Ergodic Optimization and Prevalence

Third Palis–Balzan International Symposium on Dynamical Systems

Jairo Bochi
(Catholic Univ. of Chile, Santiago)

June 18, 2015

Ergodic optimization: an overview

Basic reference: O. Jenkinson. *Ergodic Optimization*, Discrete and Cont. Dyn. Sys. A, vol. 15 (2006), pp. 197–224.

Disclaimer: I won't discuss the relations with Classical Mechanics or Thermodynamical Formalism.

Ergodic optimization: the general setting

- X = compact metric space
- $T : X \to X$ continuous map
- $f : X \to \mathbb{R}$ continuous function ("performance" or "potential")
- $\mathcal{M}_T := \{T\text{-invariant probability measures}\}$
- "ergodic supremum"

$$\begin{aligned}
\beta(f) &:= \sup_{\mu \in \mathcal{M}_T} \int f \, d\mu \\
&= \sup_{x \in X} \limsup_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) \\
&= \lim_{n \to \infty} \sup_{x \in X} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x)
\end{aligned}$$

An easy example

$X = \{0,1\}^{\mathbb{N}} = 2^{\mathbb{N}}$ Cantor set
$T \colon 2^{\mathbb{N}} \to 2^{\mathbb{N}}$ shift
$f =$ characteristic function of cylinder $C = [101]$

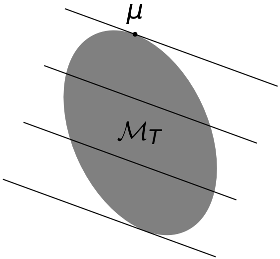
Then $\beta(f) = 1/2$. Indeed:

- Since $T^{-1}(C) \cap C = \varnothing$, for every $x \in 2^{\mathbb{N}}$, the frequency of visits to C is $\leq 1/2$;
- The T-invariant prob. μ supported on the orbit of $\overline{10} = (1,0,1,0\dots)$ has $\int f \, d\mu = 1/2$.
  Rem.: μ is the *unique* such measure.

Maximizing measures

In general, a measure $\mu \in \mathcal{M}_T$ s.t. $\int f \, d\mu = \beta(f)$ is called a **maximizing measure**.

Existence? Yes (compactness).



Generic uniqueness:

Theorem (Jenkinson and others)

*For (topologically) generic f in any "reasonable"(*) space $\mathcal{F}$ of continuous functions, the maximizing measure is **unique**.*

(*) a vector space $\mathcal{F}$ continuously and densely embedded in $C^0(X)$.

The general problem

Problem

For a fixed "nice" dynamical system T, and a fixed "nice" family/space $\mathcal{F}$ of functions f, understand the maximizing measures for all/most functions f.

Of course, the problem is uninteresting if T has *few* invariant measures.

In most of the literature, T is assumed to have strong **hyperbolicity** properties and therefore lots of periodic measures.
In all that follows we will assume T to be **uniformly expanding**.

Regularity makes a big difference

Assume $T =$ uniformly expanding.

Theorem (Bousch–Jenkinson)

For generic C^0 functions, the maximizing measures have full support.

The situation is very different if the functions are more regular:

Theorem (Subordination principle)

If $f \in C^\alpha$ (i.e. f is α-Hölder) then there exists a compact invariant set $K_f \subset X$ ("Mather set") such that

$$\mu \in \mathcal{M}_T \text{ is maximizing for } f \iff \operatorname{supp} \mu \subset K_f \, .$$

Corollary of the **Mañé Lemma** (or Mañé–Conze–Guivarc'h–Savchenko–Fathi–Contreras–Lopes–Thieullen–Bousch Lemma).

A nice example (Hunt, Ott, Jenkinson, Bousch)

The following example was first studied experimentally by Hunt and Ott (1996):

- $T(x) := 2x \bmod 1$ on $X = \mathbb{R}/2\pi\mathbb{Z}$.
- Family $\mathcal{F}$ of functions: (nonzero) linear combinations of $\cos x$ and $\sin x$.

Theorem (Bousch 2000)

In that setting, maximizing measures are always unique. Moreover, for an ***open and full measure*** *subset of $\mathcal{F}$, the maximizing measure is supported on a* ***periodic orbit****.*

(Actually the maximizing measures are *Sturmian*.)

The big conjecture

Conjecture (Hunt–Ott 1996)

*For typical chaotic systems, typical parameterized families of smooth functions, and most values of the parameter, the maximizing measure is unique and **supported on a periodic orbit**.*

(Terms in color are left undefined. . .)

