BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 42, Number 2, Pages 257–261
S 0273-0979(05)01051-7
Article electronically published on February 10, 2005

On some aspects of the theory of Anosov systems, by G. A. Margulis, with a survey, "Periodic orbits of hyperbolic flows", by Richard Sharp, Springer Monographs in Mathematics, Springer, Berlin, 2004, 138 pp., $69.95, ISBN 3-540-40121-0

Riemann's interest in the zeta function was surely in response to the observations (and calculations) of Gauss and Legendre regarding the distribution of prime numbers which suggested the asymptotic formula, now known as the Prime Number Theorem (P.N.T.):

$$\pi(x) \sim \frac{x}{\log x} \quad \text{as} \quad x \to \infty,$$

where $\pi(x) = \#\{\text{primes} \leq x\}$. But it is safe to say that the program he outlined was, to say the least, much more profound than any 'frequency' theorem he might have wished to prove, bringing as it did the analytic methods of Dirichlet to a new level. Riemann proposed that a fuller understanding of the distribution of prime numbers should result from further analysis of the zeta function

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} = \prod_p (1 - p^{-s})^{-1}$$

(where the product is over primes), which had been substantially investigated by Euler, and indeed Hadamard and de la Vallée Poussin confirmed Riemann's idea (utilising his results) following Chebychev's earlier result which states that

$$\frac{\pi(x)\log x}{x}$$

is bounded away from zero and infinity.

Riemann proved that $\zeta(s)$ has a meromorphic extension to $\mathbb{C}$ and that it satisfies a simple functional equation; hence the importance of locating zeros and poles, and in particular the Riemann Hypothesis.

These fundamental ideas proved to be so fruitful that over the last 150 years or so a plethora of zeta functions has arisen (in a variety of fields) associated with, for example, Dedekind, Epstein, Artin, Mazur, Weil, Selberg, to name but a few.

Selberg [Se] was the first to introduce a zeta function which has a direct bearing on dynamical systems, for it encapsulates the data of closed geodesic lengths for surfaces of constant curvature -1, and these geodesics are in one-one correspondence with the periodic orbits of the associated geodesic flow which takes place on the unit tangent bundle of the manifold (where a unit tangent vector is carried along its geodesic at unit speed).

The Selberg zeta function is defined by

$$Z(s) = \prod_{\gamma} \prod_{k=0}^{\infty} (1 - e^{-(s+k)l(\gamma)}),$$

where the product is over all closed geodesics γ and l denotes length. Selberg proved that $Z(s)$ has an extension to $\mathbb{C}$ which is entire and except for those on the real line (which can be specified) its zeros lie on $\Re(s) = {}^1/_2$, thus confirming the Riemann Hypothesis in this context. Moreover the zeros of $Z(s)$ are intimately

2000 *Mathematics Subject Classification.* Primary 37D20, 37D35.

related to the eigenvalues of the Laplace–Beltrami operator which, via Selberg's trace formula, are closely bound up with the lengths of the closed geodesics.

Huber and others were to use these ground-breaking results to prove an analogue of the P.N.T. (with closed geodesics in place of primes) and with error terms – the so-called Prime Orbit Theorem (P.O.T.).

This, briefly, is the background to one of the main aspects of the book under review which comprises Margulis's previously unpublished 1970 Ph.D. thesis (although see [M] for a partial account) together with a very useful and pleasantly economical survey by Richard Sharp of subsequent developments regarding the asymptotics of closed orbits of dynamical systems. The book is almost equally divided between the helpfully detailed thesis and the survey.

Margulis's thesis concerns Anosov flows. These differentiable dynamical systems (first order differential equations on compact Riemannian manifolds) generalise geodesic flows on manifolds of variable negative curvature. The formal definition requires that the tangent bundle may be continuously split into invariant subbundles E^u, E^s, E^o where, under the differential of the flow, vectors in E^u (E^s) expand (contract) exponentially and E^o is the one-dimensional bundle of vectors tangent to orbits, expansion and contraction coefficients being bounded.

