Learning Magnetic Field Structure from Trajectories

David Bindel

3 Feb 2025

Department of Computer Science
Cornell University

Who?

Simons Collaboration: “Hidden Symmetries and Fusion Energy”

`https://hiddensymmetries.princeton.edu/`

A collaboration of plasma physicists and mathematicians from:

Princeton, NYU, Maryland, IPP Greifswald, Warwick, CU Boulder, Cornell, UW Madison, EPFL, ANU, UT Austin, U Arizona.

(along with many unfunded collaborators)

- Phase 0: Aug 2017-Aug 2018
- Phase 1: Sep 2018-Aug 2022
- Phase 2: Sep 2022-Aug 2025

Some Phase 0 recollections

2017-08-22	Email from Antoine Cerfon, "would you be interested in participating in these initial conversations?"
2017-09-01	Initial conversation
2017-10-04	LOI submitted
2017-12-06	First two-day proposal meeting
2018-01-31	Second two-day proposal meeting
2018-02-15	Proposal submitted
2018-04-18	Panel pitch (Bhattacharjee, MacKay, Bindel)
2018-05-30	Award announced to collaboration (recommended change in title to add Fusion Energy).

“Fusion for a 5 Year Old”

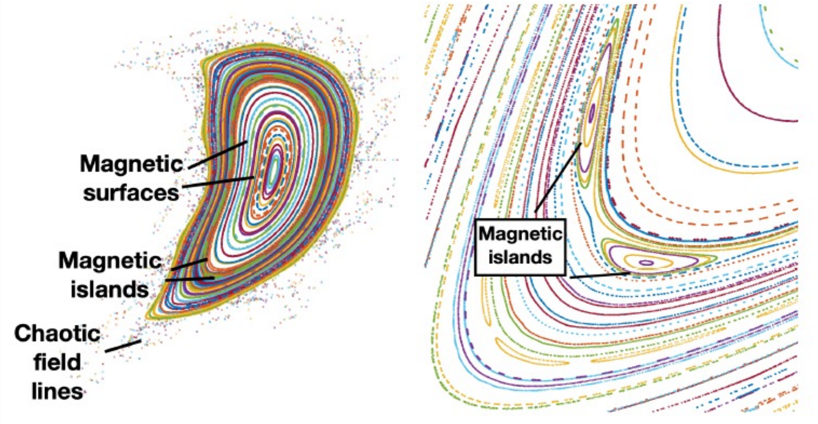
“Fusion for a 5 Year Old”

> *At the risk of sounding like a broken record, I will lobby for the addition of a paragraph in the introduction of the proposal that describes magnetically confined fusion as if it were being explained to a five year old.*
> *– Mike O’Neill (2018-02-07)*

> *“Adiabatic invariants of Hamiltonian mechanics” is well beyond the level of sophistication that should be included in the intro, in my opinion.*
> *– Response to a proposed revision (2018-02-08)*

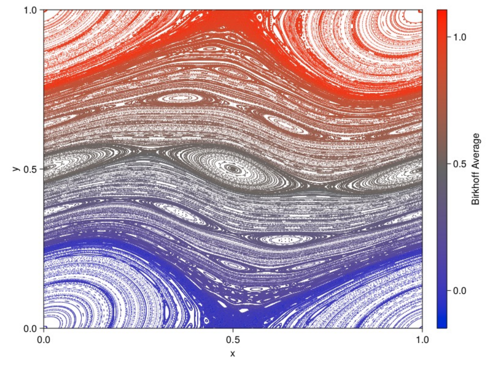
Ad: *Introduction to Stellarators* by Imbert-Gerard, Paul, Wright (recently published by SIAM)

Poincaré Features (NCSX)



“An Introduction to Stellarators”
Imbert-Gerard, Paul, and Wright.

