Super convergence of ergodic averages for quasiperiodic orbits

Suddhasattwa Das \(^1\) and James A Yorke \(^2\)

\(^1\) Department of Mathematics, University of Maryland, College Park, MD, United States of America
\(^2\) Institute for Physical Science and Technology, University of Maryland, College Park, MD, United States of America

E-mail: dass@cims.nyu.edu

Received 6 April 2016, revised 24 October 2017
Accepted for publication 10 November 2017
Published 10 January 2018

Recommended by Dr Tere M Seara

Abstract

The Birkhoff ergodic theorem asserts that time averages of a function evaluated along a trajectory of length \(N\) converge to the space average, the integral of \(f\), as \(N \to \infty\), for ergodic dynamical systems. But that convergence can be slow. Instead of uniform averages that assign equal weights to points along the trajectory, we use an average with a non-uniform distribution of weights, weighting the early and late points of the trajectory much less than those near the midpoint \(N/2\). We show that in quasiperiodic dynamical systems, our weighted averages converge far faster provided \(f\) is sufficiently differentiable. This result can be applied to obtain efficient numerical computation of rotation numbers, invariant densities and conjugacies of quasiperiodic systems.

Keywords: quasiperiodicity, Birkhoff ergodic theorem, rotation number
Mathematics Subject Classification numbers: 37A25, 37A05, 37A45

(Some figures may appear in colour only in the online journal)

1. Introduction

1.1. Quasiperiodicity

The concept of quasiperiodicity is tied to a simple class of maps called pure rotations. For every \(\rho \in (0, 1)^d\) there is a (pure) rotation on the \(d\)-dimensional torus \(\mathbb{T}^d\), i.e. the map \(T_\rho\) defined as

\[ T_\rho : \theta \mapsto \theta + \rho \mod 1 \text{ in each coordinate.} \]
This map acts on each coordinate $\theta_j$ by rotating it by some angle $\rho_j$. We call the $\rho_j$’s ‘rotation numbers’. We say $\rho = (\rho_1, \ldots, \rho_d) \in \mathbb{R}^d$ is irrational if there are no integers $k$ for which $k_1 \rho_1 + \cdots + k_d \rho_d \in \mathbb{Z}$, except when all $k_i$ are zero. In particular, this implies that each $\rho_j$ must be irrational. The simplest example of $d$-dimensional quasiperiodicity is a (pure) irrational rotation, that is, a rotation by an irrational vector $\rho$.

A map $T : X \to X$ is said to be $d$-dimensionally $C^m$ quasiperiodic on a set $X_0 \subseteq X$ for some $d \in \mathbb{N}$ if there is a $C^m$-diffeomorphism $h : T^d \to X_0$, such that,

$$T(h(\theta)) = h(T_\rho(\theta))$$

where $T_\rho$ is an irrational rotation. In this case, $h$ is a conjugacy of $T$ to $T_\rho$. It has been conjectured in [1] that for typical, randomly chosen physical systems, it is one of apparently only three kinds of maximal recurrent sets that are likely to be present, the other two being periodic orbits and chaotic sets.

1.2. An example of a quasiperiodic trajectory of a flow

To see quasiperiodicity, one needs only look at the Moon [2]. More precisely, there are quasiperiodic trajectories when modeling the Sun–Earth–Moon system as a circular restricted three-body problem in $\mathbb{R}^6$ (position + velocity) with Newtonian gravitation. Let $q_1, q_2, q_3$ be the positional coordinates where Earth–Sun rotation is in the $q_1 - q_2$ plane. The Moon is not restricted to that plane. Using rotating coordinates in which the Sun and Earth are fixed on the $q_1$ axis, we can get an excellent approximation of the Moon’s orbit as a quasiperiodic flow on $T^3 \subset \mathbb{R}^6$. ‘Restricted’ means that the Moon’s mass is modeled as negligible. Note that the masses are modeled as points so tidal forces are not included, nor is the influence of the other planets.

1.3. An example of a quasiperiodic trajectory of a map

Following Poincaré, we also can get the Poincaré return map $F$ by fixing a value for the Hamiltonian of the system and by taking trajectory points on the $q_1 - q_3$ plane, where $q_2 = 0$ and $dq_2/dt > 0$. The moon passes through this plane about once a month. Here the flow on $T^3$ is reduced to a map on $T^2$. We note that the map $F$ is real analytic.

1.4. The conjugacy problem

Given a trajectory $(x_n)$ (for $n = 0, 1, \ldots, \infty$) of a map $T : X \to X$ as above, we say that $(x_n)$ is $d$-dimensionally quasiperiodic if $(x_n)$ is in a set $X_0 \subset X$ and $T$ is $d$-dimensionally quasiperiodic on $X_0$.

