Recent results regarding geometric variation in high-dimensional dynamical systems

D. J. Albers

September 12, 2006

Roadmap

- discuss Poincaré's vision to qualitatively study nature;
- discuss practical difficulties with this vision: the dichotomy between the study of nonlinear science in traditional math and scientific communities;
- outline a framework to resolve these difficulties — identification of "sufficiently general" function spaces *endowed with measures*;
- present an application: quantification of the variation of 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 difficulties

- Concretely specifying difference: turbulence versus spatio-temporal dynamics (e.g., chaotic itinerancy)
- Relationships between examples and (complete) function spaces: example - the space of polynomials and coupled-map lattices
- Broken stability dream

Moral: nature is extremely diverse and difficult to specify

The language problem

- numerical and traditional experiments all require and imply a measure;
- commonality in most abstract dynamics results is specified with respect to the C^r Whitney topology — there is no notion of measure or probability, no "picking" mechanism to perform an experiment, instead the scope of study is narrowed using a priori geometric characteristics;
- often measure-theoretic and topological notions of common often 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;

Core difficultly: specifying a classification system, or the partitioning of function spaces

Elements of a solution: we need an scientifically minded, mathematical umbrella for the observed phenomena

- the method of unification *must* have measure-theoretic notions built in;
- the method must be able to work in conjunction with both infinite-dimensional function spaces (e.g., C^r) and finite-dimensional function spaces

Prevalence: a translation invariant "almost every" on infinite dimensional spaces

Definition 1 *Let X be a completely metrizable topological vector space and μ be a Borel measure. Given a (Borel) set $S \subset X$, we say that a* finite-dimensional *subspace $P \subset X$ is a* probe *for S provided that for all $x \in X$, μ a.e. point of the hyperplane $x + P \in S$.*

If a Borel set $S \in X$ has a probe, then S is said to be prevalent.

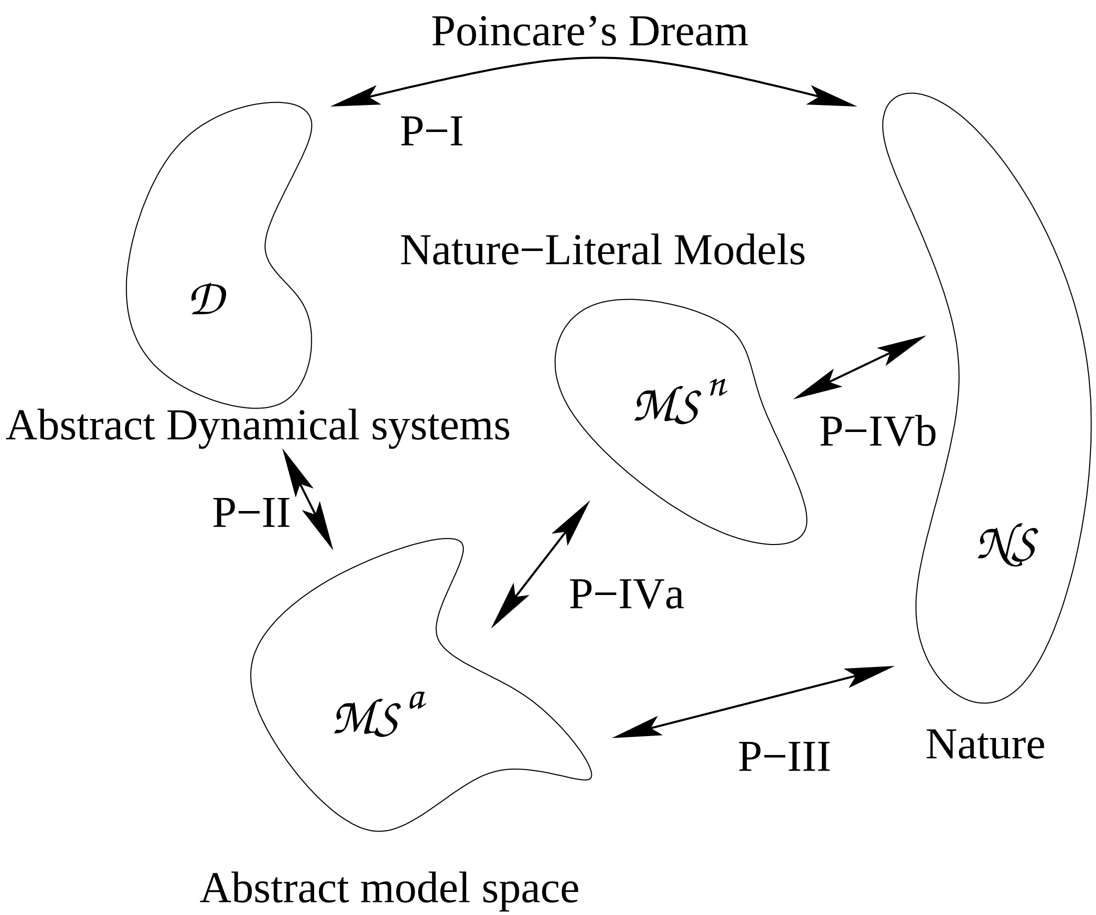
Ingredients: Topological vector space X (e.g., C^r), a Borel subset $S \subset X$, and a finite-dimensional *probe*, $P \subset X$.

Example: Nowhere differentiable functions

Theorem 1 *Almost every function in $C[0,1]$ is nowhere differentiable; that is, the nowhere differentiable functions form a prevalent subset of $C[0,1]$*

Ingredients: $X = C[0,1]$, $S = \{$nowhere Lipschitz functions $\}$, $P =$ two parallel planes.
Argument: $C[0,1]$ can be partitioned into parallel planes in such a way that in each plane, a.e. function (with respect to Lebesgue measure) is nowhere differentiable. The plane spanned by g and h forms the probe.

