MERTENS' THEOREM FOR TORAL AUTOMORPHISMS

SAWIAN JAIDEE, SHAUN STEVENS, AND THOMAS WARD

Abstract. A dynamical Mertens' theorem for ergodic toral automorphisms with error term $\mathrm{O}(N^{-1})$ is found, and the influence of resonances among the eigenvalues of unit modulus is examined. Examples are found with many more, and with many fewer, periodic orbits than expected.

1. Introduction

Discrete dynamical analogs of Mertens' theorem concern a map $T : X \to X$, and are motivated by work of Sharp [6] on Axiom A flows. A set of the form

$$\tau = \{x, T(x), \ldots, T^k(x) = x\}$$

with cardinality k is called a closed orbit of length $|\tau| = k$, and the results provide asymptotics for a weighted sum over closed orbits. For the discrete case of a hyperbolic diffeomorphism T, we always have

$$M_T(N) := \sum_{|\tau| \leqslant N} \frac{1}{\mathrm{e}^{h|\tau|}} \sim \log(N),$$

where h is the topological entropy, with more explicit additional terms in many cases. The main term $\log(N)$ is not really related to the dynamical system, but is a consequence of the fact that the number of orbits of length n is $\frac{1}{n}\mathrm{e}^{hn} + \mathrm{O}(\mathrm{e}^{h'n})$ for some $h' < h$ (see [5]). Without the assumption of hyperbolicity, the asymptotics change significantly, and in particular depend on the dynamical system. For quasihyperbolic (ergodic but not hyperbolic) toral automorphisms, Waddington [8] found asymptotics for an unweighted orbit-counting sum, and Noorani [4] found an analogue of Mertens' theorem in the form

$$M_T(N) = m \log(N) + C_1 + \mathrm{o}(1) \tag{1}$$

for some $m \in \mathbb{N}$. The constant C_1 is related to analytic data coming from the dynamical zeta function. For more general non-hyperbolic group automorphisms, the coefficient of the main term may be non-integral (see [2] for example).

In this note Noorani's result (1) with error term $\mathrm{O}(N^{-1})$ is recovered using elementary arguments, and the coefficient m of the main term in (1) is expressed as an integral over a sub-torus. This reveals the effect of resonances between the eigenvalues of unit modulus, and examples show that the value of m may be very different to the generic value given in [4].

Date: October 12, 2012.

2000 *Mathematics Subject Classification.* 37C35, 11J72.

2. Toral automorphisms

Let $T : \mathbb{T}^d \to \mathbb{T}^d$ be a toral automorphism corresponding to a matrix A_T in $\mathrm{GL}_d(\mathbb{Z})$ with eigenvalues $\{\lambda_i \mid 1 \leqslant i \leqslant d\}$, arranged so that

$$|\lambda_1| \geqslant \cdots \geqslant |\lambda_s| > 1 = |\lambda_{s+1}| = \cdots = |\lambda_{s+2t}| > |\lambda_{s+2t+1}| \geqslant \cdots \geqslant |\lambda_d|.$$

The map T is *ergodic* with respect to Lebesgue measure if no eigenvalue is a root of unity, is *hyperbolic* if in addition $t = 0$ (that is, there are no eigenvalues of unit modulus), and is *quasihyperbolic* if it is ergodic and $t > 0$. The topological entropy of T is given by $h = h(T) = \sum_{j=1}^{s} \log|\lambda_j|$.

Theorem 1. *Let T be a quasihyperbolic toral automorphism with topological entropy h. Then there are constants C_2 and $m \geqslant 1$ with*

$$\sum_{|\tau| \leqslant N} \frac{1}{\mathrm{e}^{h|\tau|}} = m \log N + C_2 + \mathrm{O}\left(N^{-1}\right).$$

The coefficient m in the main term is given by

$$m = \int_X \prod_{i=1}^{t} \left(2 - 2\cos(2\pi x_i)\right) \mathrm{d}x_1 \ldots \mathrm{d}x_t,$$

where $X \subset \mathbb{T}^d$ is the closure of $\{(n\theta_1, \ldots, n\theta_t) \mid n \in \mathbb{Z}\}$, and $\mathrm{e}^{\pm 2\pi \mathrm{i}\theta_1}, \ldots, \mathrm{e}^{\pm 2\pi \mathrm{i}\theta_t}$ are the eigenvalues with unit modulus of the matrix defining T.

