Horocycle Flow at Prime Times

Peter Sarnak
Mahler Lectures 2011

Rotation of the Circle

A very simple (but by no means trivial) dynamical system is the rotation (or more generally translation in a compact abelian group) $\alpha \in \mathbb{R}$ fixed, $T_\alpha : X \to X$, $X = \mathbb{R}/\mathbb{Z}$

$$\begin{aligned} T_\alpha(x) &= x + \alpha, \\ T^n(x) &= x + n\alpha, \quad n \in \mathbb{Z}. \end{aligned}$$

- If $\alpha \in \mathbb{Q}$, T is periodic and the orbits are finite.
- If $\alpha \notin \mathbb{Q}$, orbits are dense and equidistributed. For every x and continuous f,

$$\frac{1}{N} \sum_{n=1}^{N} f(T^n x) \to \int_0^1 f(t)\, dt.$$

(X, T) is uniquely ergodic, distal, zero entropy, ...

What if we restrict the n's to be primes?

Theorem (I. M. Vinogradov 1937)

If $\alpha \notin \mathbb{Q}$, then $p\alpha \pmod 1$, $p \le N$ prime is dense in $\mathbb{R}/\mathbb{Z}$ and equidistributed; $I \subset [0,1)$

$$\frac{\#\{p \le N : p\alpha \in I\}}{\#\{p \le N\}} \to I \quad \text{as } N \to \infty.$$

This was a major breakthrough and in quantitative form it allowed him when combined with the Hardy–Littlewood circle method to prove his famous "every large odd number is a sum of three primes". More generally to solve the linear equation

$$ax_1 + bx_2 + cx_3 = d,$$

x_j primes.

New idea was a sieve via bilinear sums.
Thinking generally of (X, T) a dynamical system; sums over primes can be addressed by

$$(I) \quad \sum_{n \leq N} f(T^{dn}x),$$

with d as large as possible, $d \leq N^{\alpha}$; α is called the level; i.e. sums on progressions.

$$(II) \quad \sum_{n \leq N} f(T^{d_1 n}x) f(T^{d_2 n}x)$$

with $\max(d_1, d_2) \leq N^{\beta}$, level β.

- In dynamical terms, the first sum is associated with (X, T^d), the second with the joining (X, T^{d_1}) with (X, T^{d_2}).
- For $(\mathbb{R}/\mathbb{Z}, T_\alpha)$, I and II are geometric series!

B. Green and T. Tao (Adelaide) have extended this to nilmanifolds. This was needed for their generalization of Vinogradov to systems of linear equations in primes, as long as there are two more variables than equations. The critical role played by nilmanifolds in this context comes from Furstenberg.

N a connected Lie group, e.g.

$$N = \begin{bmatrix} 1 & x_1 & x_2 \\ 0 & 1 & x_3 \\ 0 & 0 & 1 \end{bmatrix}, \quad \Gamma \text{ a lattice in } N, \text{ e.g. } N(\mathbb{Z}) \text{ above,}$$

$$X = \Gamma \setminus N, \quad T_\alpha(\Gamma x) = \Gamma x \alpha \quad \text{with } \alpha \in N.$$

Theorem (Green–Tao 2009)

If $\{T^n x : n \in \mathbb{Z}\}$ is dense in X, then so is $\{T^p x : p \text{ prime}\}$ and both are equidistributed with respect to dg when ordered in the obvious way.

- Their proof involves a detailed study of pieces of orbits in such a nilmanifold eventually reducing it by the derived central series to the torus case.
- The dynamics here is still distal, zero entropy, ...

Homogeneous Dynamics

$$X = \Gamma \setminus G, \quad G \text{ a semisimple Lie group}$$
$$\Gamma \text{ a lattice (discrete, } \mathrm{Vol}(\Gamma \setminus G) < \infty)$$
$$T_\alpha(\Gamma x) = \Gamma x \alpha.$$

Why examine these?

(a) They are locally symmetric Riemannian spaces.

(b) The theory of automorphic forms is built on them.

(c) For Γ "arithmetic", the spaces $\Gamma \setminus G$ are often moduli spaces. For example, $\Gamma = \mathrm{SL}_n(\mathbb{Z})$, $G = \mathrm{SL}_n(\mathbb{R})$, then $\Gamma \setminus G$ is the space of n-dimensional lattices, central in the geometry of numbers (e.g. Mahler compactness!).

(d) $\mathrm{Sp}(2g, \mathbb{Z}) \setminus \mathrm{Sp}(2g, \mathbb{R})$ is the moduli space of principally polarized abelian varieties and there are many diophantine problems connected with distribution of points in these.

Aside: One such is the André–Oort conjecture about "CM points" was recently solved by J. Pila (originally from Melbourne) and his proof uses some transcendence ideas of Mahler.

$(\Gamma \setminus G, T_\alpha)$ dynamics

- If there is to be a generalization of Vinogradov to this setting, one can show that the system must be of zero entropy. That is, α is ad-quasi-unipotent (eigenvalues are of modulus 1).
- For α unipotent, Ratner's rigidity theorem asserts that the orbit closure of any T_α orbit is structured (algebraic). The same is of quasi-unipotent α (Starkov).
- The dynamics in these cases, while of zero entropy, is much more complex — it is mixing of all orders.