Toward a pratical solution to the specification problem



Selection of a function space

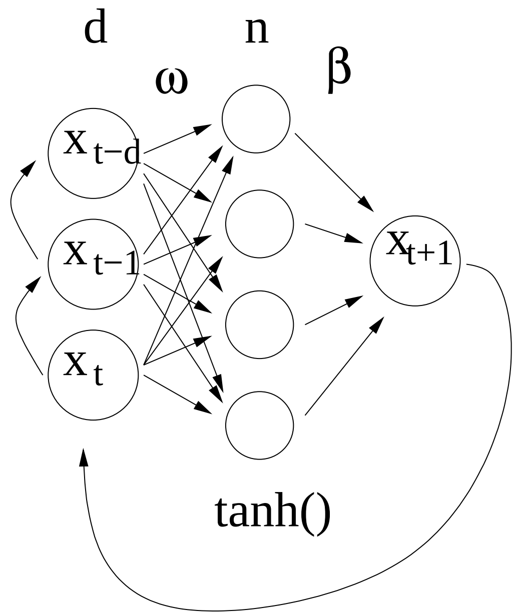
Three characteristics:

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

2. the function space must admit a *measure*;

3. the function space must be *dense* or *prevalent* in the function spaces used to yield solutions to ODEs, PDEs, and general natural systems (e.g. C^r, Sobolev spaces, etc);

Discrete-time, time-delay, feedforward, artificial neural networks



$$\Sigma(G) \equiv \{\gamma : R^d \rightarrow R | \gamma(x) = \sum_{i=1}^{n} \beta_i G(\tilde{x}^T \omega_i)\} \quad (1)$$

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

$$x_t = \beta_0 + \sum_{i=1}^{N} \beta_i G\left(s\omega_{i0} + s \sum_{j=1}^{d} \omega_{ij} x_{t-j} \right) \quad (2)$$

Measure on neural networks

The probability measure on ΣG: $\omega_{ij} \in N(0, s)$, β_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 parameters are uncorrelated;
- training 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*;

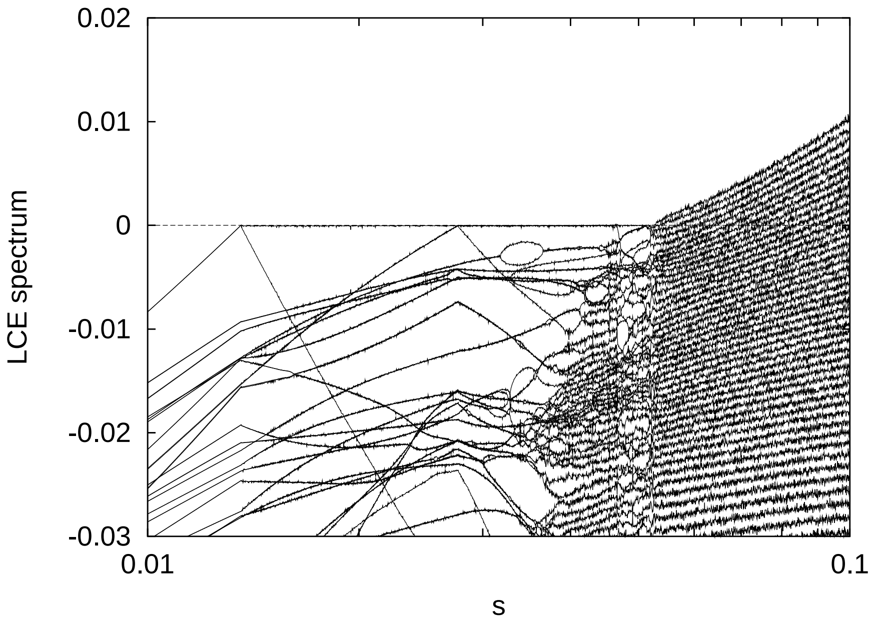
Lyapunov exponents: a geometric diagnostic

- 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);

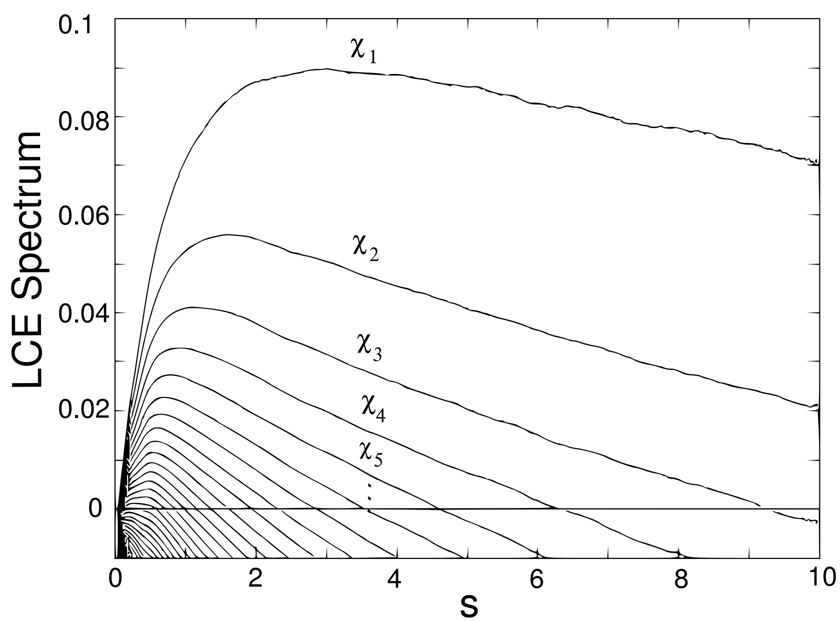
Stratification of the parameter space along a one dimensional interval: the s-parameter stratification

- 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: spatio-temporal dynamics with intermittency (chaotic itinerancy), a transition to finite state dynamics