Margulis proves the P.O.T. in this more general context which takes the form

$$\pi(x) \sim \frac{e^{hx}}{hx} \quad \text{as} \quad x \to \infty$$

where $\pi(x) = \#\{\text{closed orbits } \gamma : l(\gamma) < x\}$ and where h (the standard measure of 'chaos') is the topological entropy of the flow or equivalently the least upper bound of measure theoretic entropies. (If x is replaced by $\log(x)/h$, we get the more familiar ratio of the P.N.T. of course.) Prior to this Sinai [Si.1] had found bounds for $\log \pi(x)/x$.

Margulis's proof is intertwined with his proof of an equally important theorem which describes how closed orbits are uniformly distributed within the manifold, where the uniformity is with respect to a special flow-invariant probability measure which is now known as the Margulis measure – it is the unique measure of maximal entropy. In fact the construction of this measure dominates the first half of his thesis.

When Anosov introduced his flows he also defined discrete versions, namely diffeomorphisms (and their iterates) of compact manifolds [A]. They are defined in a similar way to the flows except that the one-dimensional sub-bundle is now absent. A prototype of an Anosov diffeomorphism is provided by an automorphism of the torus whose defining matrix has no unit modulus eigenvalues. Anosov had laid the foundations for the dynamical theory of his systems including much of the measure theory. But it was Sinai [Si.2] who provided the subsequent constructions needed for ergodic theory, and for this he developed what was to prove to be one of the most powerful methods in this field, namely the construction and utilisation of Markov partitions for Anosov diffeomorphisms. (For two-dimensional torus automorphisms this had been done by Adler and Weiss.) Markov partitions enable a diffeomorphism to be represented by a topological Markov chain (or shift of finite type), and this can be done almost faithfully.

However, Margulis's thesis could not benefit from this method as its relevance for flows was only subsequently established in the work of Ratner [Ra] and Bowen [B.1]. He therefore had to invent methods which perhaps even now have not been fully exhausted. This involved an analysis of the Riemann volume induced on the unstable (expanding) leaves associated with E^u which (via functional analysis) he utilised in the construction of special leaf measures (in general singular with respect to the Riemann measure) which are *uniformly* expanded under the flow. The expansion rate is given precisely by the topological entropy (or rather its exponential). A similar construction applies to stable (contracting) leaves associated with E^s where we have a contracting rate which exactly cancels the E^u expansion rate. By putting these measures together, so to speak, a flow-invariant measure – the Margulis measure – is produced. The argument is delicate and requires the comparison of leaf measures for different leaves. After this construction Margulis established the ergodic theory of Anosov flows with respect to this measure and in particular he shows they are K flows ('K' for Kolmogorov) (assuming there are no continuous eigenfunctions) and hence ergodic, and thereby his work subsumes the classical results of Hopf and Hedlund for geodesic flows on manifolds of negative curvature. Here one should note that when the system preserves a 'smooth' measure this ergodic theory had already been established; see for example [A.Si]. The novelty of Margulis's work centres on his special and important measure.

Around the time that Anosov introduced his systems, Smale (see [Sm] for details) was investigating his hyperbolic dynamics in the form of Axiom A diffeomorphisms and flows. These are also defined in terms of the tangent bundle structure, in fact with the same requirements when restricted to vectors tangent to the non-wandering set. (The set consists of points whose neighbourhoods necessarily return under some iteration or under the flow.) This relaxation vastly expanded the scope of the methods of the Russian school and opened up the field to new methods due in large part to Sinai, Ruelle and Bowen.

Perhaps the most important new ingredient was the introduction of statistical/thermodynamical concepts (free energy, pressure, equilibrium states, etc.) which was made possible following the Sinai-Bowen construction of Markov partitions for these more general systems, both for hyperbolic diffeomorphisms and hyperbolic flows. This enabled results for lattice gases (developed by Ruelle [R.1] and others) to be applied to Smale's systems. The full story is too long to be related here.