Digression: A Non-Stellarator Test Problem

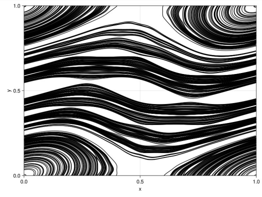


Illustrate with standard (Chirikov-Taylor) map

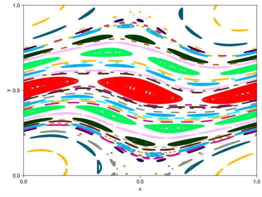
$$x_{t+1} = x_t + y_{t+1} \mod 1$$
$$y_{t+1} = y_t - \frac{0.7}{2\pi} \sin(2\pi x_t)$$

Plan in Pictures

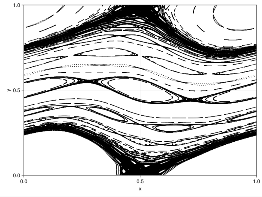
Circles



Islands



Chaos



- Iterating gives a Poincaré plot showing
  - X and O points (hyperbolic and elliptic periodic points)
  - Invariant circles and island chains (quasiperiodic orbits)
  - Chaos
- Goal: Identify these structures cheaply and automatically

Processing Poincaré Plots

1. Make a Poincaré plot and eyeball it
2. Parameterization method
3. Form a function with invariant level sets
    - Birkhoff averaging
    - Weighted Birkhoff averaging
    - Adaptive weighted Birkhoff (*)
    - Learned labels (*)
4. Model dynamics for a field line (*)

Parameterization method

Goal: $z : \mathbb{T} \to \mathbb{R}^2$ s.t.

$$F(z(\theta)) = z(\theta + \omega).$$

Discretize via Fourier:

$$\hat{z}(\theta) = \sum_{n=-m}^{m} \hat{z}_n \exp(2\pi i n \theta)$$

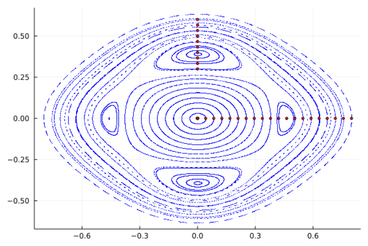
Solve nonlinear least squares problem

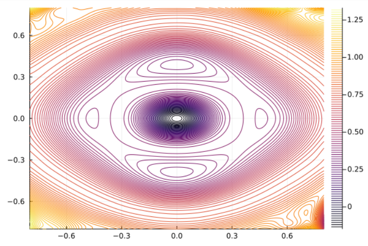
$$\min \sum_{i=0}^{N-1} \|z(i/N) - F(z(i/N + \omega))\|^2$$

with two additional constraints (phase + which circle).

Usually combine with continuation (e.g. from fixed point of F).

Learned Labels





Goal: Find (non-constant) h s.t. $h \circ F = h$.

Discretize via favorite ansatz, e.g. $h = \sum_{j=1}^{m} c_j \phi(\|x - x_j\|)$.
Define $h(x_j) = y_j$ and $h(F(x_j)) = y_j'$, solve (for example)

$$\text{minimize } \frac{\eta}{2} y^T K^{-1} y + \frac{1}{2}\|y - \tilde{y}\|^2 \text{ s.t. } y_i = y_i'$$

to encourage h smooth, non-constant, invariant under F.

Ruth and Bindel, `https://arxiv.org/abs/2312.00967`

Birkhoff Average

Consider $f : \Omega \to \Omega$ symplectic, $h \in \mathcal{C}^\infty(\Omega)$

Define *Birkhoff average*:

$$\mathcal{B}_K[h](x) = \frac{1}{K+1}\sum_{k=0}^{K}(h \circ F^k)(x).$$

Birkhoff-Khinchin: for $h \in \mathcal{L}^1$, converges a.e. to conditional expectation of an invariant measure on an invariant set.

Error behavior $\mathcal{B}_K[h](x) - \bar{h}(x)$?

- Invariant circle/island? $O(K^{-1})$
- Chaos? $O(K^{-1/2})$

Rates signal regular vs chaotic ("stochastic") trajectories.