How do we demonstrate that a trajectory $(x_n)$ is $d$-dimensionally quasiperiodic? From the definition, it is sufficient to find the conjugacy map $h : T^d \to \mathbb{R}^d$ and find the rotation numbers $\rho = (\rho_1, \ldots, \rho_d)$ for which $x_n = h(n \rho_1, \cdots, n \rho_d)$. If $h$ is found, then we will have shown that the pure rotation with rotation numbers $\rho$ is conjugate to $F$ restricted to $X_0 = h(T^d)$. We call the problem of finding $\rho$ and $h$—knowing only the trajectory $(x_n)$—the conjugacy problem for a quasiperiodic trajectory. Our approach to finding $h$ is to compute the Fourier series for $h$ using our weighted Birkhoff average $WB_N$, which we describe below.
1.5. The Birkhoff ergodic theorem

Let $T : X \to X$ be a map on a topological space $X$ with a probability measure $\mu$ for which $T$ is invariant. Given a point $x$ in $X$ and a real- or vector-valued function $f$ on $X$, we will refer to a long-time average of the form

$$ B_N(f) := \frac{1}{N} \sum_{n=0}^{N-1} f(T^n(x)), $$ \hfill (3) 

as a Birkhoff average. Throughout this paper, $f : X \to E$ where $E$ is a finite-dimensional real vector space. The limits of such sequences frequently occur in dynamical systems. If $\mu$ is a probability measure on $X$ and $T$ preserves the probability measure $\mu$ and is ergodic, then the von Neumann ergodic theorem (e.g. see theorem 4.5.2. in [3]) states that for $f \in L^2(X, \mu)$, the Birkhoff average (equation (3)) converges in the $L^1$ norm to the integral $\int_X f d\mu$. The Birkhoff ergodic theorem (see theorem 4.5.5. in [3]) strengthens von Neumann’s theorem and concludes that if $f \in L^1(X, \mu)$, then equation (3) converges to the integral $\int_X f d\mu$ for $\mu$-a.e. point $x \in X$. The Birkhoff average (equation (3)) can be interpreted as an approximation to an integral, but convergence is very slow. For any non-constant $f$, there is a constant $C > 0$ such that the following holds for infinitely many $N$.

$$ \left| \frac{1}{N} \sum_{n=0}^{N} f(T^n(x)) - \int_X f d\mu \right| \geq CN^{-1}. $$

For general ergodic dynamical systems, the rate of convergence of these sums can be arbitrarily slow, as shown in [4]. For many purposes the speed of convergence is irrelevant but it is important for numerical computations.

**Definition.** Let $(a_N)_{N=0}^\infty$ be a sequence in a normed vector space such that $a_N \to b$ as $N \to \infty$. We say $(a_N)$ has super-polynomial convergence to $b$ or super converges to $b$ if for each integer $m > 0$ there is a constant $C_m > 0$ such that

$$ |a_N - b| \leq C_m N^{-m} \text{ for all } m. $$

**Definition.** A function $w : \mathbb{R} \to [0, \infty)$ is said to be a $C^\infty$ bump function if $w$ is $C^\infty$ and the support of $w$ is $[0, 1]$ and $\int_{\mathbb{R}} w(x) dx \neq 0$ and $w$ and all of its derivatives vanish at 0 and 1. When we use a bump function as in equation (5), it is independent of scaling; hence we can assume without loss of generality that $\int_{\mathbb{R}} w(x) dx = 1$. Equation (4) below gives an example of a $C^\infty$ bump function that we will call the exponential weighting,

$$ w(t) = \begin{cases} \exp \left( \frac{-1}{t(1-t)} \right), & \text{for } t \in (0, 1) \\ 0, & \text{for } t \notin (0, 1). \end{cases} \hfill (4) $$

Instead of weighting the terms $f(T^n(x))$ in the average equally, we weight the early and late terms of the set $1, \ldots, N$ much less than the terms with $n \sim N/2$ in the middle. We insert a weighting function $w$ into the Birkhoff average to get what we call a weighted Birkhoff $\langle W_B N \rangle$ average. Let $\mathbb{T}^d$ denote a $d$-dimensional torus. For $X = \mathbb{T}^d$ and a continuous $f$ and for $\phi \in \mathbb{T}^d$, we define

$$ W_B N(f)(x) := \frac{1}{A_N} \sum_{n=0}^{N-1} w\left( \frac{n}{N} \right) f(T^n x), \text{ where } A_N := \sum_{n=0}^{N-1} w\left( \frac{n}{N} \right). \hfill (5) $$
Of course the sum of the terms \( w(\frac{n}{N})/A_N \) is 1. Continuity of \( w \) guarantees that 
\[
\lim_{N \to \infty} A_N/N = \int_0^1 w(t)dt.
\]
We introduced the exponentially weighted Birkhoff average of equation (5) for numerical investigations of quasiperiodic systems in [5]. Motivated by finding how effective our method was numerically, we discovered a simple proof of its super convergence.

