Toward a solution to a problem of Poincaré:
A Macro-analysis of geometric variation of high-dimensional dynamics

D. J. Albers

June 21, 2006
Roadmap

• discuss Poincaré’s vision to qualitatively study nature;

• discuss practical difficulties with this vision;

• outline a framework to resolve these difficulties — identification of sufficiently general function spaces endowed with measures;

• quantify variation in the geometric structure for a function space relative to a measure;
Poincare’s vision: Study nature via a qualitative geometric study of the space of all models, in particular, $C^r$ diffeomorphisms (discrete-time maps) and $C^r$ vector fields (ODEs)
Practical problems with Poincaré’s Vision

- Turbulence versus spatially extended dynamics
- Polynomials and coupled-map lattices
- Broken stability dream

Nature is extremely diverse
Core issue: the partitioning of *function spaces*

This can be complicated, but in spirit there are intrinsically three ways:

- **Abstract dynamics**: geometric conditions and assumptions (e.g. hyperbolicity) imply a defining property (e.g. ergodicity), then prove the conditions or assumptions are generic or dense in $C^r$; (here there exists no measure-theoretic notion, therefore no probabilistic language);

- **Experimental science**: perform an experiment that is repeatable; the act of performing an experiment intrinsically imposes a measure which partitions and focuses what is studied, the repeatability of the experiment implies a sort of persistence or stability;

- **Computational science**:
  - Traditional modeling using “rationalized models” of particular natural systems;
  - Monte Carlo studies of function spaces using joint and product measures on parameter space (this is what we do);
The language problem

- most abstract dynamics results are with respect to the $C^r$ Whitney topology — there is no notion of measure or probability, no “picking” mechanism to perform an experiment;

- numerical and traditional experiments all require and imply a measure;

- often measure-theoretic and topological notions of common yield conflicting results;

- the notion of prevalence, invented by Hunt, Sauer, Yorke, etc, is intended to address this problem, but it can be a difficult notion to use;
Toward a practical solution to Poincare’s problem

Poincare’s Dream

Nature–Literal Models

Abstract Dynamical systems

Abstract model space

\( P-I \)

\( P-II \)

\( P-III \)

\( P-IVa \)

\( P-IVb \)

\( D \)

\( MS^n \)

\( MS^a \)

\( NS \)

\( \mathcal{D} \)
Measurements and dynamics: discrete-time, time delay dynamical systems

$F$ is the dynamical system, $E : U \to R$ ($E$ is a $C^k$ map), where $E$ represents some empirical style measurement of $F$, and $g$ is the “Takens’s” map:

$$g(x_t) = (E(x_t), E(F(x_t)), \ldots, E(F^{2d}(x_t))) \quad (1)$$
Selection of a function space

Three characteristics:

- practical function space that can be used to model or reconstruct empirical results (i.e. it must be a discrete-time, time-delay dynamical system);

- the function space must admit a measure;

- the function space must be \textit{dense or prevalent} in the function spaces used to yield solutions to ODEs, PDEs, and general natural systems (e.g. $C^r$, Sobolev space, etc);
Artificial neural networks

\[ \Sigma(G) \equiv \{ \gamma : \mathbb{R}^d \to \mathbb{R} | \gamma(x) = \sum_{i=1}^{n} \beta_i G(\bar{x}^T \omega_i) \} \]  \hspace{1cm} (2)

Here, \( x \in \mathbb{R}^d \) is a \( d \)-vector of inputs, \( \bar{x}^T \equiv (1, x^T) \), \( n \) is the number of hidden units (neurons), \( \beta_1, \ldots, \beta_N \in \mathbb{R} \) are hidden-to-output layer weights, \( \omega_1, \ldots, \omega_N \in \mathbb{R}^{d+1} \) are input-to-hidden layer weights, and \( G : \mathbb{R}^d \to \mathbb{R} \) is the activation function (or neuron) with \( G \equiv \tanh() \);

\[ x_t = \beta_0 + \sum_{i=1}^{N} \beta_i G \left( s \omega_{i0} + s \sum_{j=1}^{d} \omega_{ij} x_{t-j} \right) \]  \hspace{1cm} (3)
Measure on neural networks

The probability measure on $\Sigma$: $\omega_{ij} \in N(0, s)$, $\beta_i$ uniform on $[0, 1]$, $x_t$ uniform on $[-1 : 1]$;