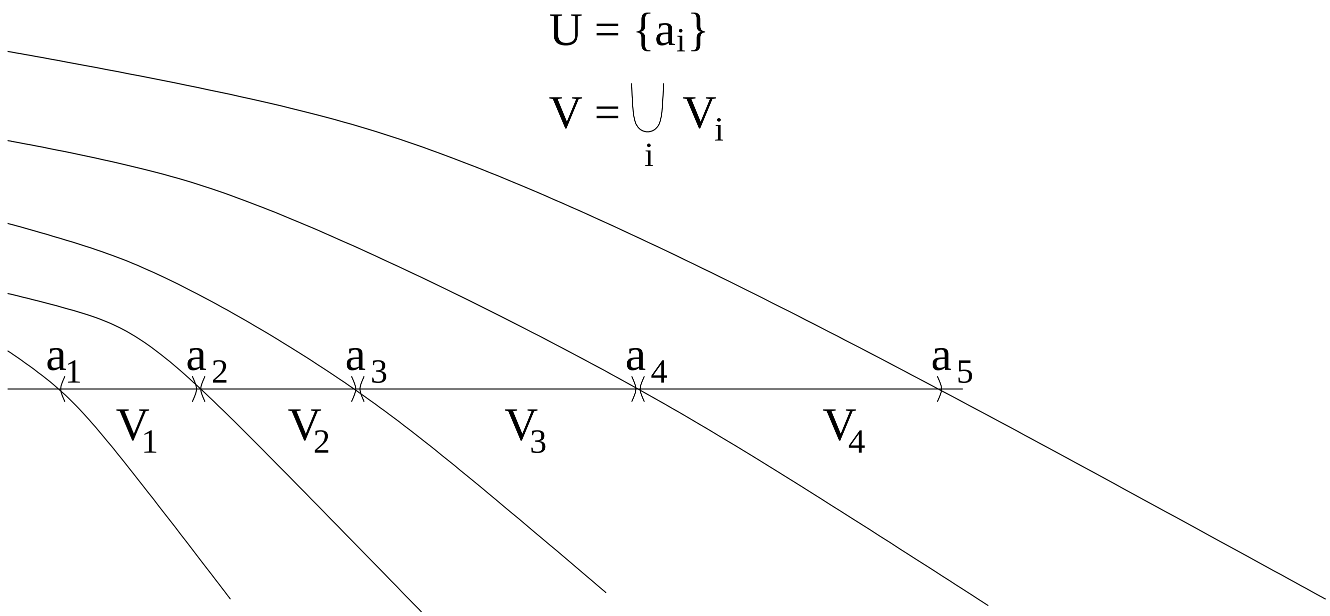
Example of the s-partition



Prototypical picture of a single, chaotic network, given the measure imposed on the parameters (weights)



Bifurcation chains structure



$V_i =$ *bifurcation link sets*;

$V =$ *chain link sets*;

$U =$ *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 δ_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

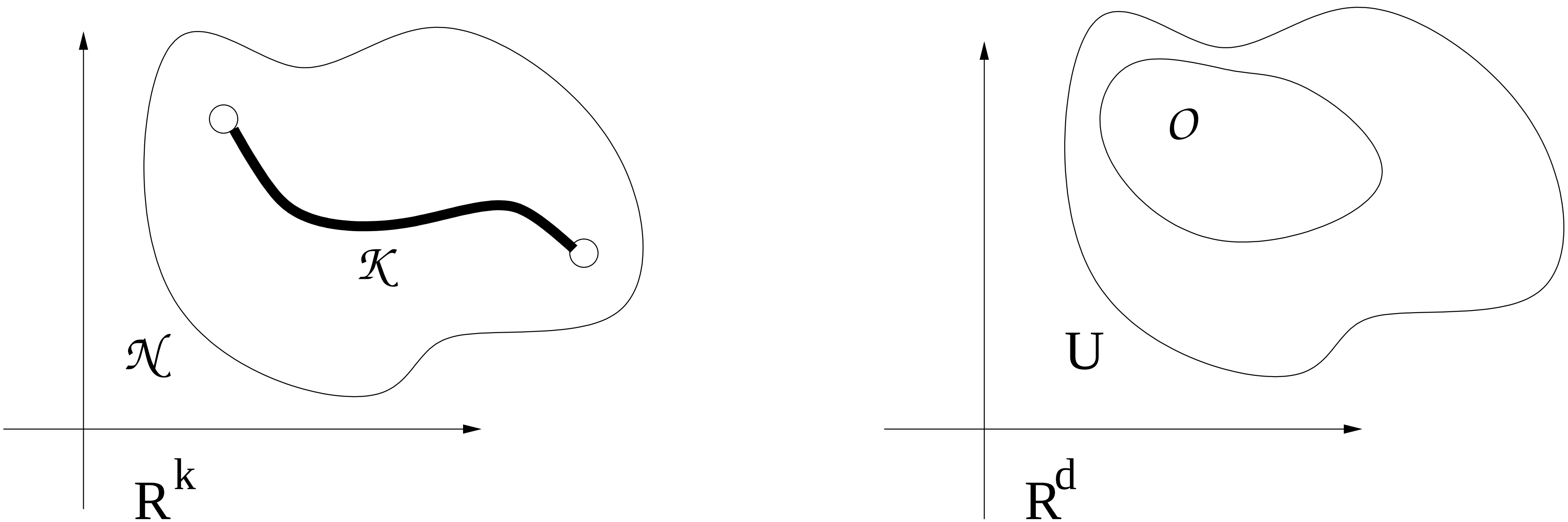
i. the following condition must be reasonably true: given the map $f_{s,\beta,\omega}$, if the parameter $s \in R^1$ is varied continuously, then the Lyapunov exponents vary continuously;

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 positive LCEs are unimodal;

Quantification of persistence chaos



Definition 2 (Degree-p Persistent Chaos) *Assume a map $f_\xi : U \to U$ ($U \subset R^d$) that depends on a parameter $\xi \in R^k$. The map f_ξ has chaos of degree-p on an open set $\mathcal{O} \subset U$ that is persistent for $\xi \in \mathcal{A} \subset R^k$ if $\exists$ a neighborhood $\mathcal{N}$ of $\mathcal{A}$ such that $\forall\ \xi \in \mathcal{N}$, the map f_ξ retains at least $p \geq 1$ positive LCEs Lebesgue a.e. in $\mathcal{O}$.*

"Observational" properties on a open set in parameter space

(a) lack of periodic windows with respect to (s, β, ω);

(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 $d/2$;

Persist 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 (β, ω) with the properties (a)-(f) is large and approaches 1 as $k, d \to \infty$.*