1.6. Unique invariant probability measure

Lebesgue probability measure \( \lambda \) on \( \mathbb{T}^d \) is the unique invariant probability measure for any irrational rotation \( T_{\rho} \). A map \( T : X_0 \to X_0 \) therefore has a unique invariant probability measure \( \mu \) which is defined as
\[
\mu(U) := \lambda(h^{-1}(U)), \quad \text{for every Borel set } U \subseteq X_0.
\]

1.7. Diophantine rotations

An irrational vector \( \rho \in \mathbb{R}^d \) is said to be Diophantine if for some \( \beta > 0 \) it is Diophantine of class \( \beta \) (see [7], definition 3.1), which means there exists \( C_\beta > 0 \) such that for every \( k \in \mathbb{Z}^d \), \( k \neq 0 \) and every \( n \in \mathbb{Z} \),
\[
|k \cdot \rho - n| \geq \frac{C_\beta}{\|k\|^{d+\beta}}.
\]
(6)

For every \( \beta > 0 \) the set of Diophantine vectors of class \( \beta \) have full Lebesgue measure in \( \mathbb{R}^d \) (see [7], 4.1). The Diophantine class is crucial in the study of quasiperiodic behavior, for example in [8, 9]. In equation (6) is central to Arnold’s study of small denominators in [10].

**Theorem 1.1.** Let \( X \) be a \( C^\infty \) manifold and \( T : X \to X \) be a \( C^\infty \) map which is \( d \)-dimensionally quasiperiodic on \( X_0 \subseteq X \), with invariant probability measure \( \mu \). Assume \( T \) has a Diophantine rotation vector. Let \( f : X \to E \) be \( C^\infty \), where \( E \) is a finite-dimensional, real vector space. Assume \( w \) is a \( C^\infty \) bump function. Then for each \( x_0 \in X_0 \), the weighted Birkhoff average \( \text{WB}_N f(x_0) \) has super convergence to \( \lim_{N \to \infty} B_N(f)(x_0) = \int_{X_0} f d\mu \). Moreover, the convergence is uniform in \( x_0 \).

Theorem 1.1 is proved as a corollary of theorem 3.1 in section 3.

1.8. Applicability to chaos

Since \( \text{WB}_N(f)(x_0) \) and \( B_N(f)(x_0) \) have the same limit as \( N \to \infty \), the weighted Birkhoff average provides a fast way of computing the limit of a Birkhoff average. If the ergodic process \( T \) is chaotic instead of quasiperiodic, then weighted Birkhoff averages provide no advantage over Birkhoff averages.

1.9. The need for fast convergence

We note that in [5] we get convergence to the limit of our 30-digit numerical precision in all of our one-dimensional (1D) quasiperiodicity studies well before \( N = 10^6 \) iterates. If we used \( B_N \), we would expect to need about \( N \sim 10^{30} \), higher by a factor of about \( 10^{23} \). Hence if we can compute \( 10^6 \) iterates per second, \( \text{WB}_N \) would require one second while \( B_N \) would require one billion billion years.
1.10. Related methods

The accelerated convergence of weighted ergodic sums have been discussed in [11–13], but
without conclusions of specific rates of convergence. In [14], a convergence rate of \(O(N^{-\alpha})\),
\((0 < \alpha < 1)\), was obtained for functionals in \(L^{2+\epsilon}\) for a certain choice of weights.

In [15], Luque and Villanueva develop methods for obtaining rotation numbers by taking
repeated averages of averages of a quasiperiodic signal. By taking \(p\) nested averages, their
method converges to the rotation number with an error bounded by \(C_p N^{-p}\), where \(C_p\) is a
constant and \(N \geq 2^p\). Their method of computation depends on the choice of convergence
rate \(p\) and as \(p\) increases the computational complexity increases for fixed \(N\). Also, their com-
putation time \(T(p, N)\) obeys \(T(p, N)/N \to \infty\) as \(p \to \infty\). In comparison, computation time
for our weighted Birkhoff average is simply proportional to \(N\) since it requires an average of
\(N\) numbers, including the evaluation of \(w\) at \(N\) points.

Jacques Laskar uses a \(C^1\) weighting function (a Hanning window filter), namely what is
effectively the function \(\sin^2(\pi t)\) on \((0, 1)\) except that he uses a two-sided version on \((-1, 1)\),
in investigating almost periodic perturbations of almost periodic flows in several papers
[16–19]. But he also states in an appendix (remark 2, p 146) to [16]) that one can consider
what is essentially our weighted Birkhoff average, equation (5). He reports that it has excel-
lent asymptotic properties, but we have not found any situations where he implemented it or
demonstrated its convergence properties.