Birkhoff Average

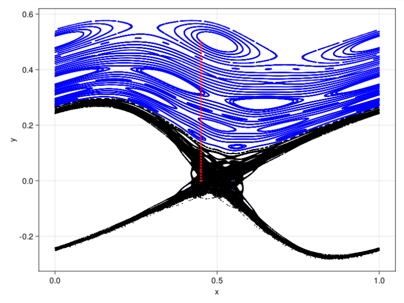
Ideas:

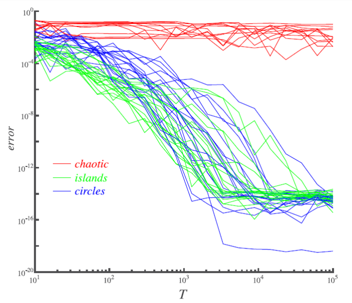
- Invariant sets as level sets of Birkhoff average
- Convergence rates as signal of regularity vs chaos

Converges in the long run – but in the long run, we are all dead. (with apologies to Keynes)

Related: Learn a continuous, nonconstant $\bar{h}$ s.t. $\bar{h} = \bar{h} \circ F$. Can do pretty well with kernel interpolation ansatz – a topic for another talk.

Weighted Birkhoff average





Sander and Meiss, Physica D, 411 (2020) p. 132569;
Das, Sander, and Yorke, Nonlinearity, 30 (2017), pp. 4111-4140

Weighting accelerates regular convergence to super-algebraic:

$$\mathcal{WB}_K[h](x) = \sum_{k=0}^{K} w_{k,K}(h \circ F^k)(x).$$

Interlude: Signals and Systems

Shift operator maps sequences to sequences:

$$(Sz)_t = z_{t+1}$$

Eigenfunctions are λ^t (generalized are $t^k \lambda^t$).

Gives solutions to constant coefficient linear difference equations:

$$(p(S)z)_t = \sum_{j=0}^{d} c_j z_{t+d-j} = 0$$

$$\implies z_t = \sum_{j=1}^{d} \alpha_j \lambda_j^t, \quad p(\lambda_j) = 0$$

(or a little more complex if p has multiple roots.)

Interlude: Signals and Systems

Consider a signal of the form

$$z_t = z_* + \sum_{j=1}^{m} \alpha_j \lambda^j.$$

Apply a *filter* associated with a polynomial

$$p(w) = \sum_{j=0}^{d} c_j w^j$$

Then

$$(p(S)z)_t = p(1)^t z_* + \sum_{j=1}^{m} \alpha_j p(\lambda_j)^t.$$

To find z_* make $p(1) = 1$ and $|p(\lambda_j)|$ small.

Signal Processing Perspective

Parameterize $z(\theta)$ for invariant circle

$$F(z(\theta)) = z(\theta + \omega), \quad z(\theta) = \sum_{n \in \mathbb{Z}} \hat{z}_n \exp(2\pi i n \theta)$$

Trajectory $z_t = z(\omega t)$ has series expansion

$$z_t = \sum_{n \in \mathbb{Z}} \hat{z}_n \xi^{nt}, \quad \xi = \exp(2\pi i \omega)$$

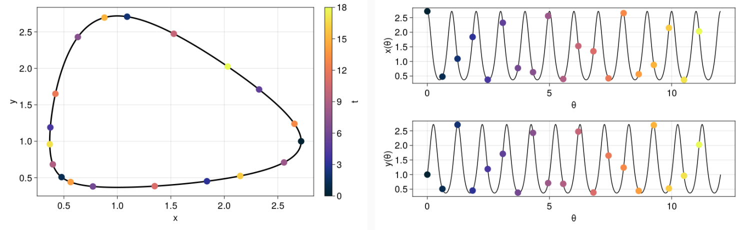
Observables $h_t = h(z_t)$ can be similarly expanded