As we will see in Example 3, the quantity m appearing in Theorem 1 takes on a wide range of values. In particular, m may be much larger, or much smaller, than its generic value 2^t.

Proof. Since T is ergodic,

$$F_T(n) = |\{x \in \mathbb{T}^d \mid T^n(x) = x\}| = |\mathbb{Z}^d/(A_T^n - I)\mathbb{Z}^d| = \prod_{i=1}^{d} |\lambda_i^n - 1|,$$

so

$$O_T(n) = \frac{1}{n} \sum_{m|n} \mu(n/m) \prod_{i=1}^{d} |\lambda_i^m - 1|.$$

Write $\Lambda = \prod_{i=1}^{s} \lambda_i$ (so the topological entropy of T is $\log|\Lambda|$) and

$$\kappa = \min\{|\lambda_s|, |\lambda_{s+2t+1}|^{-1}\} > 1.$$

The eigenvalues of unit modulus contribute nothing to the topological entropy, but multiply the approximation $|\Lambda|^n$ to $F_T(n)$ by an almost-periodic factor bounded above by 2^{2t} and bounded below by A/n^B for some $A, B > 0$, by Baker's theorem (see [3, Ch. 3] for this argument).

Lemma 2. $\left| F_T(n) - |\Lambda|^n \prod_{i=s+1}^{s+2t} |\lambda_i^n - 1| \right| \cdot |\Lambda|^{-n} = \mathrm{O}(\kappa^{-n}).$

Proof. We have

$$\prod_{i=1}^{d}(\lambda_i^n - 1) = \underbrace{\prod_{i=1}^{s}(\lambda_i^n - 1)}_{U_n} \underbrace{\prod_{i=s+1}^{s+2t}(\lambda_i^n - 1)}_{V_n} \underbrace{\prod_{i=2t+s+1}^{d}(\lambda_i^n - 1)}_{W_n}, \tag{2}$$

where U_n is equal to the sum of Λ^n and $(2^s - 1)$ terms comprising products of eigenvalues, each no larger than $\kappa^{-n}|\Lambda|^n$ in modulus, W_n is equal to the sum of $(-1)^{d-s}$ and $2^{d-2t-s} - 1$ terms bounded above in absolute value by κ^{-n}, and $|V_n| \leqslant 2^{2t}$. It follows that

$$\begin{aligned}\frac{\left|\prod_{i=1}^{d}(\lambda_i^n - 1) - (-1)^{d-s}\Lambda^n \prod_{i=s+1}^{s+2t}(\lambda_i^n - 1)\right|}{|\Lambda|^n} &= \frac{\left|V_n\left(U_n W_n - (-1)^{d-s}\Lambda^n\right)\right|}{|\Lambda|^n} \\ &= \frac{\left|V_n\left(\Lambda^n + \mathrm{O}\left(\Lambda^n/\kappa^n\right) - \Lambda^n\right)\right|}{|\Lambda|^n} \\ &= \mathrm{O}(\kappa^{-n}).\end{aligned}$$

The statement of the lemma follows by the reverse triangle inequality. □

Now

$$M_T(N) = \sum_{n=1}^{N} \frac{1}{n|\Lambda|^n}\left(F_T(n) + \sum_{d|n, d<n} \mu\left(\tfrac{n}{d}\right) F_T(d)\right)$$

and

$$\left|\sum_{n=N}^{\infty} \frac{1}{n|\Lambda|^n} \sum_{d|n,d<n} \mu\left(\tfrac{n}{d}\right) F_T(d)\right| \leqslant \sum_{n=N}^{\infty} \frac{1}{n} \cdot n \cdot \mathrm{O}(|\Lambda|^{-n/2}) = \mathrm{O}\left(|\Lambda|^{-N/2}\right),$$

so there is a constant C_3 for which

$$\left|\sum_{n=1}^{N} \frac{1}{n|\Lambda|^n} \sum_{d|n,d<n} \mu\left(\tfrac{n}{d}\right) F_T(d) - C_3\right| = \mathrm{O}\left(|\Lambda|^{-N/2}\right).$$