1.11. A companion paper

This project has two components. The rigorous foundation for our methods are reported here
and the numerical work is reported in [5]; the effectiveness of the theoretical work is dem-
onstrated through the numerical results; that is, we get really fast convergence in practice.
The reliability of the numerical work in [5] is correspondingly based on the rigorous results
presented here so that when we know a sequence is guaranteed to converge and then we get
numerical convergence to the available precision, it suggests strongly that the results are cor-
rect. In that paper we present four examples:

- The restricted circular three-body problem in \(\mathbb{R}^2\) with 1D quasiperiodic curves (dimen-
sion \(d = 1\)).
- A periodically forced van der Pol oscillator with an attracting quasiperiodic curve.
- The Chirikov-Taylor standard map (area-preserving map on \(\mathbb{T}^2\)) with quasiperiodic closed
curves.
- A map on a torus \(\mathbb{T}^2\) exhibiting \(d = 2\)-dimensional quasiperiodicity.

The first three of these investigations are for 1D quasiperiodic curves in dimension 2, while
the fourth is for two-dimensional quasiperiodic dynamics on \(\mathbb{T}^2\). All of these problems are real
analytic, so our hypotheses below of \(C^\infty\) dynamics is satisfied by these examples. In addition,
toy examples are presented in which a rotation number is known \(a\ priori\) and the methods
converge to the actual rotation number.

2. Applications of theorem 1.1

2.1. Example 1. Rotation vectors

By the Birkhoff ergodic theorem, a function can be integrated using just its values along a tra-
jectory, without knowing the points on the trajectory or where the invariant set is. Given a map
$T : X \to X$ with a quasiperiodic set $X_0$ and a trajectory $(x_n) \subset X_0$, we must convert each $x_n$ into a set of $d$-dimensional angles $\phi_n$, as shown for a 1D-case in figure 1. Our goal is to determine the rotation vector $\rho \in \mathbb{T}^d$, purely from $(\phi_n) \in \mathbb{T}^d$. There are some hidden subtleties. The standard approach towards computing $\rho$ is as the limit $\rho = \lim_{N \to \infty} N^{-1} (\bar{\phi}_N - \bar{\phi}_0)$, where the quantity $\bar{\phi}_N - \bar{\phi}_0$ is computed as $\sum_{n=0}^{N-1} \Delta \phi_n$, where $\Delta \phi_n := \angle(\phi_n+1, \phi_n)$ is some measure of the angle from $\phi_n$ to $\phi_n+1$.

Given a quasiperiodic trajectory, we need to define smooth functions $\Delta \phi$ and $\phi$, taking values in $\mathbb{R}^d$ and $\mathbb{T}^d$ respectively. This general problem can be quite difficult. See [6] for a comprehensive method addressing this general problem of finding rotation number / rates from quasiperiodic trajectories.

Even in the 1D-case, there are two ways to measure the angle between $\phi_n$ and $\phi_{n+1}$. Sometimes, $\Delta \phi$ can be chosen to be the positive angle from $\phi_n$ to $\phi_{n+1}$. That would not work however in figure 1, where $\Delta \phi$ must be the shortest angular difference from $\phi_n$ to $\phi_{n+1}$. In another example, if an astronomer measures the angular position of Mars from the earth, the angle sometimes makes small positive changes and sometimes negative, exhibiting ‘retrograde motion’. In such a case, it is appropriate to choose $\Delta \phi$ so as to minimize the absolute value of the angular change.

To avoid these problems, we will make two assumptions on $\phi$ and $\Delta \phi$. We begin with a definition.

**Definition.** Let $\phi : X_0 \to \mathbb{T}^d$ be a $C^\infty$ map. Then the lift $\tilde{\phi} : \tilde{X}_0 \equiv \mathbb{R}^d \to \mathbb{R}^d$ can be written in the form $\tilde{\phi}(\theta) = A\theta + g(\theta)$, where $g : \mathbb{R}^d \to \mathbb{R}^d$ is bounded (and periodic with period 1 in each coordinate), and $A$ is a matrix called the homology matrix of $\phi$, so that

$$A = A : H_1(X_0) \to H_1(\mathbb{T}^d).$$

Note that the entries of $A$ are again integers. We say $\phi$ is a unit degree map if the determinant of $A$ is $\pm 1$. In that case, $A^{-1}$ is also an integer valued matrix with determinant $\pm 1$. Note that, if $\phi$ is unit-degree, then there is a choice of coordinates for $\theta$ under which $A$ is the identity, and henceforth, we will assume that $A$ is the identity.
(A1) $\phi$ is a $C^\infty$ unit degree map.
(A2) $\Delta \phi : X_0 \to \mathbb{R}^d$ is a $C^\infty$ map such that for each $x \in X_0$, if $\phi(x)$ is a lift of $\phi(x)$, then $\Delta \phi(x) + \phi(x)$ is a lift of $\phi(T(x))$.

Let $\phi_n$ be the point $\phi(x_n)$. Then note that $\Delta \phi(x_n)$ is a lift of $\phi_{n+1} - \phi_n$. Fix any lift $\phi_0$ of $\phi_0$ and inductively define $\phi_n := \phi_{n-1} + \Delta \phi(\theta_{n-1})$. Note that by virtue of $\Delta \phi$, $\phi_n$ is always a lift of $\phi_n$.