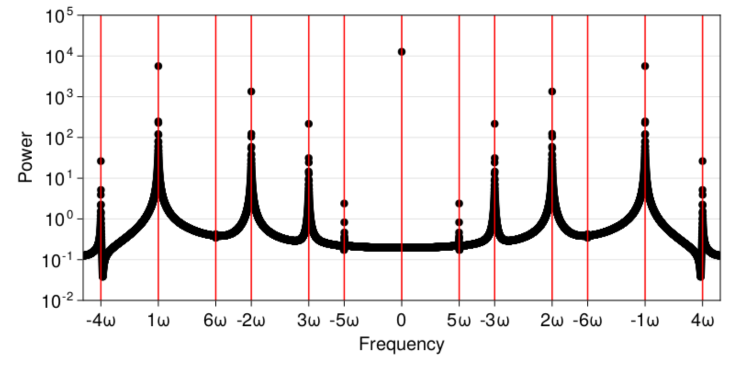
$$h_t = \sum_{n \in \mathbb{Z}} \hat{h}_n \xi^{nt}, \quad \bar{h} = \hat{h}_0$$

Weighted Birkhoff starting from x_0

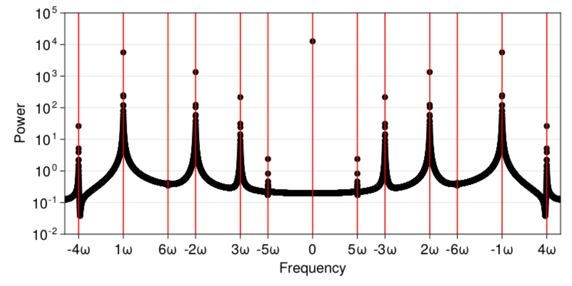
$$\mathcal{B}_K[h](x_0) = \sum_{n \in \mathbb{Z}} \hat{h}_n p_K(\xi^n), \quad p_K(z) = \sum_{k=0}^{K} w_{k,K} z^k$$

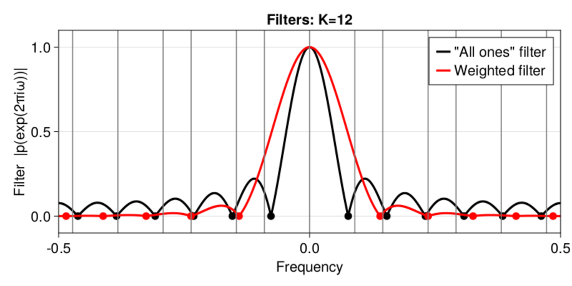
Signal Processing Perspective



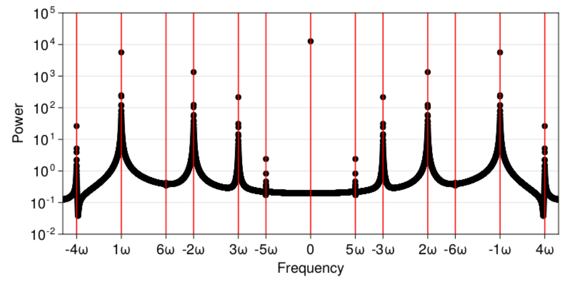


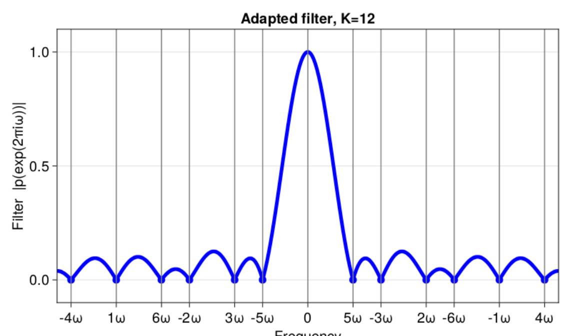
Signal Processing Perspective





Signal Processing Perspective: Adaptive Filtering





Adaptive Filtering

Series for $h_t = h(z_t)$

$$h_t = \sum_{n \in \mathbb{Z}} \hat{h}_n \xi^{nt}$$

Filtered/accelerated series with polynomial p_K:

$$\mathcal{AWB}_K[h](x_t) = \sum_{n \in \mathbb{Z}} \hat{h}_n \xi^{nt} p_K(\xi^n) \to \hat{h}_n$$

How do we adaptively choose the filter polynomial?