We consider the most basic case of the modular quotient.

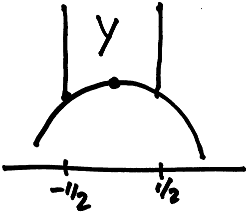
The Modular Group:

$$G = \mathrm{SL}_2(\mathbb{R}), \quad \Gamma = \mathrm{SL}_2(\mathbb{Z}).$$

If $K = \mathrm{SO}(2)$, then $G/K \cong \mathbb{H}$, the upper half plane.

$$[G \text{ acts by } z \mapsto \frac{az+b}{cz+d}]$$

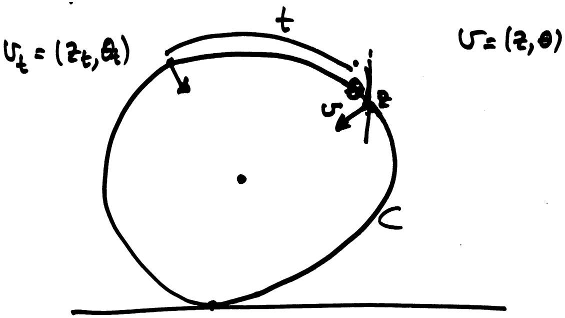
$Y = \Gamma \setminus \mathbb{H}$.



$\alpha = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ the standard unipotent.

The action of α^n or more generally $\begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}$ with $t \in \mathbb{R}$ is the horocycle flow:

$$\Gamma \setminus G \cong \left\{(z,\theta), -\frac{\pi}{2} < \theta < \frac{\pi}{2}\right\}.$$



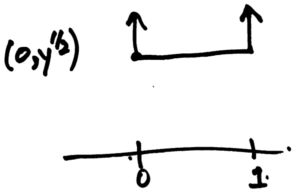
There is a unique horocycle as indicated with $v \perp$ to C at z. Now flow along the horocycle a distance t to arrive at a new $z = z_t$ and $\theta = \theta_t$. This is the action of $\begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}$ by multiplication on the right.

If $H_N = \begin{bmatrix} N^{-1/2} & 0 \\ 0 & N^{1/2} \end{bmatrix}$, $N \geq 1$ in $\mathbb{Z}$, then

$$\Gamma H_N \begin{bmatrix} 1 & N \\ 0 & 1 \end{bmatrix} = \Gamma H_N$$

so that H_N is a periodic point of period N for T_α.

More generally if $\xi = \begin{bmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{bmatrix}$, $y > 0$ $(\theta = 0)$ then



$z \to z+1 \in \Gamma$

is a closed horocycle of period $1/y$.

Hedlund–Dani Theorem (pre-Ratner)

$X = \mathrm{SL}_2(\mathbb{Z}) \setminus \mathrm{SL}_2(\mathbb{R})$, $\alpha = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, $\xi \in X$. Then we have one of

(i) *ξ is a periodic point for T_α.*

(ii) *The closure of the orbit of ξ is a closed horocycle.*

(iii) *The closure of the orbit of ξ is all of X and the orbit is equidistributed with respect to dg.*

What about the orbits at primes?

(i) Dirichlet's theorem on primes in progressions.

(ii) Vinogradov.

For ξ as in (iii) sometimes called generic, let

$$\nu(\xi, N) = \frac{1}{\pi(N)} \sum_{p \le N} \delta_{T^p \xi}$$

where $\pi(N)$ is the number of primes $\le N$, δ_η is a point mass at η. $\nu(\xi, N)$ is a probability measure on X and the conjecture is that these converge to dg as $N \to \infty$.

Theorem (Ubis–S. 2011)

Let ν_ξ be a weak limit of $\nu(\xi, N)$. Then ν_ξ is absolutely continuous with respect to dg; in fact,

$$\nu_\xi \le 10 dg.$$

Corollary

If U is open (nice), $\mathrm{Vol}_{dg}(U) > \frac{9}{10}$, then the orbit of ξ at primes enters U for a positive proportion of primes.

For $\xi = H_N$ we can do better.

$$\nu_N := \frac{1}{\pi(N)} \sum_{p \leq N} \delta_{T^p(H_N)}.$$

Theorem

Assume the Ramanujan–Selberg conjectures for $\mathrm{GL}_2/\mathbb{Q}$*. Then any limit of the* ν_N*'s, call it* ν*, satisfies*

$$\frac{dg}{5} \leq \nu \leq \frac{9}{5} dg.$$

In particular, the images $T^p H_N$*,* $p \leq N$*,* p *prime are dense in* X *as* $N \to \infty$*.*

We can resolve a related problem for $(\Gamma \setminus G, T_\alpha)$.