Macro-geometric *quantification*

For a *particular* neural network:

$$M^{f_{s,\beta,\omega}}(s) = \sum_{i=1}^{d} \nu(\chi_i(s)) \tag{3}$$

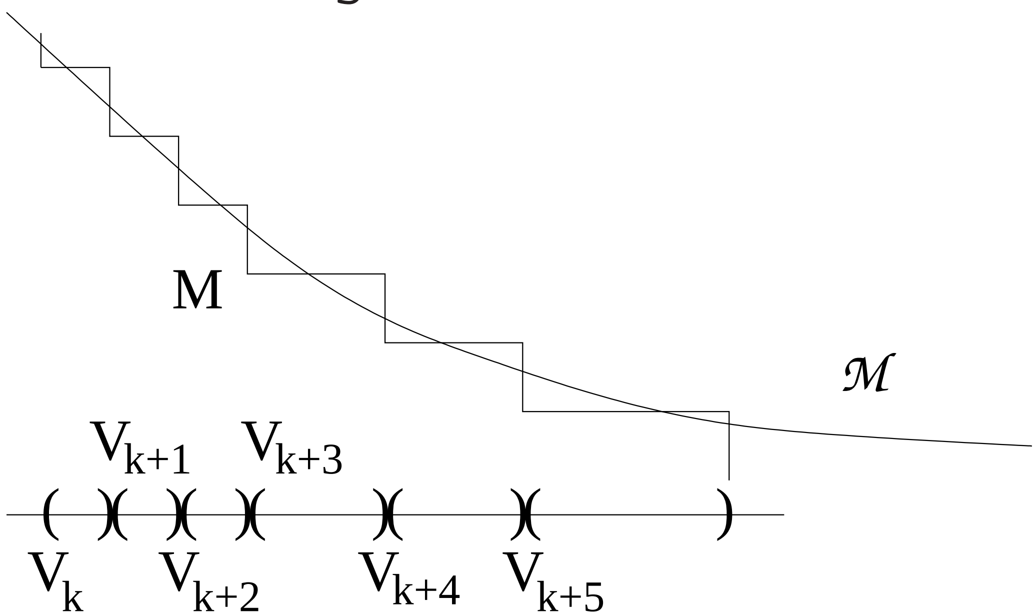
where $\nu(\chi_i(s)) = 1$ if $\chi_i > 0$, and 0 otherwise;

For an ensemble, $[M^{f_{s,\beta,\omega}}(s)]_{i \in I}$:

$$M(s) = E[M^{f_{s,\beta,\omega}}(s)]_{i \in I} \tag{4}$$

Standard deviation: $[M^{f_{s,\beta,\omega}}(s)]_{i \in I}$ as σ_M.

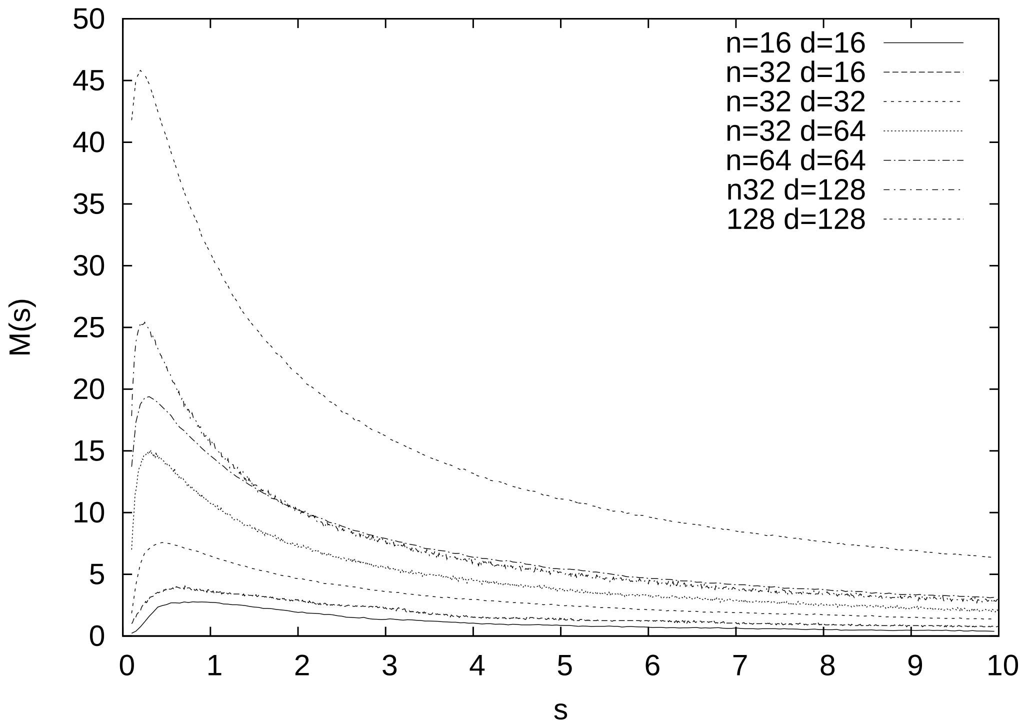
Macro-geometrical variation



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 variation: counting the number of positive Lyapunov exponents versus parameter variation, $M(s)$



Macro-geometric *quantification* with universal scaling

For a *particular* neural network:

$$M^{f_{s,\beta,\omega}}(s) = \sum_{i=1}^{d} \nu(\chi_i(s)) \tag{5}$$

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

For an ensemble, $[M^{f_{s,\beta,\omega}}(s)]_{i \in I}$:

$$M(s) = E[M^{f_{s,\beta,\omega}}(s)]_{i \in I} \tag{6}$$

Standard deviation: $[M^{f_{s,\beta,\omega}}(s)]_{i \in I}$ as σ_M.