**Corollary 2.1.** Let $X, X_0, T,$ and $w$ be as in theorem 1.1. Let $\phi : X_0 \to \mathbb{T}^d$ and $\Delta \phi : X_0 \to \mathbb{R}^d$ satisfy assumptions (A1) and (A2). Then for every initial point $x_0 \in X_0,$

$$WB_N(\Delta \phi) := \frac{1}{A_N} \sum_{n=0}^{N-1} w\left(\frac{n}{N}\right) \Delta \phi_n$$

has super convergence to a vector $\hat{\rho}$, which is a lift of the rotation vector $\rho$, uniformly in $x_0$ as $N \to \infty$.

In other words, $\rho = \hat{\rho} \mod 1$, (i.e. mod 1 in each coordinate).

**Proof.** By theorem 1.1, the sum has super convergence to the integral $\int_{x_0} \Delta \phi \, d\mu$. By the Birkhoff ergodic theorem, $\int_{x_0} \Delta \phi \, d\mu$ is also the limit of the unweighted Birkhoff sum $\frac{1}{N} \sum_{n=0}^{N-1} \Delta \phi_n$, which is equal to $\frac{1}{N} \sum_{n=0}^{N-1} (\phi_{n+1} - \phi_n)$, which is equal to $(\phi_N - \phi_0)/N$. Therefore, the sum in the claim has super convergence to the limit $\lim_{N \to \infty} \frac{\phi_N - \phi_0}{N}$. We will now show that this limit is the rotation number $\rho$.

Note that since $(x_n)$ is a quasiperiodic trajectory, there exists an unknown rotation number $\rho$ and an unknown continuous, periodic function $h : \mathbb{R}^d \to \mathbb{R}^d$ so that for each $n$,

$$\bar{x}_n = n \rho + h(n \rho)$$

is a lift of $x_n$. Recalling the notation $\bar{\phi}(\bar{x}) = A\bar{x} + g(\bar{x})$, we get,

$$\bar{\phi}_n = A\bar{x}_n + g(\bar{x}_n) = nA\rho + Ah(n \rho) + g(n \rho + h(n \rho)).$$

Subtracting $\bar{\phi}_0$ from both sides and dividing by $N$ gives

$$\frac{\bar{\phi}_N - \bar{\phi}_0}{N} = A\rho + \frac{Ah(n \rho)}{N} + \frac{g(n \rho + h(n \rho))}{N} - \frac{\rho_0}{N}.$$

Note that as $N \to \infty$, the last three terms $\to 0$ and as a result, $\lim_{N \to \infty} \frac{\phi_N - \phi_0}{N} = A\rho$ as claimed. \[\square\]

**Remark.** Consider the case of a 1-dimensional quasiperiodic set $X_0$ embedded in $X = \mathbb{R}^2$.
Let $C := C_B \cup C_U$ be the complement of $X_0$ in $\mathbb{R}^2$, where $C_B$ and $C_U$ are the bounded and unbounded components of $C$ respectively. For $p \in \mathbb{R}^2$, define

$$\phi(x) = (x - p)/\|x - p\|.$$
Hence \( \phi(\theta) \in \mathbb{T}^d \) where \( d = 1 \). If \( \rho \in C_b^0 \), then \( \phi \) is a unit degree map. Suppose that the range of values of \( \phi(T(x)) - \phi(x) \mod 1 \) does not include an angle \( \theta_0 \). Then \( \Delta \phi \) can be taken to be the difference 

\[
\Delta \phi(x) : \phi(T(x)) - \phi(x) - \theta_0 \mod 1 + \theta_0.
\]

So \( \Delta \phi \in [0, 1) \) is the length of the arc joining \( \phi_n \) to \( \phi_{n+1} \) and avoiding the angle \( \theta_0 \). See figure 1.

### 2.2. Computing Fourier series

#### 2.2.1. Example 2, Fourier series of the embedding

Assume \( X \subset \mathbb{R}^D \). If a map \( T : X \to X \) is quasiperiodic on \( X_0 \), then \( X_0 \) has a parameterization \( x = h(\theta) \), where \( h : \mathbb{T}^d \to \mathbb{R}^D \), for which \( x_{n+1} = T(x_n) \) is conjugate to the pure rotation \( \theta_{n+1} = \theta_n + \rho \).

We can compute a Fourier series for \( h(\theta) \) provided the rotation number \( \rho \) for \( \theta \) is known, but that can be calculated as a weighted average, as explained above in corollary 2.1. The map \( h \) is not known explicitly, but its values \( (x_n := h(n\rho \mod 1))_{n=0,1,2,...} \) are known. For every \( k \in \mathbb{Z}^d \), the \( k \)th Fourier coefficient of \( h \) is described below.

\[
a_k(h) := \int_{\mathbb{T}^d} h(\theta)e^{-i2\pi k \cdot \theta} \, d\theta, \quad \text{so} \quad h(\theta) = \sum_{k \in \mathbb{Z}^d} a_k e^{i2\pi k \cdot \theta}.
\]

The degree of differentiability of \( h \) can be estimated from the decay rate of \( a_k \) with \( ||k|| \). Therefore, the crucial step in this estimation is an accurate calculation of each \( a_k \), which can be obtained as the limit of a weighted Birkhoff average, as stated in the following corollary.

