lead to a topological description of these modular knots.

Recall that a Lorenz knot is a knot isotopic to a periodic orbit of the Lorenz differential equation. We will establish a close connection between the Lorenz knots and the modular dynamics:

**Isotopy classes of Lorenz knots and modular knots coincide.**

We first deform the embedding of \( \text{PSL}(2, \mathbb{Z}) \) in \( \text{PSL}(2, \mathbb{R}) \) in order to produce a discrete subgroup of infinite covolume. Recall that \( \text{PSL}(2, \mathbb{Z}) \) is isomorphic to a free product of \( \mathbb{Z}/2\mathbb{Z} \) and \( \mathbb{Z}/3\mathbb{Z} \) corresponding to the elements of order 2 and 3:
\[
U = \pm \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}; \quad V = \pm \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}.
\]
Consider two points \( x, y \) in the Poincaré disk at distance \( \rho \geq 0 \). This defines a homomorphism \( i_\rho : \text{PSL}(2, \mathbb{Z}) \to \text{PSL}(2, \mathbb{R}) \) sending \( U \) to the symmetry with respect to \( x \), and \( V \) to the rotation of angle \( 2\pi/3 \) around \( y \). Note that, up to conjugacy, \( i_\rho \) only depends on \( \rho \), and that the canonical embedding corresponds to some explicit value \( \rho_0 \) (the hyperbolic distance between \( \sqrt{-1} \) and \( (-1 + \sqrt{-3})/2 \) in Poincaré’s upper half plane). When \( 0 < \rho < \rho_0 \), the image is a dense subgroup. When \( \rho > \rho_0 \), the image of \( i_\rho \) is a discrete subgroup with infinite covolume: “the cusp has been opened”. The quotient \( \Sigma_\rho \) of \( \mathbb{D} \) by \( i_\rho \) \( \text{PSL}(2, \mathbb{Z}) \) is a noncompact orbifold with a “funnel”.

![Figure 6. Deforming the modular surface.](image)

Of course, for \( \rho \geq \rho_0 \), all the quotients \( M_\rho = \text{PSL}(2, \mathbb{R}) / i_\rho \text{PSL}(2, \mathbb{Z}) \) are homeomorphic to the complement of the trefoil knot.

For \( \rho > \rho_0 \), there is also a flow \( \phi_t^\rho \) on \( M_\rho \) given by left translations by diagonal matrices; this is the geodesic flow on the orbifold \( \Sigma_\rho \) of infinite area. The limit set \( K_\rho \subset \partial \mathbb{D} \) of the Fuchsian group \( i_\rho(\text{PSL}(2, \mathbb{Z})) \) is a Cantor set. The action of \( i_\rho(\text{PSL}(2, \mathbb{Z})) \) on the convex hull \( \hat{K}_\rho \subset \mathbb{D} \) is cocompact: the quotient is a compact orbifold \( \Sigma_\rho^{\text{conv}} \subset \Sigma_\rho \) with one geodesic boundary component and two singular points.

Geodesics in \( \mathbb{D} \) whose two limit points are in \( K_\rho \) define a compact set \( \Omega_\rho \in \tau^1 \Sigma_\rho = \text{PSL}(2, \mathbb{R}) / i_\rho \text{PSL}(2, \mathbb{Z}) \) which is invariant under \( \phi_t^\rho \); this is the nonwandering set. Of course, this invariant set is hyperbolic in the sense of dynamical systems, and the now classical hyperbolic theory of Hadamard–Morse–Anosov–Smale implies that the restrictions of \( \phi_t^\rho \) to \( \Omega_\rho \) are all equivalent by some homeomorphisms (sending orbits to orbits, respecting their orientations, but of course not respecting time). Periodic orbits of \( \phi_t^\rho \) are contained in \( \Omega_\rho \) so that, in particular, all flows \( \phi_t^\rho \) in \( \mathbb{S}^3 \setminus \ell \) carry the same (isotopy classes of) links (as soon as \( \rho > \rho_0 \)). Clearly, the original flow \( \phi_t = \phi_t^\rho_0 \) is not topologically conjugate to \( \phi_t^\rho \) (for \( \rho > \rho_0 \)) since most orbits of \( \phi_t \) are dense, and this is not the case for \( \phi_t^\rho \) (\( \rho > \rho_0 \)). However, when \( \rho \) decreases to \( \rho_0 \), closed orbits of \( \phi_t^\rho \), which correspond to closed geodesics in \( \Sigma_\rho^{\text{conv}} \), converge to periodic orbits of \( \phi_t \), with the exception of (the multiples of) the geodesic boundary of \( \Sigma_\rho^{\text{conv}} \) which “escapes at infinity in the cusp”.