Tildes denote rescaled coordinates

Curve fit to $M(s)$: $\mathcal{M}(s)$ in *rescaled* coordinates

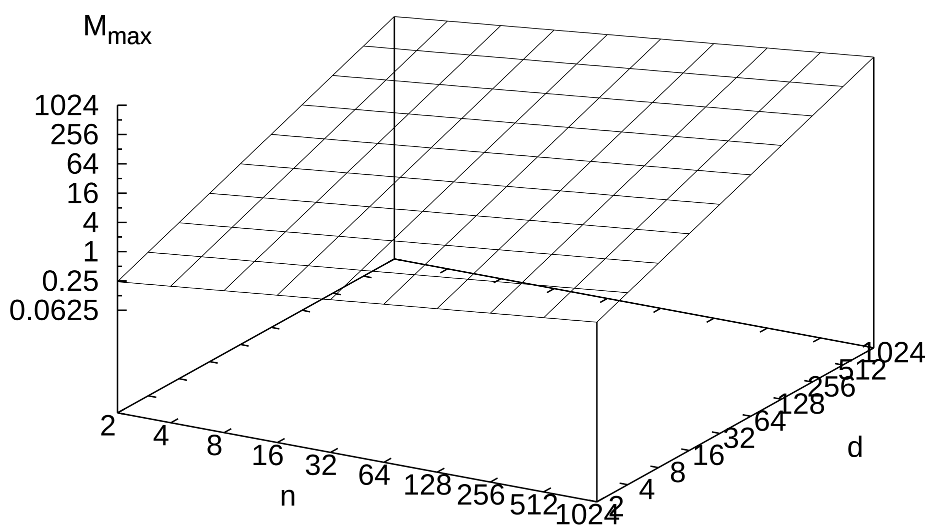
Game plan for macro-geometric analysis

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

2. fit the rescaled curve (using a rational function) $\tilde{\mathcal{M}}(s)$;

3. blow up the rescaled curve, $\tilde{\mathcal{M}}(s)$, to study the geometric variation as n and $d \to \infty$;

n and d peak rescaling

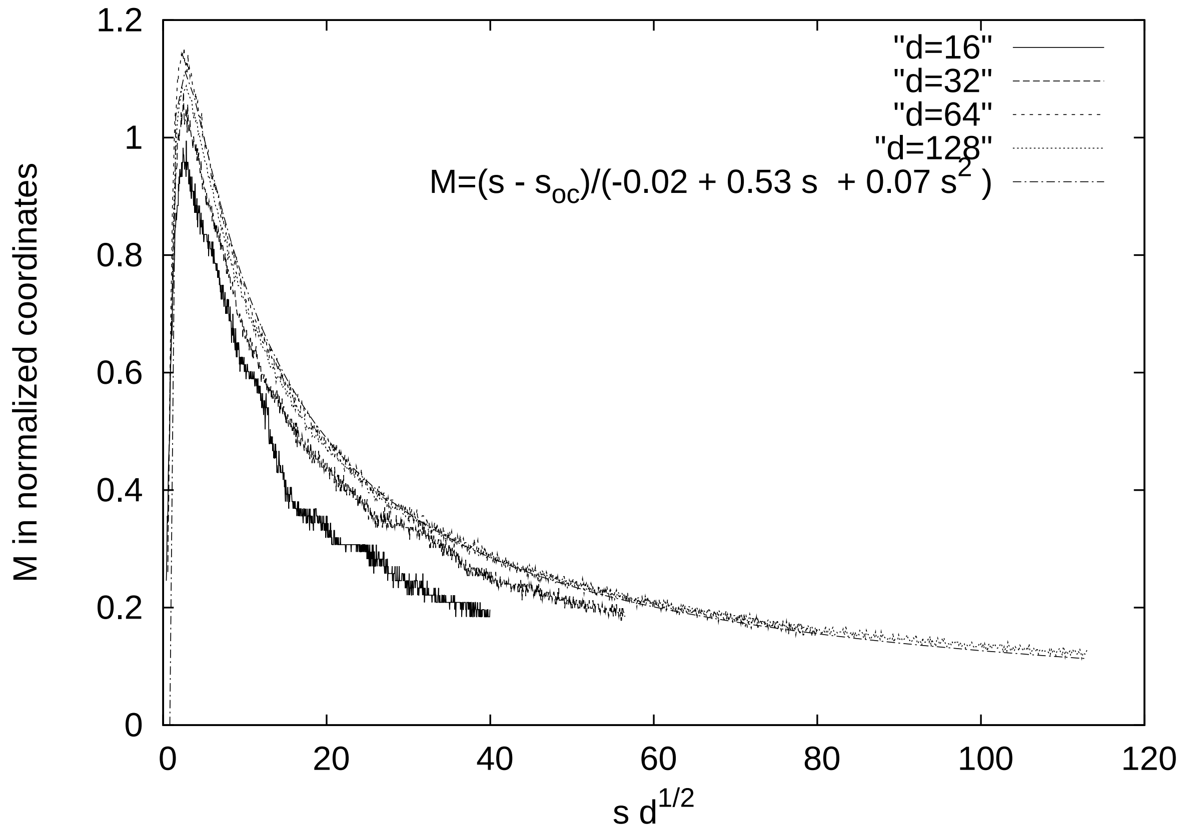


- $M(s)$ scaling in n and d:

$$M_{max}(s) = 0.11 n^{0.37} d^{0.84} \tag{7}$$

- s is rescaled to $\tilde{s} = s\sqrt{d}$

Rescaling of $\tilde{M}(s)$ (and $\tilde{\mathcal{M}}(s)$)



Considering the various plots of $M(s)$, the fitting function $\mathcal{M}(s)$ must satisfy the following properties at s_{oc}, $s_{\mathcal{M}_{max}}$, and s_{ip}:

i. $0 < s_{oc} < s_{\mathcal{M}_{max}} < s_{ip}$;

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

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

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

Less precisely, $\mathcal{M}$ 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 $\mathcal{M}$ 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 $\tilde{\mathcal{M}}(s)$:

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

Mean geometric variation:

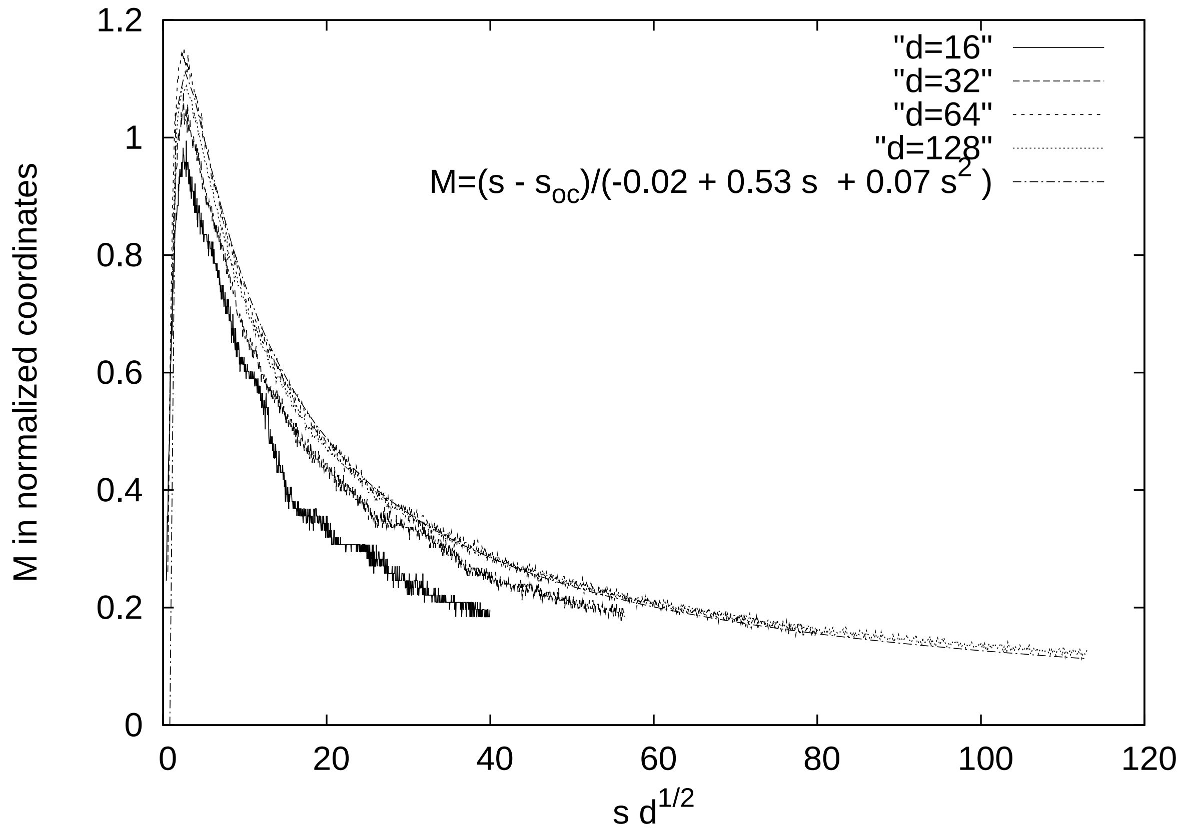
$$\tilde{\Gamma} = \frac{d\tilde{\mathcal{M}}}{d\tilde{s}} = \frac{1}{a_0 + a_1\tilde{s} + a_2\tilde{s}^2}(1 - \frac{(\tilde{s} - \tilde{s_{oc}})(a_1\tilde{s} + 2a_2\tilde{s})}{a_0 + a_1\tilde{s} + a_2\tilde{s}^2}) \quad (9)$$

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

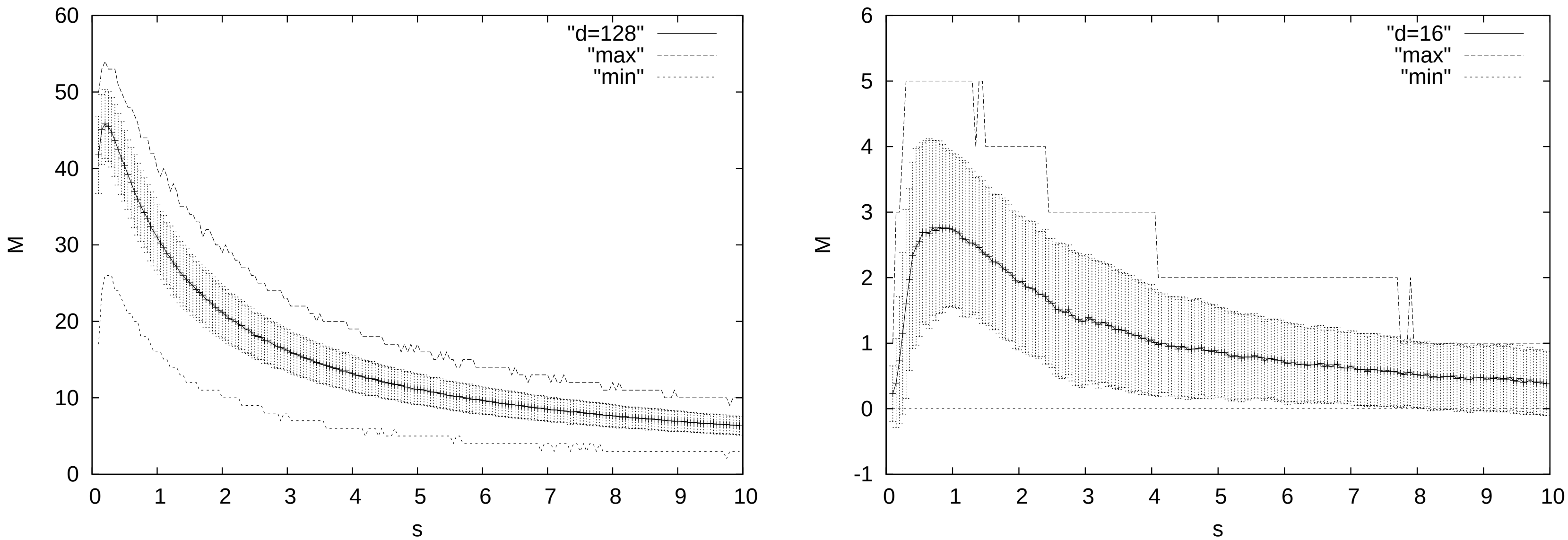
$$\tilde{\mathcal{M}}_{n=32}(s) = \frac{\tilde{s} - 0.53}{-0.02 + 0.532\tilde{s} + 0.0732\tilde{s}^2} \quad (10)$$

$$\tilde{\Gamma}_{n=32} = \frac{1}{-0.02 + 0.532\tilde{s} + 0.0732\tilde{s}^2}(1 - \frac{(\tilde{s} - 0.016)(1.38 + (2)(0.1875)\tilde{s})}{(-0.02 + 0.532\tilde{s} + 0.0732\tilde{s}^2)^2}) \quad (11)$$