Desiderata for this to work:

- Fast enough decay of $\hat{h}_n$
- “Sufficiently irrational” ω (Diophantine condition)

(Vector) Reduced Rank Extrapolation

Assume

$$h_t = \hat{h}_0 + \sum_{n \neq 0} \lambda_n^t \quad \text{(e.g. } \lambda_n = \xi^n\text{)}$$

Difference sequence removes mean:

$$u_t = h_{t+1} - h_t = \sum_{n \neq 0} (\lambda_n - 1)\hat{h}_m \lambda_m^t$$

Seek coeffs c_k to minimize

$$\sum_{t=0}^{T-1} \left(\sum_{k=0}^{K} c_k u_{k+t} \right)^2 \text{ s.t. } \sum_{k=0}^{K} c_k = 1.$$

Accelerated series is

$$\tilde{h}_t = \sum_{k=0}^{K} c_k h_{k+t}.$$

Details

- Can (and do) use vector observables
- Rectangular Hankel matrix $\implies$ fast matvecs via FFT
- Solve least squares problem with LSQR
- Constrain for time reversibility $\implies$ palindromic polynomial:

  $$c_j = c_{K-j}$$

  Roots come in inverse pairs (generally on unit circle)
- Measure convergence adaptively via residual

(Vector) Reduced Rank Extrapolation

Standard vector RRE convergence (Sidi, *Vector Extrapolation Methods with Applications*): if $|\lambda_j|$ are in descending order, error for Kth extrapolated average goes like

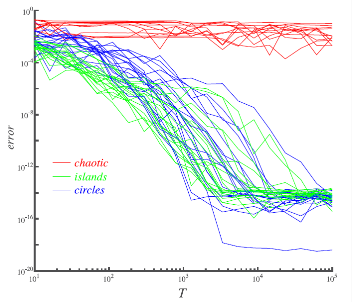
$$\hat{h}_{0,K} - \hat{h}_0 = O(\lambda_{K+1}^{2K}).$$

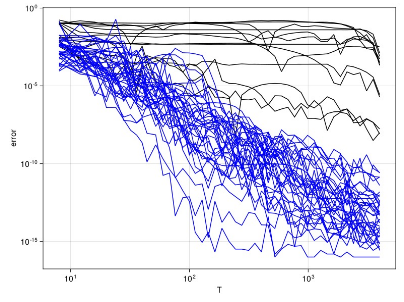
But for us everything is on the unit circle!

Alternate analysis gives super-algebraic convergence given

- Enough smoothness of circle (decay of $|\hat{h}_n|$ with $|n|$)
- "Sufficient irrationality" (Diophantine condition) so ξ_n doesn't get too close to 1 too fast.

Weighted Birkhoff vs RRE

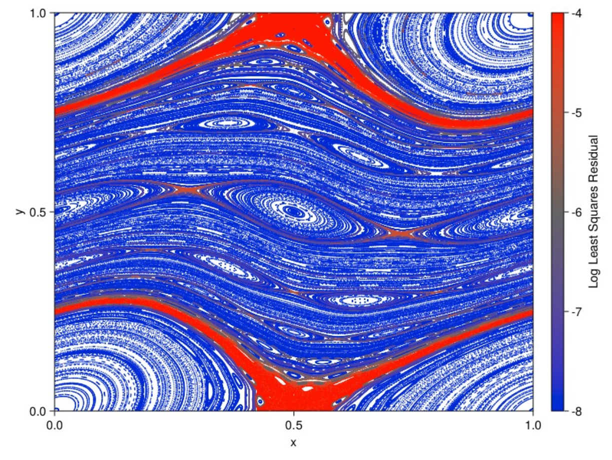




Still good for classification.convergence slightly faster than weighted Birkhoff.

Residuals and Regularity

Use least squares residual to judge “circleness.”