With the aid of Markov partitions it is possible to transfer many problems to an associated topological Markov chain where for each Lipschitz function (a potential) there is a positive Perron–Frobenius operator which is quasi-compact when acting on a suitable Banach space. In the absence of a geometrically defined Laplace–Beltrami operator, which was central to Selberg's work, this transfer operator provides the necessary data for a host of problems. In particular properties of the zeta function are deduced from properties of this operator (eigenvalues are related to periodic orbits) and rates of convergence (in 'mixing' problems) are calculated using properties of the operator, and each potential is associated with an invariant probability measure – its equilibrium state.

Margulis's uniform distribution theorem and an analogue of the Chebychev theorem were proved in this more general setting by Bowen [B.2]. Subsequently the sharper P.O.T. was proved by Parry and Pollicott [P.P] following an earlier more restricted version. The methods here were adaptations of the Wiener–Ikehara Tauberian method and relied on many of Ruelle's thermodynamic results, and of course Bowen's Markov partitions were indispensable. One of the advantages of the transfer operator procedure is that one can design it (using an appropriate potential) to serve a variety of problems. For example the uniform distribution theorem can be generalised by replacing lengths of closed orbits by integrals of the potential around closed orbits which results in a limit corresponding to the potential's equilibrium state. The method can also be applied to skew-products (or group extensions with compact group) whose quotient dynamics are hyperbolic. Here we encounter problems which are analogous to the Galois extension problems of algebraic number theory, which were first investigated by Sarnak. When the group is countable and abelian (no longer compact) the same methods apply (through subtle improvements and different Tauberian theorems) and can be used to obtain asymptotic information about closed orbits in a fixed homology class.

Evidently there are several levels of generality here, beginning with geodesic flows on manifolds of constant negative curvature, then moving to variable negative curvature and on to Anosov flows and hyperbolic (Axiom A) flows. The symbolic method behind all these is a suspension flow (with Hölder return times) defined over a topological Markov chain. Generality, say at the level of hyperbolic flows, is bought at a cost regarding error estimates. Or to put it another way, the more geometry we have (the negative curvature situation, especially when the curvature is constant) the less we have to rely on the Markov partition method which, although sufficient for interesting results, generally lacks precision for higher order asymptotics.

Nevertheless Dolgopyat [D] has shown that Markov methods are still capable of yielding surprisingly good results. For Anosov flows he has obtained a uniform bound on the norms of iterates of the transfer operator L^n_{-shr} (fixed function r), in a domain $\Re(s) > \sigma_0$, where $\sigma_0 < 1$. This has many consequences and in particular Pollicott and Sharp [P.S] have used it for the geodesic flow case (when the manifold is either two-dimensional or '$^1/_4$ pinched') to show

$$\pi(x) = \mathrm{li}(\exp hx) + O(\exp cx)$$

for some $c < h$. Here li is the usual analytic number function $\mathrm{li}(y) = \int_2^y \frac{\mathrm{d}u}{\log u} \sim \frac{y}{\log y}$.

Sharp's lucid account covers some successful attempts to improve the error terms for asymptotic results in the context of general hyperbolicity (i.e. the less than ideal settings lacking 'geometry'), but his main emphasis concerns distribution theorems under a variety of constraints. There are references to Ruelle's results (see [R.2] for an overview) when the system is assumed to be analytic with analytic foliations and to the results of Rugh [Ru] and Fried [F] which depend on weaker (but nevertheless quite strong) assumptions.

Such hypotheses lead to improved extendability results for the zeta functions either to the whole of $\mathbb{C}$ or to a strip to the left of $\Re(s) = 1$ (where there is a simple pole at $s = 1$). From there, error terms can be made more precise and convergence theorems such as rates of decay of correlations are sharpened. However, there are many problems requiring further investigation. For some time now it has seemed desirable to find a way of avoiding the convoluted procedure of passing to shifts of finite type (and then on to one-sided shifts) in order to employ a transfer operator, and some new ideas of Gouezel and Liverani have met with some success. Transfer operators have a dual relationship to what are surely the most natural operators, those induced by the diffeomorphisms of the flow. The one-parameter group of unitary operators (composition with the flow followed by a suitable multiplication) has self-adjoint infinitesimal generators unbounded on Hilbert space but perhaps more amenable on a suitable subspace. The general problem of finding optimal conditions for full extendability and for a functional equation seems to be wide open. Even when limited to shifts of finite type we do not know the best conditions on ceiling functions (return times) which will lead to ideal zeta functions.