Recall $\tilde{M}(s)$ and $\tilde{\mathcal{M}}(s)$



Representation of $M(s)$ by $\tilde{M}(s)$: "Whitney-like" picture for an ensemble



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

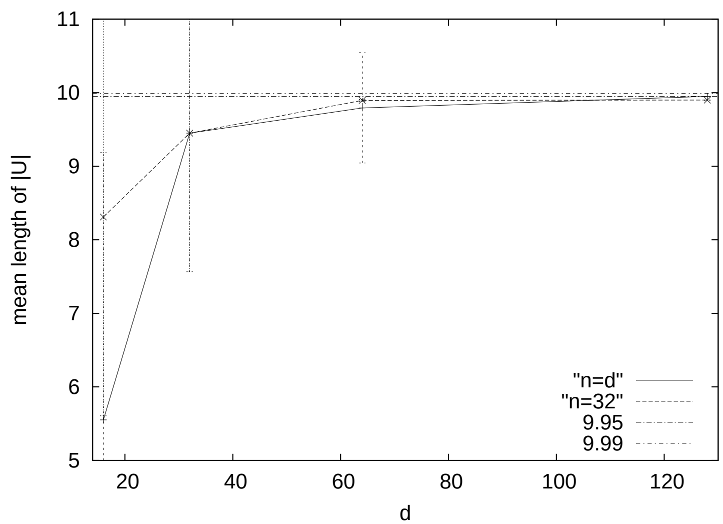
$M(s)$ argument outline

1. show $|U|$ increases monotonically with d;

2. show the mean geometric variation (on U) increases with d;

3. 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, $|U_k| = |a_1 - a_k|$

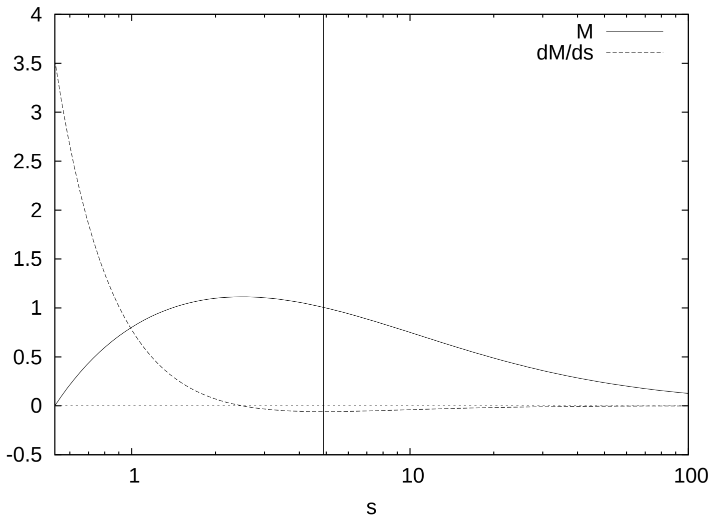


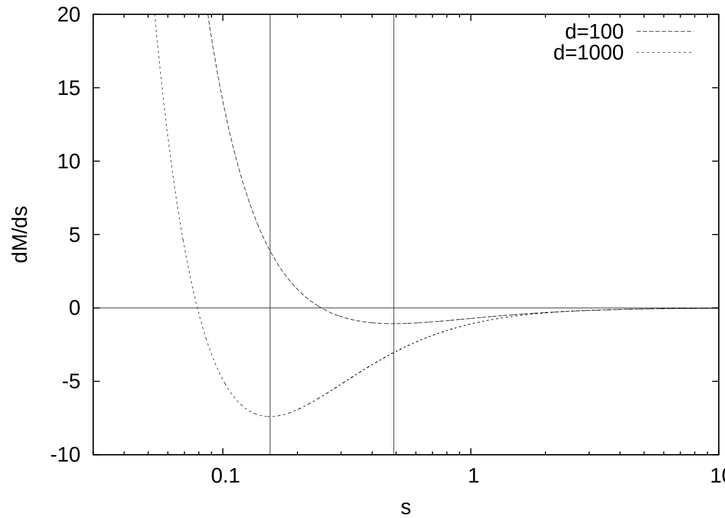
- mean length of the bifurcation chain subset $|U_k| = |a_1 - a_k|$ ($a_1 = s_{oc}$ and $a_k = s_{ip}$) with increasing dimension for $n = 32$ and $n = d$; as the dimension is increases, the mean and standard deviation of $|U_k|$ for $s \in [0.1 : 10]$ tend toward the full length of the interval;

- $\tilde{s}_{ip} \approx 4.89$, it is likely that a more accurate cutoff would be ≈ 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, $\Gamma(s) = \frac{dM}{ds}$