**Corollary 2.2.** Under the assumptions of corollary 2.1, for each \( k \in \mathbb{Z}^d \), the quantity

\[
\text{WB}_N[h(\theta)e^{-i2\pi k \cdot \theta}] = \frac{1}{AN} \sum_{n=0}^{N-1} w(N) x_n e^{-i2\pi nk \cdot \rho}
\]

has super convergence as \( N \to \infty \) to \( a_k(h) \), the \( k \)th Fourier coefficient of \( h \).

**Proof.** By theorem 1.1, the limit of the weighted Birkhoff average of the smooth function \( h(\theta)e^{-i2\pi k \cdot \theta} \) has super convergence to the integral \( \int_{\mathbb{T}^d} h(\theta)e^{-i2\pi k \cdot \theta} \, d\theta \), which is precisely \( a_k(h) \), the \( k \)th Fourier coefficient of \( h \).

Determining the smoothness of conjugacy to a pure rotation is an old problem. While we can only determine the degree of differentiability of the conjugacy function \( h \) computationally (non-rigorously) by observing how quickly its Fourier series coefficients \( a_k \) go to 0 as \( ||k|| \to \infty \), the papers [7, 20–22] arrive at rigorous conclusions on the differentiability of \( h \) by making various assumptions on the smoothness of the quasiperiodic map \( T \) and the Diophantine class of its rotation number \( \rho \). In our case, we conclude that \( h \) is real analytic if \( ||a_k|| \) decreases exponentially fast, i.e. \( \log ||a_k|| \leq A + B|k| \) for some \( A \) and \( B \), to the extent checkable with a given computer precision. Also see [5] for a discussion on computing Fourier coefficients of maps between tori instead of maps from the torus to \( \mathbb{R}^D \).
2.2.2. Example 3, computing the integral of a periodic $C^\infty$ function

We designed the above theorem as a tool for investigating quasiperiodic sets, but sometimes we can artificially create the quasiperiodic dynamics. To integrate a $C^\infty$ map with respect to the Lebesgue measure when the map is periodic in each of its $d$ variables, we can rescale its domain to a $d$-dimensional torus $\mathbb{T}^d = [0, 1]^d$ mod 1 in each coordinate. Choose $\rho = (\rho_1, \cdots, \rho_d) \in (0, 1)^d$ of Diophantine class $\beta > 0$. Let $T = T_\rho$ be the rotation by the Diophantine vector $\rho$ on $\mathbb{T}^d$. Let $w$ be the exponential weighting function equation (4).

Then by theorem 1.1, for every $\theta \in \mathbb{T}^d$, $W_B^{N}(f)(\theta)$ has super convergence to $\int_{\mathbb{T}^d} f \, d\mu$ and convergence is uniform in $\theta$.

3. A more general $C^m$ version of theorem 1.1

Theorem 1.1 is a special case of the following more detailed theorem 3.1.

$C^m$ bump functions. Let $m \geq 0$ be an integer. We will say that a function $w : \mathbb{R} \to \mathbb{R}$ is a $C^m$ bump function if $w$ is a $C^m$ function, is 0 outside $[0, 1]$ and $\int_0^1 w(t) \, dt > 0$. Hence $w$ and its first $m$ derivatives are 0 at 0 and 1. For example, the function which is $x(1-x)$ on $[0, 1]$ is $C^0$, while $\sin^2(\pi t)$ is $C^1$.

**Theorem 3.1.** Let $m > 1$ be an integer and let $w$ be a $C^{m^*}$ bump function for some $m^* \geq m$. For $M \in \mathbb{N}$, let $X$ be a $C^{M}$ manifold and $T : X \to X$ be a $C^{M}$ $d$-dimensional quasiperiodic map on $X_0 \subseteq X$, with invariant probability measure $\mu$ and a rotation vector of Diophantine class $\beta$. Let $f : X \to E$ be $C^M$, where $E$ is a finite-dimensional, real vector space. Then there is a constant $C_m$ depending upon $w, f, m, M,$ and $\beta$ but independent of $x_0 \in X_0$ such that

$$\left| \left( W_B^{N}f \right)(x_0) - \int_{X_0} f \, d\mu \right| \leq C_m N^{-m},$$

provided the ‘smoothness’ $M$ satisfies

$$M > d + m(d + \beta).$$

**Proof of theorem 1.1 as a corollary of theorem 3.1.** Assume the hypotheses of theorem 1.1. Then the hypotheses of theorem 3.1 are satisfied with $M = \infty$. Therefore, for every integer $m > 1$, in equation (8) holds for some $C_m > 0$. Therefore, $W_B^{N}(f)$ has super convergence to $\int_{X_0} f \, d\mu$, uniformly in $x_0$. ■

3.1. Proof of theorem 3.1

We need the following lemma in our proof.