- each neural network can be identified by a point in the parameter space, $R^k$;

- imposing a measure on the parameter space imposes a measure on the space of neural networks $\Sigma(\tanh)$;

- $m_\beta \times m_\omega \times m_s \times m_I$ form a product measure on $R^k \times U$, this means the parameter are all uncorrelated;

- training the an ensemble of neural networks will impose a joint probability distribution on $R^k$, thus correlating the parameters;

- many imposed measures carve out manifolds directly in the parameter space, equivalence analysis can then be done in the space of measures (using Amari's information geometry);
Neural network approximation characteristics

Neural networks form a very diverse function space; they can approximate any $C^r$ mapping on compacta, they are dense in many Sobolev spaces used to solve ODEs and PDEs; neural networks are *universal approximators*;
Lyapunov exponents: a geometric diagnostic

- measurement or quantification of global expansion and contraction along an *orbit*;
- correspondence between positive (negative) Lyapunov exponents and global unstable (stable) manifolds;
- defines the global geometric structure of the attractor;
- independent of local coordinates or norm;
- calculated relative to a measure (physical, natural, SRB, Lebesgue, etc);
Stratification of the parameter space along a one dimensional interval: the $s$-parameter stratification

- existence of four “regions”
  - Region I: fixed point to first bifurcation
  - Region II: routes to chaos
  - Region IV: bifurcation chains (possibly turbulent-like, self-similar dynamics)
  - Region V: spatially-extended dynamics with intermittency, a transition to finite state dynamics
Example of the $s$-partition
Prototypical picture of a single, chaotic network, given the measure imposed on the weights
Bifurcation chains structure

\[ U = \{ a_i \} \]
\[ V = \bigcup_i V_i \]

\[ V_i = \text{bifurcation link sets}; \]
\[ V = \text{chain link sets}; \]
\[ U = \text{bifurcation chain sets}; \]
Two micro-geometric conjectures

Conjecture 1 (Existence of bifurcation chains) Assume $f_{s, \beta, \omega}$ with a sufficiently high number of dimensions, $d$. There exists at least one bifurcation chain subset $U$. 
Conjecture 2 (Characterization of geometric variation on the bifurcation chain subset) Assume \( f_{s,\beta,\omega} \) with a sufficiently high number of dimensions, \( d \), and a bifurcation chain set \( U \) as per conjecture (1). The two following (equivalent) statements hold:

i. In the infinite-dimensional limit, the cardinality of \( U \) will go to infinity, and the length \( \max |a_{i+1} - a_i| \) for all \( i \) will tend to zero on a one dimensional interval in parameter space. In other words, the bifurcation chain set \( U \) will be \( a-\)dense in its closure, \( \overline{U} \).

ii. In the asymptotic limit of high dimension, for all \( s \in U \), and for all \( f \) at \( s \), an arbitrarily small perturbation \( \delta_s \) of \( s \) will produce a topological change. The topological change will correspond to a different number of global stable and unstable manifolds for \( f \) at \( s \) compared to \( f \) at \( s + \delta \).

It means, as \( d \to \infty \), there will be an \( s \) interval for such the length of the bifurcation chain sets shrinks, this implies at arbitrarily small \( s \)-perturbations will produce topological change;

It is sort of “ugly” and complicated;
Necessary properties for the micro-geometric arguments

i. The following condition must be reasonably true: given the map $f_{s, \beta, \omega}$, if the parameter $s \in R^1$ is varied continuously, then the Lyapunov exponents vary continuously;

ii. The number of positive LCEs increases with dimension;

iii. The length of the $U_i$'s must decrease in a relatively uniform way as the dimension is increased;

iv. The LCEs that are positive are unimodal;
“Observational” properties on a open set in parameter space

(a) lack of periodic windows with respect to \((s, \beta, \omega)\);

(b) LCEs vary continuously with \(s\);

(c) they have a single maximum (up to statistical fluctuations);

(d) \(f_{s,\beta,\omega}\) has SRB measure(s) that yields a distribution of LCEs whose variance obeys \(\sigma^2_{\chi_i} < \inf_{j=\pm 1}|\chi_i - \chi_j|\) at fixed \(s\);