Consider

$$\sum_{n \le N} \mu(n) f(T^n x) \tag{1}$$

for $x \in X$ fixed and $f \in C(X)$. Here $\mu(n)$ is the Möbius function

$$\mu(n) = \begin{cases} (-1)^t & \text{if } n \text{ is a product of } t \text{ distinct primes,} \\ 0 & \text{otherwise.} \end{cases}$$

Estimates for (1) of the form $1/(\log N)^A$ are more or less equivalent to sums of $f(T^n x)$ over n prime (with rates).

A fundamental question is whether (1) is $o(N)$, "μ is disjoint or orthogonal to (X, T)".

This is so for

$$(X, T) = (\mathbb{R}/\mathbb{Z}, T_\alpha), \qquad \text{Vinogradov/Davenport},$$
$$(X, T) = (\Gamma \setminus N, T_\alpha), \quad N \text{ nilpotent Lie group}, \qquad \text{Green and Tao}.$$

Theorem (Bourgain–S. 2011)

Let $G = \mathrm{SL}_2(\mathbb{R})$, Γ a lattice in G, $X = \Gamma \setminus G$, $T : X \to X$ a unipotent transformation. Then for $x \in X$, $f \in C(X)$,

$$\sum_{n \leq N} \mu(n) f(T^n x) = o(N),$$

i.e. Möbius is disjoint from a horocycle flow.

Ideas of Proofs:

For $X = \mathrm{SL}_2(\mathbb{Z}) \setminus \mathrm{SL}_2(\mathbb{R})$, $T_\alpha, \alpha = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$

- The periodic orbits are dense in X.
- We approximate (quantitatively) any long horocycle (discrete) flow by (suitably long) pieces of periodic horocycles. This is similar to Dirichlet's theorem giving approximations to reals by rationals.

For $\xi = \begin{bmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{bmatrix}$ a periodic horocycle of period $1/y$, the sums over the orbit are controlled by period integrals of the form, $h \in \mathbb{Z}$, $|h| \leq y^{-1}$,

$$I = \int_0^1 f\left(\begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix}\begin{bmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{bmatrix}\right) e(-hx)\, dx$$

for $f \in C^\infty(X)$ and $\int_X f\, dg = 0$.

Decomposing f spectrally into automorphic forms shows that

(1) If $h = 0$, I measures the speed of equidistribution of the closed horocycle of length $1/y$. It is known this is closely connected to the Riemann Hypothesis!
[It is $O(y^{3/4})$ iff RH is true]

(2) If $h \neq 0$, then the best one can hope for is square-root cancellation.

If $0 \leq \theta \leq 1/2$ is an admissible bound towards Ramanujan $\mathrm{GL}_2/\mathbb{Q}$,

$$\theta = \frac{1}{2} \text{is trivial}$$
$$\theta = 0 \text{is the conjecture}$$
$$\theta = \frac{7}{64} \text{is the record (Kim–S. 2003),}$$

then

$$I \ll y^{1/2-\theta} \|f\|_{W^6}.$$

W^k is the Sobolev k-norm.

This is the source of controlling the level of distribution for the type I sums for the horocycle flow. It gives a level of essentially $1/2$ and also yields when properly stated an effective Dani Theorem.

Note: For (X, T) a unipotent flow on $\Gamma \setminus \mathrm{SL}_2(\mathbb{R})$ which is compact (uniquely ergodic), an effective equidistribution for the continuous horocycle flow is due to M. Burger ('88) and A. Venkatesh (Perth) for the discrete case (2007).

The difficulty to get primes (fully) for $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ are the type *II* sums. This requires the analysis of the joinings of (X, T^{d_1}) and (X, T^{d_2}) which is a one-parameter unipotent orbit in $\Gamma \setminus \mathrm{SL}_2(\mathbb{R}) \times \Gamma \setminus \mathrm{SL}_2(\mathbb{R})$.

There is no apparent spectral treatment of this. The only treatment is that pioneered by Ratner (already in 1983 for this case) and it is ineffective.

- For the H_N (periodic) orbit, we are able to handle this joining using the theory of "shifted convolutions".
- For the qualitative theorem with Bourgain (no rates), we actually combine a finite version of Vinogradov with Ratner's methods. Critical is the computation of a certain subgroup of $\mathbb{Q}^*$ connected with joining (X, T^{d_1}), (X, T^{d_2}).

Some references:

- J. Bourgain and P. Sarnak, "Disjointness of Möbius from horocycle flows", preprint, 2011.
- M. Burger, Duke M. J. **61**(3) (1990), 779–803.
- S. Dani, Erg. Th. and Dynam. Sys. **2** (1982), 139–158.
- B. Green and T. Tao, "Linear equations in primes", Ann. Math. **171** (2010), 1753–1850.
- B. Green and T. Tao, "The Möbius function is orthogonal to nilsequences", to appear in Ann. Math.
- J. Pila, Ann. Math. **173**(3) (2011), 1779–1840.
- M. Ratner, Ann. Math. **134** (1991), 545–607.
- A. Ubis and P. Sarnak, "The horocycle flow at prime times", 2011.

I. M. Vinogradov, Recueiv Math. **8** (1937), 313–320.