An important result

Improving on the work of previous authors (Yuan–Hunt, Contreras–Lopes–Thieullen, Bousch, Bressaud–Quas, Morris, Quas–Siefken), Contreras managed to prove the following:

Theorem (Contreras 2013)

For uniformly expanding dynamics, and (topologically) generic Lipschitz functions, the maximizing measure is (unique and) supported on a periodic orbit.

Actually the conclusion holds for an open and dense subset of $C^{\mathrm{Lip}}(X)$, and the "locking property" holds: the maximizing measures are robust under perturbations.

Goal

We would like to obtain results like Contreras', but with genericity being not only in the topological sense, but in a **probabilistic** sense as well (thus being a little closer to the spirit of the Hunt–Ott conjecture).

Setting for our main result (details later):

- $T =$ one-sided shift on 2 symbols;
- $\mathcal{F} =$ space of "super-continuous" functions (very strong modulus of regularity);
- "probabilistic genericity" is expressed in terms of **prevalence**.

Motivation for prevalence

Is is possible to speak of probabilities in infinite-dimensional vector spaces?

- ☹ There is no useful (say, σ-finite) translation-invariant measure.
- ☹ There is no useful (say, σ-finite) translation-invariant class of measures;
- ☺ However there is a translation-invariant notion of "almost every point", called **prevalence** [Hunt–Sauer–Yorke, Christensen].

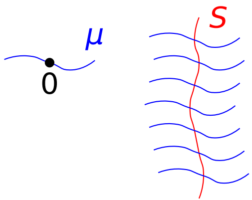
Measure transversality and shyness

- $\mathcal{F}$ = complete metrizable vector space;
- $S \subset \mathcal{F}$ Borel set;
- μ = Borel probability measure on $\mathcal{F}$ with compact support.

μ is called **transverse** to S ($\mu \pitchfork S$) if:

$$\forall f \in \mathcal{F}, \quad \mu(S - f) = 0.$$

I.e. summing to *any* $f \in \mathcal{F}$ a random perturbation we get outside of S with μ-probability 1.



$S \subset \mathcal{F}$ is called **shy** if $\exists\ \mu \pitchfork S$.

Prevalent sets

A Borel subset of a complete metrizable vector space is called **prevalent** if its complement is shy.

Less formally: In order to prove that a set $P \subset \mathcal{F}$ is prevalent, we need to find a compactly supported measure μ such that given *any* $f \in \mathcal{F}$, if we perturb f by adding a μ-random term g, then $f + g \in P$ with μ-probability 1.

In that case, we can always replace μ by another with small support. Thus $f + g$ can be thought as a **random perturbation** of f.

Properties of prevalence

- $\dim \mathcal{F} < \infty \Rightarrow$ the prevalent sets are exactly those of full Lebesgue measure.
- Prevalence is preserved under translation.
- Prevalence is preserved under augmentation.
- Prevalence is preserved under countable intersection.
- Prevalence implies denseness.

Some spaces of functions on $2^{\mathbb{N}}$

Given a sequence of positive numbers $\mathbf{a} = (a_n) \searrow 0$, define a metric on $2^{\mathbb{N}} = \{0,1\}^{\mathbb{N}}$:

$$d_{\mathbf{a}}(x,y) := a_{n(x,y)} \quad \text{where} \quad n(x,y) := \inf\{i \in \mathbb{N};\ x_i \neq y_i\}.$$

Space of functions:

$$C^{\mathbf{a}}(2^{\mathbb{N}}) := \left\{ f \colon X \to \mathbb{R};\ f \text{ is Lipschitz w.r.t. } d_{\mathbf{a}} \right\}$$

(The faster $a_n \to 0$, the smaller the space $C^{\mathbf{a}}$.)
This is a (nonseparable) Banach space with the norm:

$$\|f\|_{\mathbf{a}} := \|f\|_\infty + \mathrm{Lip}_{\mathbf{a}}(f)\,.$$

Example: $d_{\mathbf{a}}$ with $\mathbf{a} = (2^{-n})$ is the “usual” metric on X.
The space of α-Hölder functions w.r.t. the usual metric is $C^{\mathbf{b}}(2^{\mathbb{N}})$ where $\mathbf{b} = (2^{-\alpha n})$.

The main theorem

Theorem (joint with Yiwei Zhang. ArXiv:1501.00961)

The locking property (*) is prevalent in $C^{\mathbf{a}}(2^{\mathbb{N}})$, provided $\mathbf{a} = (a_n) \searrow 0$ sufficiently fast (**).