In other words, with the exception of boundary geodesics, corresponding to parabolic elements in \( \text{PSL}(2, \mathbb{Z}) \), the periodic knots associated to \( \phi_t^\rho \) (\( \rho > \rho_0 \)) are (isotopic to) the modular knots we want to describe. We are therefore led to give a description of the topology of periodic orbits of \( \phi_t^\rho \) (\( \rho > \rho_0 \)).
Look at Figure 7. Consider a geodesic \( u: \mathbb{R} \to \mathbb{D} \) with endpoints \( u(-\infty) \) in the interval \( I \) and \( u(+\infty) \) in the interval \( J \). It intersects the central hexagon on a compact arc, and the union of these arcs defines an embedding \( j \) of \( I \times J \times [0, 1] \) in \( T^1\mathbb{D} \cong \text{PSL}(2, \mathbb{Z}) \). Projecting this parallelepiped in \( \text{PSL}(2, \mathbb{R})/i_{\rho}\text{PSL}(2, \mathbb{Z}) \), one gets an embedding of \( I \times J \times (0, 1) \) in \( T^1\Sigma_\rho \), but the top and the bottom faces do intersect in the projection. Figure 8 describes the projected parallelepiped \( P \subset \text{PSL}(2, \mathbb{R})/i_{\rho}\text{PSL}(2, \mathbb{Z}) \), which is a compact manifold with boundary and corners.

The maximal \( \phi^t_\rho \) invariant set contained in \( P \) is of course the nonwandering set \( \Omega_\rho \). The restriction of \( \phi^t_\rho \) to \( \Omega_\rho \) is therefore conjugate to the suspension of a full shift on two symbols \( \{\text{left, right}\} \). A nonwandering geodesic travels in the convex hull \( \hat{K}_\rho \subset \mathbb{D} \), intersects successively \( \text{PSL}(2, \mathbb{Z})\)-translates of the hexagon, and might exit by the right or left exit, as seen from the entrance side. Any bi-infinite sequence is possible and the sequence characterizes the geodesic.

We now use the main idea of Birman–Williams’ template theory. In each of the rectangles \( j(I \times J_{\text{left}} \times [0, 1]) \) and \( j(I \times J_{\text{right}} \times [0, 1]) \), collapse the strong stable manifolds. This produces two rectangles forming a branched manifold which is embedded in \( M_\rho \cong S^3 \setminus \ell \).

We still have to explain why it is embedded in the way described in Figure 9. Assuming this for a moment, we recognize the Lorenz template, which carries Lorenz knots and links. The process of collapsing the stable manifolds can be done in a smooth way, so that the periodic orbits move by some isotopy (note that a periodic orbit intersects a strong stable manifold in at most one point, so that the collapse does not introduce double points). In other words, the periodic links of \( \phi^t_\rho \) are precisely the periodic links on the template, i.e. the Lorenz links.

We briefly explain why the template is indeed embedded as in the picture. We basically have to prove that the two Mickey Mouse ears represent a trivial two com-
ponent link, and that the ears are untwisted. The template consists of two rectangles (symmetric with respect to the involution $v \mapsto -v$ in $T^1 \Sigma_\rho$). Each consists of the periodic orbit corresponding to (one orientation of) the boundary of $\Sigma_\rho^{\text{conv}}$ and a piece of the unstable manifold of this orbit. This rectangle projects in $\Sigma_\rho^{\text{conv}}$ as a neighborhood of the boundary curve. In the original modular surface, the rectangle projects as in Figure 10, that one can push as close as one wants towards the cusp.

From the lattice point of view, the first rectangle consists of (rescaled) lattices of the form $Z + Z \cdot \tau$ with $\Im \tau > 1$. The Weierstrass invariants $(g_2, g_3)$ of such lattices are given by the classical formulas

$$g_2(q) = \frac{4\pi^4}{3} (1 + 240q + \cdots); \quad g_3(q) = \frac{8\pi^6}{27} (1 - 504q + \cdots)$$

where, as usual, $q = \exp(2i\pi \tau)$. This means that the rectangle sits inside (a rescaling of) the holomorphic disk $q \mapsto (g_2(q), g_3(q)) \in \mathbb{C}^2$ which is an embedding for $|q|$ small enough, and intersects transversally the curve $\{\Delta = 0\}$ since $\Delta(q) = (2\pi)^{12} (q - 24q^2 + \cdots)$ for small $q$. One concludes first of all that the periodic orbit corresponding to the boundary of the rectangle is unknotted in the sphere since it can be isotoped in this embedded disk. Second of all, it implies that the rectangle is untwisted, since it can also be pushed in an embedded disk. Finally, this implies that the linking number between the boundary curve and the trefoil knot is 1.