Therefore, by Lemma 2 and using the notation from (2),

$$M_T(N) = \sum_{n=1}^{N} \frac{1}{n}\left(V_n + \mathrm{O}\left(\kappa^{-n}\right)\right) + C_3 + \mathrm{O}\left(|\Lambda|^{-N/2}\right).$$

Clearly there is a constant C_4 for which

$$\left|\sum_{n=1}^{N} \frac{1}{n} \mathrm{O}\left(\kappa^{-n}\right) - C_4\right| = \mathrm{O}\left(\kappa^{-N}\right), \tag{3}$$

so by (2) and (3),

$$M_T(N) = \sum_{n=1}^{N} \frac{1}{n} V_n + C_3 + C_4 + \mathrm{O}(R^{-N}) \tag{4}$$

where $R = \min\{\kappa, |\Lambda|^{1/2}\}$. Since the complex eigenvalues appear in conjugate pairs we may arrange that $\lambda_{i+t} = \bar{\lambda}_i$ for $s+1 \leqslant i \leqslant s+t$, and then

$$|\lambda_i - 1||\lambda_{i+t} - 1| = (\lambda_i - 1)(\lambda_{i+t} - 1).$$

It follows that $V_n = \prod_{i=s+1}^{s+2t}(\lambda_i^n - 1)$. Put

$$\Omega = \left\{\prod_{i\in I} \lambda_i \mid I \subseteq \{s+1, \ldots, s+2t\}\right\},$$

write

$$\mathcal{I}(\omega) = \{I \subset \{s+1, \dots, s+2t\} \mid \prod_{i \in I} \lambda_i = \omega\},$$

$$K(\omega) = \sum_{I \in \mathcal{I}(\omega)} (-1)^{|I|},$$

and let $m = K(1)$ (notice that $\mathcal{I}(\omega) = \emptyset$ unless $\omega \in \Omega$). Then $V_n = \sum_{\omega \in \Omega} K(\omega)\omega^n$ so, by (4),

$$\begin{aligned} M_T(N) &= \sum_{n=1}^{N} \frac{1}{n} \sum_{\omega \in \Omega} K(\omega)\omega^n + C_5 + \mathrm{O}\left(R^{-N}\right) \\ &= m \sum_{n=1}^{N} \frac{1}{n} + \sum_{\omega \in \Omega \setminus \{1\}} K(\omega) \sum_{n=1}^{N} \frac{\omega^n}{n} + C_5 + \mathrm{O}\left(R^{-N}\right) \\ &= m \log N - \sum_{\omega \in \Omega \setminus \{1\}} K(\omega) \log(1-\omega) + C_6 + \mathrm{O}(N^{-1}), \end{aligned}$$

since $\sum_{n=1}^{N} \frac{1}{n} = \log N + \gamma + \mathrm{O}(N^{-1})$, and$\sum_{n=1}^{N} \frac{\omega^n}{n} = -\log(1-\omega) + \mathrm{O}(N^{-1})$ for $\omega \neq 1$ by the Abel continuity theorem and partial summation.

If the eigenvalues of modulus one are $\mathrm{e}^{\pm 2\pi \mathrm{i}\theta_1}, \dots, \mathrm{e}^{\pm 2\pi \mathrm{i}\theta_t}$ then

$$V_n = \prod_{i=1}^{t} (1 - \mathrm{e}^{2\pi \mathrm{i}\theta_i n})(1 - \mathrm{e}^{-2\pi \mathrm{i}\theta_i n}) = \prod_{i=1}^{t} \left(2 - 2\cos(2\pi\theta_i n)\right).$$