(*) A function $f \in C^{\mathbf{a}}(2^{\mathbb{N}})$ satisfies the **locking property** if:

- f has a unique maximizing measure μ (w.r.t. the shift), and it is periodic;
- μ is also the unique maximizing measure for every $g \in C^{\mathbf{a}}(2^{\mathbb{N}})$ sufficiently close to f.

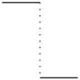
(**) Unfortunately, we need really fast convergence to 0, namely:

$$\frac{a_{n+1}}{a_n} = O\left(2^{-2^{n+2}}\right)$$

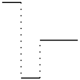
Haar functions

The **Haar functions** are continuous and form an orthogonal basis of $L^2(2^{\mathbb{N}}, \mathrm{bernoulli}_{\frac{1}{2},\frac{1}{2}})$; they are 1 and

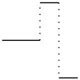
$$h_\varnothing := \frac{1}{2}(\chi_{[0]} - \chi_{[1]}) =$$



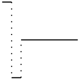
$$h_0 := \frac{1}{2}(\chi_{[00]} - \chi_{[01]}) =$$



$$h_1 := \frac{1}{2}(\chi_{[10]} - \chi_{[11]}) =$$



$$h_{00} := \frac{1}{2}(\chi_{[000]} - \chi_{[001]}) =$$



$$\ldots$$

$$h_\omega := \frac{1}{2}(\chi_{[\omega 0]} - \chi_{[\omega 1]}) \qquad (\omega = \text{word}).$$

Haar series

Every continuous function f on the Cantor set $2^{\mathbb{N}}$ has a uniformly convergent (*) **Haar series**:

$$f(x) = c + \sum_{\omega} c_\omega h_\omega(x) \, ,$$

where ω runs on the (finite) words on the letters 0, 1.

(*) In that sense Haar series are better behaved that Fourier series.

The spaces $C^{\mathbf{a}}(2^{\mathbb{N}})$ introduced before can be essentially characterized in terms of the decay of the Haar coefficients (c_ω).

The random perturbations

Given a family of positive numbers $\mathbf{b} = (b_\omega)$ indexed by words ω, we define a set of functions:

$$\mathcal{H}_{\mathbf{b}} := \left\{ \sum_{\omega} c_\omega h_\omega;\ c_\omega \in [-b_\omega, b_\omega] \right\} = \textbf{Hilbert brick.}$$

Then, for appropriate $\mathbf{b}$ (e.g. $b_\omega = a_n/(n+1)$, $n = |\omega|$):

- $\mathcal{H}_{\mathbf{b}}$ is a compact subset of $C^{\mathbf{a}}(2^{\mathbb{N}})$;
- taking random independent coefficients $c_\omega \sim \text{Uniform}([-b_\omega, b_\omega])$ we obtain a probability $\mu_{\mathbf{b}}$ supported on $\mathcal{H}_{\mathbf{b}}$;
- **these are the random perturbations in our Main Theorem**, i.e., the measure $\mu_{\mathbf{b}}$ is transverse to the set of functions that don't have the locking property.

Strategy of the proof of the Main Theorem

A **step function** of level n is a function on $2^{\mathbb{N}}$ that is constant on cylinders of rank n. We will see that **step functions have periodic maximizing measures**.

Since $\mathbf{a} = (a_n) \to 0$ very fast, the functions f in $C^{\mathbf{a}}(2^{\mathbb{N}})$ are well-approximated by step functions f_n (which can be obtained by truncating the Haar series).

We will show that **with probability** 1 (in any translated Hilbert brick. . .), **the maximizing measure for f coincides with the (periodic) maximizing measure for f_n** for some n.

We need quantitative information on the ergodic optimization of step functions. . .

Finite dimensional ergodic optimization

Let F be a finite-dimensional vector space of functions, with basis $\{f_1, \dots, f_n\}$.
Define a "projection" linear map $\pi\colon \mathcal{M} \to \mathbb{R}^n$ on the vector space of signed measures $\mathcal{M}$ by

$$\pi(\mu) := \left(\int f_1 \, d\mu, \dots, \int f_n \, d\mu \right).$$

Define a compact convex set:

$$R := \pi(\mathcal{M}_T) = \textbf{rotation set}$$

(the projection of the T-invariant probability measures).

Origin of the name: $(f_1, \dots, f_n) =$ displacement function of a map $T\colon \mathbb{T}^n \to \mathbb{T}^n$ homotopic to id.