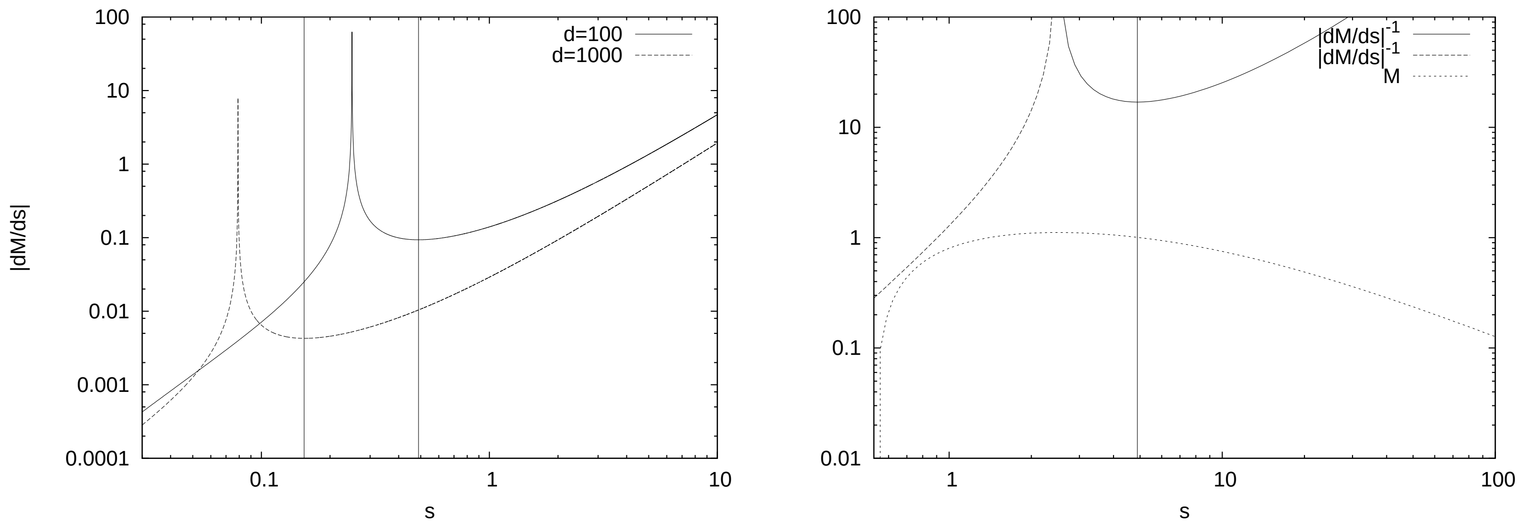




Left plot: both $\tilde{\mathcal{M}}$ and $\frac{d\tilde{\mathcal{M}}}{d\tilde{s}}$ with a vertical line drawn at s_{ip}

Right plot: $\frac{d\tilde{\mathcal{M}}}{d\tilde{s}}$ 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:$|\frac{d\mathcal{M}}{ds}^{-1}|$ versus s for $d = 100$ and $d = 1000$; right plot:$|\frac{d\mathcal{M}}{ds}^{-1}|$ 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 $\delta\mathcal{M} \in N$ where δs is defined by increments of $\delta\mathcal{M}$ yielding

$$|V_k| = \frac{\delta s}{\delta\mathcal{M} - 1} \tag{12}$$

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

$$|V_k| \approx |\frac{ds}{d\mathcal{M}}| \tag{13}$$

Estimation of the number of persistent positive exponents, p

Estimate for p is based on $\mathcal{M}$:

$$p_{\mathcal{M}}(s, \delta s) = \mathcal{M}(s) - |\frac{d\mathcal{M}}{ds}(s)|\delta s \quad (14)$$

Conservative estimate of p is provided by

$$p_{\text{min}}(s, \delta s) = \text{min}[M^{f_{s,\beta,\omega}}(s)]_{i \in I} \quad (15)$$

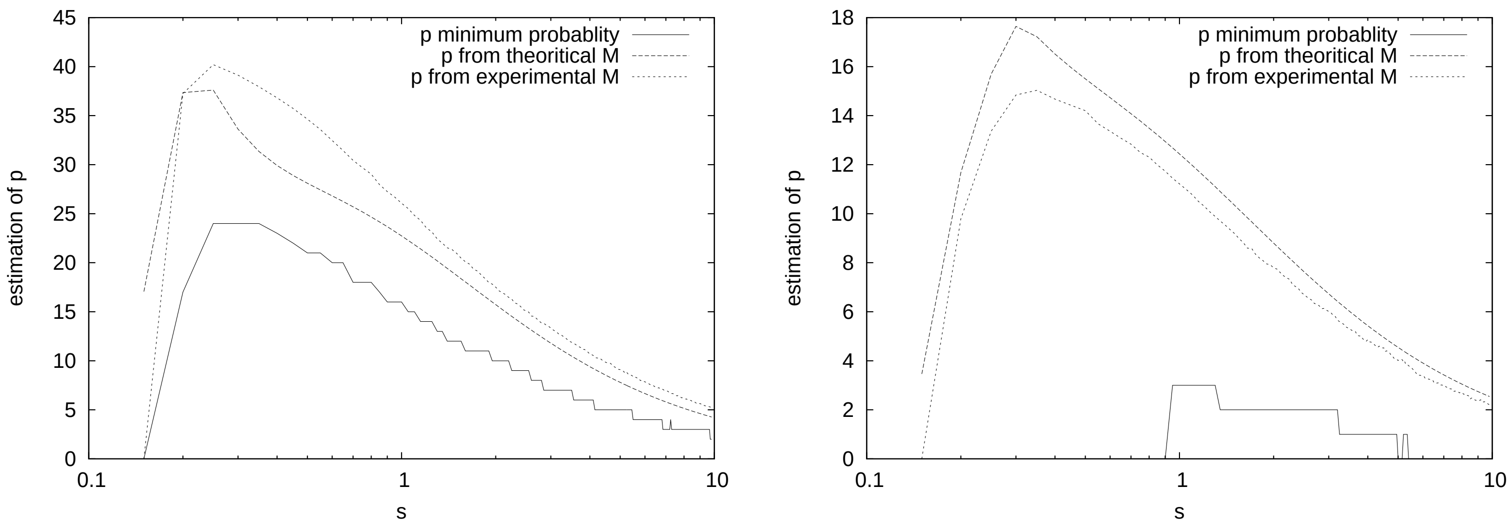
A more moderated empirical estimate of p based on the mean and standard deviation of M

$$p_{\sigma}(s, \delta s) = M(s_{M_{\text{min}}}) - \sigma_M(s_{M_{\text{min}}}) \quad (16)$$

where

$$s_{M_{\text{min}}} = \text{arg}\min_{s \in \mathcal{S}} M(s) \quad (17)$$

Comparing p-estimates



Estimates of p in accordance with Eqns. 14 - 15 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 3 (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. 6, $\mathcal{M}(s)$ that satisfied properties (i)-(iv). As n and d diverge to infinity, M will converge to $\tilde{\mathcal{M}}$ 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. 6 and $\mathcal{M}(s)$ 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 $\mathcal{M}(s)$;

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