Let $X \subset \mathbb{T}^t$ be the closure of $\{(n\theta_1, \dots, n\theta_t) \mid n \in \mathbb{Z}\}$, so that by the Kronecker–Weyl lemma we have

$$\frac{1}{N} \sum_{n=1}^{N} \prod_{i=1}^{t} \left(2 - 2\cos(2\pi\theta_i n)\right) \longrightarrow \int_X \prod_{i=1}^{t} \left(2 - 2\cos(2\pi x_i)\right) \mathrm{d}x_1 \dots \mathrm{d}x_t$$

as $N \to \infty$. Then, by partial summation,

$$\begin{aligned} \sum_{n=1}^{N} \frac{1}{n} V_n &= \sum_{n=1}^{N} \left(\frac{1}{n} - \frac{1}{n+1}\right) \sum_{m=1}^{n} V_m + \frac{1}{N+1} \sum_{m=1}^{N} V_m \\ &\sim \left(\int_X \prod_{i=1}^{t} \left(2 - 2\cos(2\pi x_i)\right) \mathrm{d}x_1 \dots \mathrm{d}x_t\right) \log N, \end{aligned}$$

so that m has the form stated. □

The exact value of m is determined by the structure of the group X, which in turn is governed by additive relations among the arguments of the eigenvalues of unit modulus. Here are some illustrative examples.

Example 3. (a) If all the arguments θ_i are independent over $\mathbb{Q}$ (the generic case), then $X = \mathbb{T}^t$, so

$$m = \int_0^1 \cdots \int_0^1 \prod_{i=1}^{t} \left(2 - 2\cos(2\pi x_i)\right) \mathrm{d}x_1 \dots \mathrm{d}x_t = \left(\int_0^1 \left(2 - 2\cos(2\pi x_1)\right) \mathrm{d}x_1\right)^t = 2^t.$$

(b) A simple example with $m > 2^t$ is the following. Let T_2 be the automorphism of $\mathbb{T}^8$ defined by the matrix $A \oplus A$, where

$$A = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 8 \\ 0 & 1 & 0 & -6 \\ 0 & 0 & 1 & 8 \end{pmatrix}. \tag{5}$$

Here X is a diagonally embedded circle, and

$$\begin{aligned} m &= \iint_{\{x_1=x_2\}} \prod_{j=1}^{2} (2 - 2\cos(2\pi j x_j)) \, \mathrm{d}x_1 \, \mathrm{d}x_2 \\ &= \int_0^1 (2 - 2\cos(2\pi x))^2 \, \mathrm{d}x = 6 > 2^2. \end{aligned}$$

Extending this example, let T_n be the automorphism of $\mathbb{T}^{4n}$ defined by the matrix $A \oplus \cdots \oplus A$ (n terms). The matrix corresponding to T_n has $2n$ eigenvalues with modulus one (comprising two conjugate eigenvalues with multiplicity n). Then X is again a diagonally embedded circle, and

$$m = \int_0^1 (2 - 2\cos(2\pi x))^t \, \mathrm{d}x = \frac{(2t)!}{(t!)^2} \sim \frac{2^{2t}}{\sqrt{\pi t}}$$

by Stirling's formula. This is much larger than 2^t, reflecting the density of the syndetic set on which the almost-periodic factor is close to 2^{2t}. Indeed, this example shows that $\frac{m}{2^t}$ may be arbitrarily large.

(c) A simple example with $m < 2^t$ is the following. Let S be the automorphism of $\mathbb{T}^{12}$ defined by the matrix $A \oplus A^2 \oplus A^3$, with A as in (5). Again X is a diagonally embedded circle, and

$$\begin{aligned} m &= \iiint_{\{x_1=x_2=x_3\}} \prod_{j=1}^{3} (2 - 2\cos(2\pi j x_j)) \, \mathrm{d}x_1 \mathrm{d}x_2 \, \mathrm{d}x_3 \\ &= \int_0^1 (2 - 2\cos(2\pi x))(2 - 2\cos(4\pi x))(2 - 2\cos(6\pi x)) \, \mathrm{d}x = 6 < 2^3. \end{aligned}$$

Extending this example, the value of m for the automorphism of $\mathbb{T}^{4t}$ defined by the matrix $A \oplus A^2 \oplus \cdots \oplus A^t$ as t varies gives the sequence

$$2, 4, 6, 10, 12, 20, 24, 34, 44, 64, 78, 116, 148, 208, 286, 410, 556, 808, 1120, 1620, \ldots$$

(we thank Paul Hammerton for computing these numbers). This sequence, entry A133871 in the Encyclopedia of Integer Sequences [7], does not seem to be readily related to other combinatorial sequences.