(e) as \(d\) increases, the length of the \(s\)-intervals, denoted \(U_i\), between LCE zero-crossings decreases as \(\sim d^{-1.92}\);

(f) the maximum number of positive LCEs increases monotonically as \(d/4\) and the attractor’s Kaplan-Yorke dimension scales as \(d/2\);
Persistence chaos conjecture

**Conjecture 3 (Persistent chaos in high dimensions)** Given $f_{s, \beta, \omega}$, if $k$ and $d$ are large enough, the probability with respect to $m_\beta \times m_\omega$ of the set $(\beta, \omega)$ with the properties (a)-(f) is large and approaches 1 as $k, d \rightarrow \infty$. 
Definition 1 (Degree-\(p\) Persistent Chaos) Assume a map \(f_\xi : U \to U\) (\(U \subset \mathbb{R}^d\)) that depends on a parameter \(\xi \in \mathbb{R}^k\). The map \(f_\xi\) has chaos of degree-\(p\) on an open set \(O \subset U\) that is persistent for \(\xi \in A \subset \mathbb{R}^k\) if \(\exists\) a neighborhood \(\mathcal{N}\) of \(A\) such that \(\forall \xi \in \mathcal{N}\), the map \(f_\xi\) retains at least \(p \geq 1\) positive LCEs Lebesgue a.e. in \(O\).
Macro-geometric variation: counting the number of positive Lyapunov exponents versus parameter variation, $M(s)$
Macro-geometrical variation

What is gained?

- no need for continuity of LCEs with respect to parameter variation;
- completely ignore the variation in the LCEs with parameter variation with the exception of sign changes;
- the characterization of the geometry is much more simple and based on much less restrictive assumptions with nearly no loss of information;
Macro-geometric quantification

For a particular neural network:

\[
M_{f,s,\beta,\omega}^r(s) = \sum_{i=1}^{d} \nu(\chi_i(s))
\]  
(4)

where \( \nu(\chi_i(s)) = 1 \) if \( \chi_i > 0 \), and 0 otherwise;

For an ensemble, \([M_{f,s,\beta,\omega}^r(s)]_{i \in I}\):

\[
M(s) = E[M_{f,s,\beta,\omega}^r(s)]_{i \in I}
\]  
(5)

Standard deviation: \([M_{f,s,\beta,\omega}^r(s)]_{i \in I}\) as \(\sigma_M\).

Curve fit to \(M(s)\): \(\mathcal{M}(s)\)

(Tildes denote rescaled coordinates)
Game plan for macro-geometric analysis

- find a universal scaling for $M(s)$ independent of $n, d$;

- fit the rescaled curve (using a rational function);

- blow up the rescaled curve to study the geometric variation as $n$ and $d \to \infty$;
• \( M(s) \) scaling in \( n \) and \( d \):

\[
M_{\text{max}}(s) = 0.11n^{0.37}d^{0.84}
\]  

(6)

• \( s \) is rescaled to \( \tilde{s} = s\sqrt{d} \).
Rescaling of $\tilde{M}(s)$ (and $\tilde{M}(s)$)

\[ M = \frac{s - s_{oc}}{-0.02 + 0.53 s + 0.07 s^2} \]
Considering the various plots of $M(s)$, the fitting function $M(s)$ must satisfy the following properties at $s_{oc}$, $s_{M_{\text{max}}}$, and $s_{ip}$:

i. $0 < s_{oc} < s_{M_{\text{max}}} < s_{ip}$;

ii. $s_{oc}$ such that $M(s_{oc}) = 0$ with $\frac{dM}{ds}(s_{oc}) > 0$;

iii. $s_{M_{\text{max}}}$ such that $M(s_{M_{\text{max}}}) = \max(M(s))$ for all $s > 0$;

iv. $s_{ip}$ such that $\frac{d^2M}{ds^2}(s_{ip}) = 0$;

Less precisely, $M$ needs to have a zero at $s_{oc}$ and be unimodal for $s > s_{oc}$; it is not an oversight that we did not specify another $s > s_{ip}$ value such that $M$ is zero, this is because numerical analysis of neural networks for very large $s$ values is a disaster.
$M(s)$ fitting

Rational function representation of $\tilde{M}(s)$:

$$\tilde{M}(\tilde{s}) = \frac{\tilde{s} - \tilde{s}_{oc}}{a_0 + a_1\tilde{s} + a_2\tilde{s}^2} \quad (7)$$