**Lemma 3.2 (Poisson summation formula, [23]).** For each $g \in L^2(\mathbb{R})$,

$$\Sigma_n g(n) = \Sigma_n \int_{\mathbb{R}} g(t) e^{-i2\pi nt} \, dt.$$

To prove theorem 3.1, we use the coordinates $\theta$ on $X_0$ for which equation (2) holds. So for the rest of this section, $X_0 = \mathbb{T}^d$ and $T$ is an irrational rotation $T_\rho$, as described in equation (1). Of course $T_\rho^n = T_\rho^n : \theta \mapsto \theta + np \pmod{1}$. As a result, $f(T_\rho^n(\theta)) = f(\theta + np)$. Therefore, in this dynamical system, equation (5) takes the form
so that
\[ \| f_{\mathbf{C}w} - n \|_1 \leq \frac{B}{N} \cdot \| \mathbf{C}w \cdot \Sigma - (31) \text{ is finite.} \]

Hence there is a constant \( B \) for some \( \mathbf{C}w \cdot \Sigma \). Note that the Fourier coefficient \( a_0 \) of \( f \) is the integral \( \int_{\mathbb{T}^d} f d\theta \) and is also \( \mathbf{W}_B(a_0 \sigma_0)(\theta) \) for every \( N \) since \( \sigma_0 \equiv 1 \). Therefore,
\[
E_N := \{ \mathbf{W}_B f \}(\theta) - \int_{\mathbb{T}^d} f(t) dt
\]
\[
= \sum_{k \in \mathbb{Z}^d} a_k \mathbf{W}_B \sigma_k(\theta)
\]
\[
= A_N^{-1} \sum_{k \in \mathbb{Z}^d} a_k \Sigma^n \left[ w(n/N) \sigma_k(\theta + n\rho) \right]
\]
\[
= A_N^{-1} \sum_{k \in \mathbb{Z}^d} a_k \left[ \sum_n \int_{\mathbb{R}} w(t/N) \sigma_k(\theta + t\rho) e^{-i2\pi m t} dt \right] \cdot \text{by lemma 3.2}
\]
\[
= A_N^{-1} \sum_{k \in \mathbb{Z}^d} a_k \left[ \int \sigma_k \left[ \sum_n \int_{\mathbb{R}} w(t/N) e^{i2\pi m (k \cdot \rho - n)} dt \right] \right].
\]

Set \( s = t/N \) so that \( N ds = dt \) and
\[
|E_N| \leq (N/A_N) \sum_{k \in \mathbb{Z}^d} |a_k| \sum_n \left| \int_{\mathbb{R}} w(s) e^{i2\pi s \rho - n} ds \right|.
\]

But \( N/A_N \) is bounded by some constant \( C_w > 0 \) since it converges to \( \int_0^1 w(s) ds \).

We establish a bound for \( \int_0^1 w(s) e^{\Omega s} ds \), where \( \Omega := i2\pi N(k \cdot \rho - n) \). Since \( w(s) \) and its first \( m \) derivatives vanish at 0 and 1, integrating by parts \( m \) times yields
\[
\int_0^1 w(s) e^{\Omega s} ds = \Omega^{-m} \int_0^1 w^{(m)}(s) e^{\Omega s} ds \leq |\Omega|^{-m} \| w^{(m)} \|,
\]
where \( \| w^{(m)} \| \) is the \( L^1 \) norm of the \( m \)th derivative of \( w \).

Since \( f \) is a \( C^M \)-function, differentiating it \( M \) times gives \( |a_k| \leq C_{f,M} \| k \|^{-M} \), for each \( k \in \mathbb{Z}^d \) for some \( C_{f,M} > 0 \) that depends only on \( f \) and \( M \). In equation (10) gives
\[
|E_N| \leq C_w \sum_{k \in \mathbb{Z}^d} C_{f,M} \| k \|^{-M} \| \Sigma^n (2\pi)^{-m} N^{-m} \| k \cdot \rho - n \|^{-(m)} \| w^{(m)} \|,
\]
\[
= C_n \| \Sigma^n \| \| \Sigma^n \| \| k \cdot \rho - n \|^{-(m)}
\]
where we have collected constants into \( C_n := C_n (2\pi)^{-m} C_{f,M} \| w^{(m)} \| \). Therefore our proof is finished when we show that the quantity in the above brackets \( B := \Sigma_{k \in \mathbb{Z}^d} \| k \|^{-M} \Sigma_n \| k \cdot \rho - n \|^{-(m)} \) is finite. We will bound \( B \) by \( \Sigma_{k \in \mathbb{Z}^d} \| k \|^{-d^*} \), which is finite if and only if \( d^* > d \).