(d) Generalizing the example in (c), for any sequence (a_n) of natural numbers, we could look at the automorphisms S_n of $\mathbb{T}^{4n}$ defined by the matrices $\bigoplus_{k=1}^n A^{a_k}$, with A as in (5). In order to make m small, we need a "sum-heavy" sequence, that is, one with many three-term linear relations of the form $a_i + a_j = a_k$. More precisely, one would like many linear relations with an odd number of terms, and few with an even number of terms. Constructing such sequences, and understanding how dense they may be, seems to be difficult.

Taking (a_n) to be the sequence whose first eight terms are $1, 2, 3, 5, 7, 8, 11, 13$ and whose subsequent terms are defined by the recurrence $a_{n+8} = 100a_n$, we find that the automorphism S_{8n} of $\mathbb{T}^{32n}$ has $m = 2^{4n} = 2^{t/2}$. Thus $\frac{m}{2^t}$ may be arbitrarily small.

We close with some remarks.

(1) In the quasihyperbolic case the $\mathrm{O}(1/N)$ term is oscillatory, so no improvement of the asymptotic in terms of a monotonic function is possible. The extent to which the exponential dominance of the entropy term fails in this setting is revealed by the following. Let $F_T(n)$ denote the number of points fixed by the automorphism T^n. On the one hand, Baker's theorem implies that $F_T(n)^{1/n} \to e^h$ as $n \to \infty$. On the other hand Dirichlet's theorem shows that $F_T(n+1)/F_T(n)$ does not converge (see [1, Th. 6.3]).

(2) The formula for m in the statement of [4, Th. 1] is incorrect in a minor way; as stated in [4, Rem. 2] and as illustrated in the examples above, m should be $K(1)$, which is not necessarily the same as 2^t.

(3) The proof of Theorem 1 also gives an elementary proof of the asymptotics in the hyperbolic case: in the notation of the proof, $V_n = 1$ so $m = 1$. The Euler-MacLaurin summation formula gives an asymptotic of the shape

$$\sum_{|\tau| \leqslant N} \frac{1}{\mathrm{e}^{h|\tau|}} = \log N + C_2 + \sum_{r=0}^{k-1} \frac{B_{r+1}}{(r+1)N^{r+1}} + \mathrm{O}\left(N^{-(k+1)}\right),$$

where $B_1 = -\frac{1}{2}$, $B_2 = \frac{1}{6}, \dots$ are the Bernoulli numbers, for any $k \geqslant 1$.

References

[1] V. Chothi, G. Everest, and T. Ward. S-integer dynamical systems: periodic points. *J. Reine Angew. Math.* **489**, 99–132, 1997.

[2] G. Everest, R. Miles, S. Stevens, and T. Ward. Orbit-counting in non-hyperbolic dynamical systems. *J. Reine Angew. Math.* **608**, 155–182, 2007.

[3] G. Everest and T. Ward. *Heights of polynomials and entropy in algebraic dynamics.* Springer-Verlag London (1999).

[4] Mohd. Salmi Md. Noorani. Mertens theorem and closed orbits of ergodic toral automorphisms. *Bull. Malaysian Math. Soc. (2)* **22**(2), 127–133, 1999.

[5] A. Pakapongpun and T. Ward. Functorial orbit counting *J. Integer Sequences*, **12**, Article 09.2.4, 2009.

[6] R. Sharp. An analogue of Mertens' theorem for closed orbits of Axiom A flows. *Bol. Soc. Brasil. Mat. (N.S.)*, **21**(2), 205–229, 1991.

[7] N. J. A. Sloane. An on-line version of the encyclopedia of integer sequences. *Electron. J. Combin.* **1**:Feature 1, approx. 5 pp., 1994. `www.research.att.com/~njas/sequences/`.

[8] S. Waddington. The prime orbit theorem for quasihyperbolic toral automorphisms. *Monatsh. Math.* **112**, no. 3, 235–248, 1991.

(SJ) Department of Mathematics, 123 Mittraphab Road, Khon Kaen University 40002,Thailand

(SS & TW) School of Mathematics, University of East Anglia, Norwich NR4 7TJ, UK