Mean geometric variation:

$$\tilde{\Gamma} = \frac{d\tilde{M}}{d\tilde{s}} = \frac{1}{a_0 + a_1\tilde{s} + a_2\tilde{s}^2}(1 - \frac{(\tilde{s} - \tilde{s}_{oc})(a_1\tilde{s} + 2a_2\tilde{s})}{a_0 + a_1\tilde{s} + a_2\tilde{s}^2}) \quad (8)$$

The fit produced $a_0 = -0.02$, $a_1 = 0.53$, and $a_2 = 0.0732$, yielding:

$$\tilde{M}_{n=32}(s) = \frac{\tilde{s} - 0.53}{-0.02 + 0.532\tilde{s} + 0.0732\tilde{s}^2} \quad (9)$$

$$\tilde{\Gamma}_{n=32} = \frac{1}{-0.02 + 0.532\tilde{s} + 0.0732\tilde{s}^2}(1 - \frac{(\tilde{s} - 0.016)(1.38 + (2)(0.1875)\tilde{s})}{(-0.02 + 0.532\tilde{s} + 0.0732\tilde{s}^2)^2}) \quad (10)$$
Recall $\tilde{M}(s)$ and $\tilde{m}(s)$

$$M = (s - s_{oc})/(-0.02 + 0.53 s + 0.07 s^2)$$
Intuition: “Whitney-like” picture of the ensemble

$M(s)$ with the standard deviation of $M(s)$, $M_{\text{max}}(s)$ and $M_{\text{min}}(s)$ for ensembles of networks with $n = d = 128$ and $n = d = 16$. 
$M(s)$ argument outline

• show $|U|$ increases monotonically with $d$;

• show the mean geometric variation (on $U$) increases with $d$;

• show the mean length of $V_k$’s decreases on $U$, this defines the type of geometric variation — the bifurcation chains structure;
Asymptotic length of (crudely defined) bifurcation chains region, $|U_k| = |a_1 - a_k|$

- mean length of the bifurcation chain subset $|U_k| = |a_1 - a_k|$ ($a_1 = s_{oc}$ and $a_k = s_{ip}$) with increasing dimension for $n = 32$ and $n = d$; as the dimension is increases, the mean and standard deviation of $|U_k|$ for $s \in [0.1 : 10]$ tend toward the full length of the interval;

- $\bar{s}_{ip} \approx 4.89$, it is likely that a more accurate cutoff would be $\approx 10$;

- $0 < s_{oc} < 1$ and $s_{ip} > 1$ where both scale like $d^{1/2}$, thus $|U_k|$ will increase like $|s_{ip} - s_{oc}|\sqrt{d}$ ($4.36\sqrt{d}$ in particular), thus the length of the bifurcation chains region increases;
Mean rate of geometric variation, $\Gamma(s) = \frac{dM}{ds}$

Left plot: both $\tilde{M}$ and $\frac{d\tilde{M}}{ds}$ with a vertical line drawn at $s_{ip}$

Right plot: $\frac{d\tilde{M}}{ds}$ in $s$ coordinates for $d = 100$ and $d = 1000$; the $d = 1000$ (versus the $d = 100$) graph is transformed up by $0.11d^{0.84}$ in the $y$-coordinate while it is transformed down by $d^{-1/2}$ in the $x$-coordinate, therefore $\frac{dM}{ds}$ increases monotonically with $d$ on $V = (s_{oc}, s_{ip})$;
Mean length of the chain link sets $V_k$:

Left plot: $|\frac{dM}{ds}|^{-1}$ versus $s$ for $d = 100$ and $d = 1000$; right plot: $|\frac{dM}{ds}|^{-1}$ simultaneously with $M(s)$ in the rescaled coordinates

$|V_k| = |s_{\chi_k-1} - s_{\chi_k}|$ not uniform as $d$ increases for all $s$; approximate these lengths by taking $\delta M \in \mathbb{N}$ where $\delta s$ is defined by increments of $\delta M$ yielding

$$|V_k| = \frac{\delta s}{\delta M - 1}$$  \hspace{1cm} (11)