Finite dimensional ergodic optimization

Functions $f \in F$ can be "integrated" with respect to vectors $v \in R = \pi(\mathcal{M}_T)$:

$$\langle f, v\rangle := \int f \, d\mu \quad \text{where } \mu \text{ is s.t. } \pi(\mu) = v.$$

To compute the "ergodic supremum" becomes a finite-dimensional problem:

$$\beta(f) := \sup_{\mu \in \mathcal{M}_T} \int f \, d\mu = \sup_{v \in R} \langle f, v\rangle \, .$$

If the extreme points of the rotation set R happen to have unique preimages in $\mathcal{M}_T$ then every $f \in F$ has a unique maximizing measure.

Finite dimensional ergodic optimization

Conclusion

Ergodic optimization of functions in an n-dimensional space $F \subset C^0(X)$ is basically equivalent to:

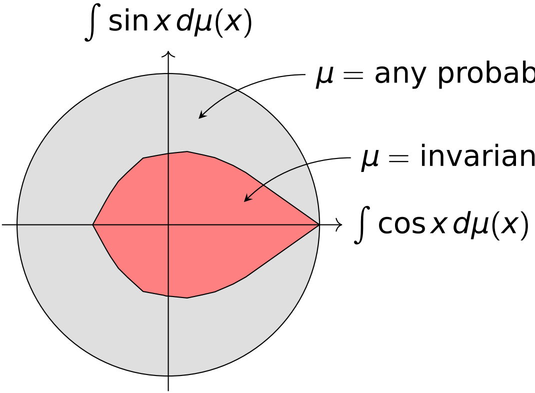
- regarding F as $(\mathbb{R}^n)^*$;
- determining the extreme points of the compact convex set $R := \pi(\mathcal{M}_T) \subset \mathbb{R}^n$;
- determining their preimages under $\pi \colon \mathcal{M}_T \to \mathbb{R}^n$.

Remark

For $T =$ shift, every compact convex set $R \subset \mathbb{R}^n$ can be realized as a rotation set (for suitable C^0 functions). (Kucherenko–Wolf)

The fish on the dish

Example #1 (Hunt, Ott, Jenkinson, Bousch):
$T(x) = 2x \bmod 1$ on $\mathbb{R}/2\pi\mathbb{Z}$, $F := \{\text{trig. poly. deg 1}\}$.



Note: "sharper" extreme points of the fish are more likely to be maximizing...

Example #2: step functions of level 2

$F := \{\text{step functions on } 2^{\mathbb{N}} \text{ of level } 2\}$, with basis $\chi_{[00]}, \chi_{[01]}, \chi_{[10]}, \chi_{[11]}$.
The projection $\pi\colon \mathcal{M} \to \mathbb{R}^4$ is:

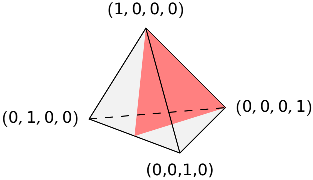
$$\mu \mapsto (\mu([00]), \mu([01]), \mu([10]), \mu([11])).$$

The "dish" $\pi(\{\text{prob. measures}\}) =$ unit simplex:

$$\Delta = \left\{(p_{ij}) \in \mathbb{R}^4;\ p_{ij} \geq 0,\ \sum p_{ij} = 1\right\}.$$

The "fish" $R = \pi(\{\text{inv. prob.}\})$ is

$$R = \left\{(p_{ij}) \in \Delta;\ p_{01} = p_{10}\right\}.$$



The vertices have unique pre-images in $\mathcal{M}_T$, which are measures supported on periodic orbits:

Vertex of R	per. orb.
$(1,0,0,0)$	$\overline{0}$
$(0,0,0,1)$	$\overline{1}$
$(0,\frac{1}{2},\frac{1}{2},0)$	$\overline{01}$

Generalization: Step functions of level n

For the shift $T\colon 2^{\mathbb{N}} \to 2^{\mathbb{N}}$, consider:

- $F_n :=$ {step functions of level n} $\simeq \mathbb{R}^{2^n}$;
- $R_n :=$ associated rotation set.