There are many other problems which I am not competent to sensibly comment on. There is a great deal of current research activity engaging physicists and number theorists regarding dynamical zeta functions. Most of this has to do with the Hilbert–Polya 'dream' of relating the Riemann zeta function to some as yet undiscovered operator and its eigenvalues. Much has to do with an attempt to understand the process of transition from quantum to classical mechanics. There is every reason to hope that deep connections between dynamics and number theory are yet to be revealed.

Margulis's thesis is an excellent text for graduate students looking for problems, and Sharp's update will surely guide ambitious research students toward 'rich pickings'.

References

[A] D. V. Anosov. Geodesic flows on closed Riemannian manifolds of negative curvature. *Proc. Steklov Inst. Math.* **90** (1967). Transl. A. M. S. MR0224110 (36:7157)

[A.Si] D. V. Anosov and Ya. G. Sinai. Certain smooth ergodic systems. *Russ. Math. Surv.* **22** No. 5 (1967), 103–167. MR0224771 (37:370)

[B.1] R. Bowen. Symbolic dynamics for hyperbolic flows. *Amer. J. Math.* **95** (1973), 429–460. MR0339281 (49:4041)

[B.2] R. Bowen. The equidistribution of closed geodesics. *Amer. J. Math.* **94** (1972), 413–423. MR0315742 (47:4291)

[D] D. Dolgopyat. On decay of correlations in Anosov flows. *Ann. of Math.* **147** (1998), 357–390. MR1626749 (99g:58073)

[F] D. Fried. Meromorphic zeta functions for analytic flows. *Comm. Math. Phys.* **174** (1995), 161–190. MR1372805 (97a:58148)

[M] G. Margulis. Certain measures associated with У-flows on compact manifolds. *Func. Anal. Appl.* **4** (1969), 55–67. MR0272984 (42:7865)

[P.P] W. Parry and M. Pollicott. An analogue of the prime number theorem for closed orbits of Axiom A flows. *Ann. of Math.* **118** (1983), 573–591. MR0727704 (85i:58105)

[P.S] M. Pollicott and R. Sharp. Exponential error terms for growth functions on negatively curved surfaces. *Amer. J. Math.* **120** (1998), 1019–1042. MR1646052 (99h:58148)

[Ra] M. Ratner. Markov partitions for Anosov flows on n-dimensional manifolds. *Israel J. Math.* **15** (1973), 92–114. MR0339282 (49:4042)

[R.1] D. Ruelle. *Thermodynamic Formalism*. Addison Wesley, Reading, MA, 1978. MR0511655 (80g:82017)

[R.2] D. Ruelle. *Dynamical zeta functions for piecewise monotone maps of the interval*. CRM monograph series No. 4. A. M. S., 1994. MR1274046 (95m:58101)

[Ru] H. H. Rugh. The correlation spectrum for hyperbolic analytic maps. *Nonlinearity* **5** (1992), 1237–1263. MR1192517 (93i:58121)

[Se] A. Selberg. Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series. *J. Indian Math. Soc.* **20** (1956), 47–87. MR0088511 (19:531g)

[Si.1] Ya. G. Sinai. The asymptotic behaviour of the number of closed geodesics on a compact manifold of negative curvature. *Transl. A. M. S.* **73** (1968), 227–250. MR0229178 (37:4752)

[Si.2] Ya. G. Sinai. The construction of Markov partitions. *Func. Anal. Appl.* **2** (1968), 70–80. MR0250352 (40:3591)

[Sm] S. Smale. Differentiable dynamical systems. *Bull. A. M. S.* **73** (1967), 747–817. MR0228014 (37:3598)

William Parry
University of Warwick
E-mail address: b.parry@warwick.ac.uk