The second rectangle is the image of the first one by the symmetry $v \mapsto -v$ (which, from the lattice side, corresponds to one quarter turn). One has to consider now lattices of the form $i(Z + Z \cdot \tau)$ with $\Im \tau > 1$ for which the invariants are $g_2(q)$, $-g_3(q)$. The situation is exactly the same as before except that the boundary geodesic is now described with the other orientation, and has a linking number $-1$ with the trefoil. Moreover, we see that the two boundary periodic orbits define a trivial two component link (since they bound disjoint embedded disks). From this information, one can deduce that the template is indeed embedded as in Figure 9.

This finishes the sketch of proof that (isotopy classes of) Lorenz and modular knots coincide. To be precise, we should be careful with the two boundary trivial knots that we just discussed, which appear on the template, but not in the modular surface (since they were pushed to infinity). However, since some modular knots are

![Figure 9. Modular template.](image1)

![Figure 10. Cusp neighborhood.](image2)
trivial knots, one can state that modular knots and Lorenz knots coincide. Of course, one does not have to restrict to knots, and we could also discuss links as well. The same proof shows that \textit{all modular links are isotopic to Lorenz links and, conversely, that a Lorenz link with no exceptional component is isotopic to a modular link}.\footnote{Note added in proof. The reader may look at the AMS feature Column by É. Ghys and J. Leys: Lorenz and modular knots, a visual introduction, AMS Feature Column, November 2006, http://www.ams.org/featurecolumn/archive/lorenz.html}

Figure 11 represents the simultaneous position of the template and the trefoil knot (easy to prove). \textit{This picture provides a third computation} of \textit{link} \((k_A, \ell)\). Indeed, up to conjugacy, any hyperbolic element \(A\) in \(\text{PSL}(2, \mathbb{Z})\) can be written as a product

\[ A = UV^{\varepsilon_1}UV^{\varepsilon_1} \ldots UV^{\varepsilon_n} \]

where each \(\varepsilon_i\) is equal to \(\pm 1\). From the dynamical point of view, this means that the corresponding closed geodesic follows the template, turning left or right successively according to the signs of the \(\varepsilon_i\)’s. Since we know that the trefoil knot has linking number +1 with the first ear and −1 with the second, we obviously get:

\[ \text{link} \,(k_A, \ell) = \sum_{1}^{n} \varepsilon_i. \]

This is a third version of the Rademacher function [6], [11]! The reader will notice some analogy between this left-right count and the signature invariant that we discussed earlier. This is not surprising since it turns out that the spherical braid group \(B_4(S^2)\) is isomorphic to \(\text{SL}(2, \mathbb{Z})\), and that the signature is (a multiple of) the Rademacher function [39].

As a corollary of the description of modular knots and links, we get the following:

\textit{Modular links are fibered links and have nonnegative signature. Modular knots are prime. The knot} \(k_A\) \textit{is trivial if and only if} \(A\) \textit{is conjugate to a word of the form} \((UV)^a(UV^{-1})^b\) \((a, b \geq 1)\).

\textit{Indeed, these properties hold for Lorenz knots} [17], [91]!

\textit{It would be nice to understand those fibrations from the modular side: for instance, can one find some “arithmetical” description of the fibrations of the complements}
of $k_A$? In [17], the authors suggest that there could be some “natural limit” to the fibrations of $S^3 \setminus L$ as $L$ describes all Lorenz links. Maybe the modular point of view will answer this question, and build a bridge between Riemann’s $\zeta$ function and dynamical $\zeta$ functions (see for instance [89]).

Another question would be to give an arithmetical or combinatorial computation of the linking numbers of two knots $k_A$ and $k_B$ as a function of $A, B$ in $\text{PSL}(2, \mathbb{Z})$ (compare [54]). One could also try to understand more sophisticated link invariants for these modular links.

**Final remorse.** Many interesting questions should have been discussed in this survey, like energy bounds and asymptotic crossing numbers, Hofer metric, global geometry of groups of symplectic diffeomorphisms etc. This is a good excuse to suggest [35], [74], [75] as additional reading!

**Acknowledgment.** It is a pleasure to thank Jean-Marc Gambaudo for his friendly collaboration.

**References**


Knots and dynamics


École Normale Supérieure de Lyon, Unité de Mathématiques Pures et Appliquées, UMR CNRS 5669, 46, Allée d’Italie 69360 Lyon, France
E-mail: etienne.ghys@umpa.ens-lyon.fr