(Hard cases near rational rotational transform)

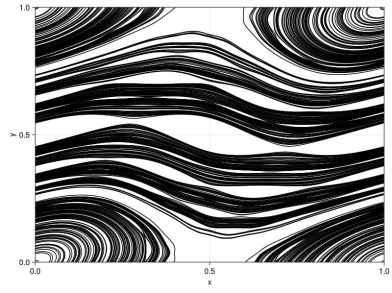
Post-Processing (Filter Diagonalization)

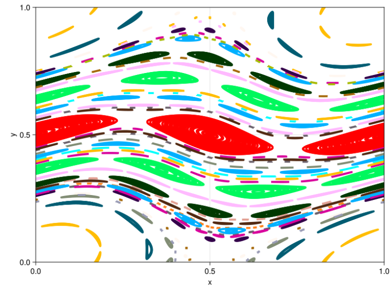
Why use the RRE model just for averaging?

1. Form filter polynomial with coefficients c
2. Find natural frequencies / polynomial roots
3. Sort by contribution to signal
4. Of 10 most contributing frequencies, identify rationals (Sander & Meiss)
    - Yes: island chain — RRE on qth step
    - No: call largest the rotational transform
5. Project signal onto Fourier modes

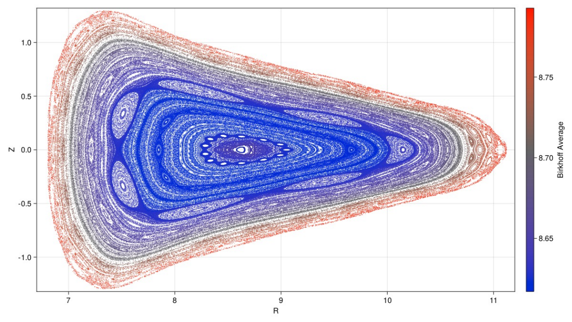
Get shape and characteristics of circles and islands.

Island Identification





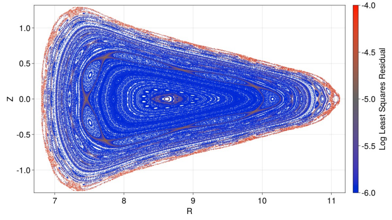
Wistell Stellarator Configuration



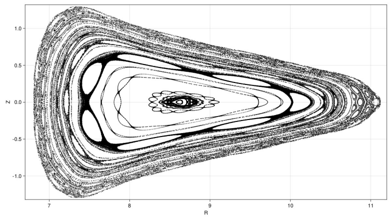
- 1000 random trajectories (via RK4 on interpolated B field)
- $K_{\text{max}} = 300, T_{\text{max}} = 900$
- Residual tolerance = 10^{-6}
- Rational tolerance = 10^{-6}

Wistell Analysis

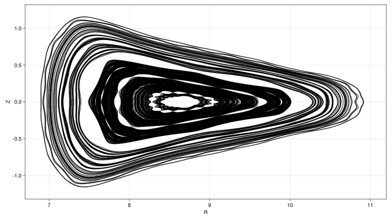
Residual



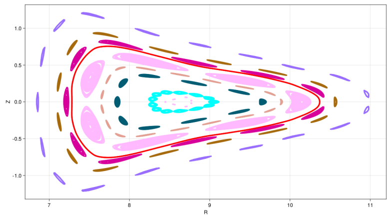
Chaos



Circles



Islands



Pros and Cons

- Extrapolation pros
  - Classifies chaos vs regular trajectories
  - Recovers invariant circles/islands
  - No need for continuation or initial guesses
  - Parallelizable over trajectories
- Cons
  - Problems near low-order rationals
  - Linear algebra adds extra cost vs weighted Birkhoff
- Higher dimensions?
  - Relevant beyond field line flow (guiding center approx)
  - Invariant sets are more complicated
  - The "model the trajectory" philosophy should still work

Ruth and Bindel, `https://arxiv.org/abs/2403.19003`