To bound the sum \( \Sigma \| k \cdot \rho - n \|^{-(m)} \), note that there is a unique integer \( n_0 \) which is closest to the number \( k \cdot \rho \). In equation (6) implies \( \| k \cdot \rho - n_0 \|^{-m} \leq C_{\rho,m} \| k \|^{m(d+\beta)} \). The rest of the sum satisfies \( \Sigma_{n \neq n_0} \| k \cdot \rho - n \|^{-(m)} \leq \Sigma_{n \neq 0} (|n| + 0.5)^{-m} \), which is bounded since \( m > 1 \). Hence there is a constant \( C_{\rho,m} > 1 \) such that \( \Sigma_n \| k \cdot \rho - n \|^{-(m)} \leq C_{\rho,m} \| k \|^{m(d+\beta)} \). Hence \( B \leq \text{constant} \times \Sigma_{k \in \mathbb{Z}^d} \| k \|^{-d^*} \) where \( -d^* := -M + m(d + \beta) \), since by in equation (9), \( d^* > d \)
and \( B \) is finite.
Acknowledgments

JY was partially supported by National Research Initiative Competitive grants 2009-35205-05209 and 2008-04049 from the USDA. We thank a commentator who wishes to remain anonymous for suggesting the use of the Poisson Summation Formula to simplify our original proof.

ORCID iDs

Suddhasattwa Das https://orcid.org/0000-0003-2085-4743

References

[1] Sander E and Yorke J A 2015 The many facets of chaos Int. J. Bifurcation Chaos 25 15300
[2] Simon J L, Bretagnon P, Chapront J, Chapront-Touze M, Francou G and Laskar J 1994 Numerical expressions for precession formulae and mean elements for the Moon and the planets Astron. Astrophys. 282 663
[3] Brin M and Stuck G 2002 Introduction to Dynamical Systems (Cambridge: Cambridge University Press)
[4] Krengel U 1978 On the speed of convergence in the ergodic theorem Mon.hefte Math. 86 3–6
[5] Das S, Saiki Y, Sander E and Yorke J 2017 Quantitative quasiperiodicity Nonlinearity 30 4111
[6] Das S, Saiki Y, Sander E and Yorke J 2017 Solving the Babylonian problem of quasiperiodic rotation rates (arXiv:1706.02595)
[7] Herman M R 1979 Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations Publ. Math. l’Inst. Hautes Études Sci. 49 5–233
[8] Simó C 1994 Averaging Under Fast Quasiperiodic Forcing (Boston, MA: Springer) pp 13–34
[9] Simó C, Sousa-Silva P and Terra M 2013 Practical Stability Domains Near L4,5 in the Restricted Three-Body Problem: Some Preliminary Facts vol 54 (Berlin: Springer) (https://doi.org/10.1007/978-3-642-38830-9_23)
[10] Arnold V 1965 Small denominators. I. Mapping of the circumference onto itself Am. Math. Soc. Transl. 46 213–84
[11] Lin M and Weber M 2007 Weighted ergodic theorems and strong laws of large numbers Ergod. Theor. Dynam. Syst. 27 511–43
[12] Bellow A, Jones R L 1992 Almost everywhere convergence of weighted ergodic averages Mathematische Annalen 293 399–426
[13] Bellow A and Losert V 1985 The weighted pointwise ergodic theorem and the individual ergodic theorem along subsequences Trans. Am. Math. Soc. 288 307–45
[14] Durand P and Shneider D 2002 Ergodic averages with deterministic weights Ann. Inst. Fourier 52 561–83
[15] Luque A and Villanueva J 2014 Quasi-periodic frequency analysis using averaging-extrapolation methods SIAM J. Appl. Dyn. Syst. 13 1–46
[16] Laskar J 1999 Introduction to Frequency Map Analysis (Dordrecht: Springer) pp 134–50
[17] Laskar J 1993 Frequency analysis of a dynamical system Celest. Mech. Dyn. Astron. 56 191–6
[18] Laskar J 1993 Frequency analysis for multi-dimensional systems global dynamics and diffusion Physica D 67 257–83
[19] Laskar J 2003 Frequency map analysis and particle accelerators 20th Proc. IEEE Particle Accelerator Conf. (Portland, Oregon, 2–16 May 2003) pp 378–82
[20] Katznelson Y and Ornstein D 1989 The absolute continuity of the conjugation of certain diffeomorphisms of the circle Ergod. Theor. Dynam. Syst. 9 681–90
[21] Katznelson Y and Ornstein D 1989 The differentiability of the conjugation of certain diffeomorphisms of the circle Ergod. Theor. Dynam. Syst. 9 643–80
[22] Yoccoz J C 1984 Conjugaison différentiable des difféomorphismes du cercle dont le nombre de rotation vérifie une condition diophantienne Ann. Sci. l’École Norm. Supér 17 333–59
[23] Gasquet C and Witomski P 1999 Sampling Signals and Poisson’s Formula (Berlin: Springer) (https://doi.org/10.1007/978-1-4612-1598-1_37)