Theorem (Ziemian)

- *The rotation set R_n is a **polytope** in $\mathbb{R}^{2^n}$;*
- *each vertex of R_n is the projection of a **unique** shift-invariant measure, which is supported on a periodic orbit.*

The polytopes R_n

	dim	# vertices	assoc. periodic orbits
R_1	1	2	$\overline{0}, \overline{1}$
R_2	2	3	$\overline{0}, \overline{1}, \overline{01}$
R_3	4	6	$\overline{0}, \overline{1}, \overline{01}, \overline{001}, \overline{011}, \overline{0011}$
R_4	8	19	
R_5	16	179	
R_6	32	30166	

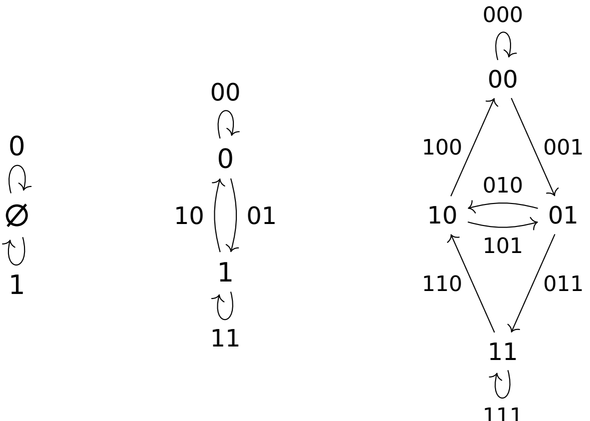
The number of vertices grows super-exponentially; there is no exact formula.

To describe the polytopes R_n, we need to introduce a combinatorial object.

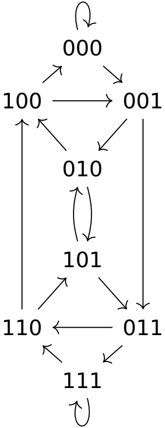
de Bruijn graphs

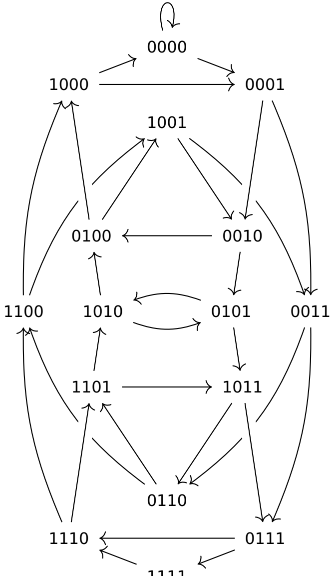
The **de Bruijn graph** G_n has:

- nodes labelled by words on length $n-1$;
- arrows labelled by words ω on length n, of form $\text{prefix}(\omega) \xrightarrow{\omega} \text{suffix}(\omega)$;



G_4 and G_5


000
100
001
010
101
110
011
111


0000
1000
0001
1001
0100
0010
1100
1010
0101
0011
1101
1011
0110
1110
0111

The graph G_n and the rotation set R_n

Recall: $F_n := \{\text{step functions of level } n\}$
Given $f \in F_n$ assigns **weights** of the arrows of G_n. The maximizing measure μ for f can be obtained as follows:

- find the (simple closed) **cycle of G_n of maximum mean weight**;[1]
- this cycle can be seen as a periodic orbit for the shift;
- μ is the measure supported on this orbit.

Conclusion

The set R_n is indeed a polytope; its vertices correspond to the (simple closed) cycles on the graph G_n.

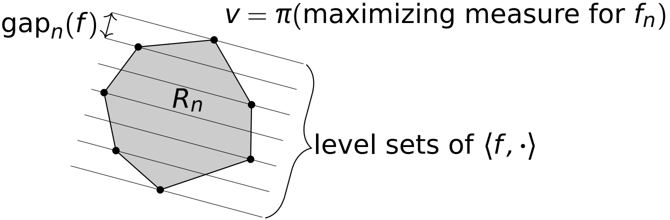
[1] This problem is studied in applied math (Karp algorithm . . .)

A "measure" of uniqueness

Suppose $f\colon 2^{\mathbb{N}} \to \mathbb{R}$ is a step function of level n, or equivalently, an attribution of weights to the arrows of G_n .

Compute $\langle f, v\rangle$ for each vertex v of the polytope R_n.

Let $\mathrm{gap}_n(f) :=$ the difference between the maximum and the second maximum:



So $\mathrm{gap}_n(f) \geq 0$, and $\mathrm{gap}_n(f) > 0$ iff the maximizing measure is unique.