As $d \to \infty$ in regions of $s$ where small changes in $s$ lead to large changes in $M$, approximate the length of $|V_k|$ with:

$$|V_k| \approx \left| \frac{ds}{dM} \right|$$  \hspace{1cm} (12)
Estimation of $p$

Estimate for $p$ is based on $M$:

$$p_M(s, \delta s) = M(s) - \left| \frac{dM}{ds}(s) \right| \delta s$$  \hspace{1cm} (13)

Conservative estimate of $p$ is provided by

$$p_{\text{min}}(s, \delta s) = \min [M^{f_{s,\beta,\omega}}(s)]_{i \in I}$$  \hspace{1cm} (14)

A more moderated empirical estimate of $p$ based on the mean and standard deviation of $M$

$$p_\sigma(s, \delta s) = M(s_{\text{min}}) - \sigma_M(s_{\text{min}})$$  \hspace{1cm} (15)

where

$$s_{\text{min}} = \arg \min_{s \in S} M(s)$$  \hspace{1cm} (16)
Comparing $p$-estimates

Estimates of $p$ in accordance with Eqns. 13 - 14 for $n = d = 128$ (left plot) and $n = d = 64$ (right plot) with a radius $\delta s = 0.1$. 
New definition of bifurcation chains region

Definition 2 (Bifurcation chains region) Assume the mapping $f_{s,\beta,\omega}$ with a chain link set $(V)$. The mapping $f_{s,\beta,\omega}$ is said to have a **bifurcation chains region** if there exists an $s$-interval, denoted $V_{BC}$, with positive Lebesgue measure such that:

a. the probability (on $V_{BC}$ with respect to $m_\beta \times m_\omega \times m_s$) that $M > 0$ increases to unity as $d$ and $n$ approach infinity for $s \in V_{BC}$;

b. the mean of the length of all the bifurcation link sets $(V_k)$ in $V_{BC}$ decreases monotonically as as $d$ and $n$ approach infinity;

Conservative estimate: $a_1 = s_{oc}$, $a_k = s_{ip}$;
$M$ conjectures

Conjecture 4 (Persistence of $M$) Assume the mapping $f_{\beta,\omega,s}, M(s)$ as defined in Eq. 5, $M(s)$ that satisfied properties (i)-(iv). As $n$ and $d$ diverge to infinity, $M$ will converge to $\tilde{M}$ in rescaled coordinates and thus satisfy properties (i)-(iv) Lebesgue a.e. on $s$ where $M > 0$. Moreover, $\frac{\sigma M}{M}$ will decrease monotonically with increases in $d$.

Conjecture 5 (Existence of bifurcation chains) Assume the mapping $f_{\beta,\omega,s}, M(s)$ as defined in Eq. 5 and $M(s)$ that satisfied properties (i)-(iv). As $n$ and $d$ diverge to infinity the probability that there will exist an $s$-interval with positive Lebesgue measure for $f_{\beta,\omega,s}$ that corresponds to a bifurcation chains region approaches unity.
What is gained, what is lost

Gained:

- precise, quantifiable definition of the bifurcation chains interval;
- specification of the requirements for the bifurcation chains structure to persist; in particular the conditions for persistence of bifurcation chains are significantly weakened compared with previous results;

Lost:

- all control over the LCEs away from zero;
- no statement about open balls in parameter space;
- observations less precisely characterized (but with similar consequences);
Relationship to other conjectures

Bifurcation chains:

- weakening and generalization of the needed hypothesis of the micro-geometric analysis with the same overall conclusions;

Persistent chaos:

- $M$-conjecture implies property (a);
- $M$-conjecture says nothing about properties (b)-(d);
- $M$-conjecture quantifies property (e) (length of $U_k$’s);
- $M$-conjecture is constructed using property (f);
Summary

We:

- identified a construction where a function space can be studied relative to a measure;
- defined a non-restrictive tool \( M(s) \) for characterizing geometric variation for an ensemble of mappings;
- quantified a geometric structure (bifurcation chains) that is existent in high-dimensional dynamical systems and persists on an interval of parameter space;

**Conclusion:** for the construction we utilize (i.e. relative to the measure we impose), chaos becomes more persistent as the number of degrees of freedom are increased; this is due to the increasing number of unstable manifolds whose transition to stability is characterized by \( M(s) \);

Collaborators: J. P. Crutchfield (UC-Davis CSE), J. C. Sprott (UW-Madison Physics)