Proof of the prevalence theorem

Let us recall the main theorem:

Theorem (B., Zhang)

Fix a space of "super-continuous" functions $C^{\mathbf{a}}(2^{\mathbb{N}})$, and an appropriate Hilbert brick

$$\mathcal{H}_{\mathbf{b}} := \left\{ \sum_{\omega} c_\omega h_\omega ;\ c_\omega \in [-b_\omega, b_\omega] \right\} .$$

Let $g \in C^{\mathbf{a}}(2^{\mathbb{N}})$, and take a random function f in the translated Hilbert brick $g + \mathcal{H}_{\mathbf{b}}$.
Then there exists a "periodic measure" μ which is the unique maximizing measure for f and for all $\tilde{f} \in C^{\mathbf{a}}(2^{\mathbb{N}})$ sufficiently close to f.

Main Lemma: the Gap criterion

Lemma (Gap criterion)

*Given an **arbitrary continuous function** f, truncate its Haar series to obtain a step function f_n:*

$$f(x) = c(f) + \sum_{\omega} c_\omega(f) h_\omega(x) \;\Rightarrow\; f_n(x) := c(f) + \sum_{|\omega|<n} c_\omega(f) h_\omega(x).$$

*If the following **gap condition** holds:*

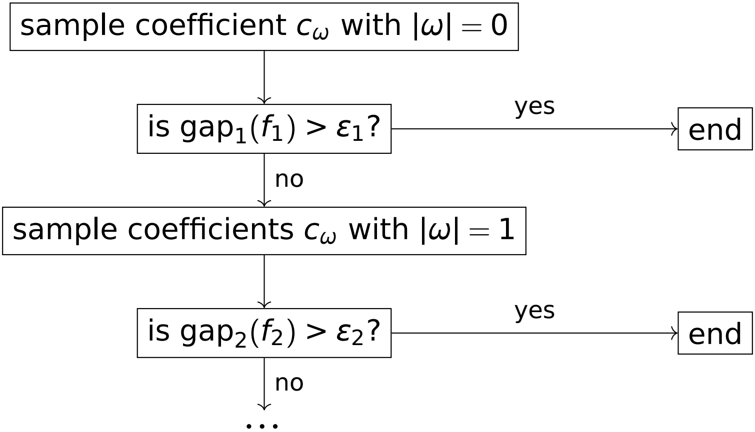
$$\mathrm{gap}_n(f_n) > \sum_{k=n}^{\infty} (k-n+1) \max_{|\omega|=k} |c_\omega(f)|$$

then the maximizing measure for f_n (which is unique and periodic) is also the maximizing measure for f.

Proof: Combinatorial arguments with the de Bruijn graphs (4 pages).

Proof of the theorem

Let ε_n be an upper bound for the RHS in the gap condition. The following “algorithm” finds the maximizing measure (provided it stops):



Proof of the theorem

We need to show that the algorithm stops with probability 1, i.e., $\mathrm{Prob}\big[\exists n;\ \mathrm{gap}_n(f_n) > \varepsilon_n\big] = 1$.

- $\mathrm{gap}_n(f_n)$ depends on the Haar coefficients of level $n-1$;
- $\varepsilon_n = O(\text{Haar coefficients of level } n)$;
- the Haar coefficients of level n are much smaller than the **variance** of the Haar coefficients of level $n-1$.

It follows that:

- $\mathrm{variance}(\mathrm{gap}_n(f_n)) \gg \varepsilon_n$;
- $\mathrm{Prob}\big[\,\mathrm{gap}_n(f_n) > \varepsilon_n\big] \to 1$ (overkill)
- $\mathrm{Prob}\big[\text{algorithm stops at a level} \leq n\big] \to 1$

Proof of the theorem

Why do we need super-exponential decay of the Haar coefficients (strong modulus of continuity)?

Because:

- the polytope R_n has a super-exponential number of vertices;
- these vertices are the candidates for maximizing measures for f_n;
- and we need to guarantee a gap between the top 2 vertices.

How to improve the main result?

What about the **Hölder case** (exponential decay of Haar coefficients)? Recall:

Lemma (Gap criterion)

$$\mathrm{gap}_n(f_n) > \varepsilon_n \geq \sum_{k=n}^{\infty} (k-n+1) \max_{|\omega|=k} |c_\omega(f)| \quad \Rightarrow$$

the maximizing measure for f_n (which is unique and periodic) is also the maximizing measure for f.

Hölder case $\Rightarrow \varepsilon_n \to 0$ exponentially, while computer experiments indicate that $\mathrm{gap}_n(f_n) \to 0$ polynomially (i.e. $O(1/n^\alpha)$) a.s. (despite the super-exponential number of candidate maximizers.)

A finer understanding of the geometry of the polyhedra R_n may help. . .