

Parameter properties of biochemical systems

Ralph L. Stoop[†], Florian Gomez[‡], Stefan Martignoli* and Ruedi Stoop[‡]

[†]Institute of Physics, University of Basel, 4056 Basel, Switzerland

[‡]Institute of Neuroinformatics, ETH Zurich and University of Zurich, 8057 Zurich, Switzerland

* Department of Mathematics, University of Applied Sciences HSR, 8640 Rapperswil, Switzerland

Email: ruedi@ini.phys.ethz.ch

Abstract—Systems depending on two or more parameters are investigated for the existence of some emergent properties. As an example, we may ask whether they show a particular periodicity in their dynamics. Traditionally, it is assumed that this happens on a direct product of the intervals where this property is displayed when looking at one parameter alone. This, however, is wrong for real-world systems, the origin of the phenomenon being based in the latter's nonlinearities. We demonstrate that the emergence of joint properties is generically confined on shrimp-shaped domains in the combined parameter space. This entrains difficulties for clustering approaches that are generally based on the Gaussian cartesian product point of view.

1. Introduction

Biological organisms are able to fabricate intricate machineries from the molecular scale up to the macroscopic scale, without the obvious need to store and to explicitly handle the corresponding information. Synthetic biology, molecular programming, and nucleic acid nanotechnology have thus become an experimental playground for the search for systems that carry out human-defined molecular programs, to input, output, and manipulate molecular structures. For chemistry to become the next information technology substrate, improved tools for designing, simulating, and analyzing complex molecular circuits and systems are necessary. On the DNA nanotechnology model system, corresponding knowledge is presently quickly growing and the area of alternative computing paradigms starting to take shape. From a physics point of view, biological and physical processes start to converge, so that to describe biochemical computation, concepts from physics can be borrowed and applied.

Most real-world systems exhibit a nontrivial behavior of some observables in time. Many such processes exhibit periodicity (the circadian rhythm, the cell cycle, reproduction), which therefore has often been regarded as a key expression of the essential mechanisms of life. Conversely, irregular behavior is often related to abnormal stimuli, or to a defect or disorder of the generating mechanism. Modern methods of measurements and modeling have now provided techniques that permit the observation of dynamical aspects of processes, which in the past, due to a lack of such technology, were described as steady-state. Genetic

expression processes are an example thereof [1]. Recently, it has been possible to measure down to single cell expression, which revealed different kinds of rhythmic to irregular expression patterns [2, 3]. In our study, we will put forward a generic model that demonstrates that regular and 'stochastic' expression may result from the same nonlinear system and that the transition among these states may require small parameter changes only.

2. Emergence of Shrimps in Parameter Space

How is this multitude of scaled versions of the same shrimp template generated? In the case of smooth systems, shrimps are the result of the interaction of two or more largely independent parameters in creating points with a full set of zero partial derivatives. From this observation, the shrimps phenomenon can be explained in a simple way, for flows (the Rössler system [4]) and for maps (the dissipative Hénon map [5] in [6]). For simplicity of argument we will consider the discrete formulation and follow the exposition given in [7]. Note that the dissipative Hénon map is the paradigmatic two-dimensional discrete map accounting for the universality properties of dissipative nonlinear systems. The Hénon map can be written in its standard form as $f_h : \{x, y\} \rightarrow \{c - dy - x^2, x\}$. After cycling through the coordinates by means of two iterations, the two-dimensional system can be condensed into the approximative one-dimensional map

$$f : x \rightarrow b - (a - x^2)^2,$$

which incorporates the two parameters a, b for the offset and the leading term nonlinearity in one equation.

Stable k -periodic islands arise whenever

$$x_k = f^k(x_k), \quad |m_k| = |f^{k'}(x_k)| < 1 \quad (1)$$

holds, where f^k denotes the k -fold iterated map f and the prime ' denotes the derivative with respect to x . A superstable locus requires that $m_k = 0$. More explicitly, we have

$$f^{k'}(x_k) = \prod_{i=1}^k 4x_i \prod_{i=1}^k (a - x_i^2). \quad (2)$$

This implies that all k -superstable solutions need to pass either through $x_k = 0$ or $x_k = \pm \sqrt{a}$. For the case $x_k = 0$,

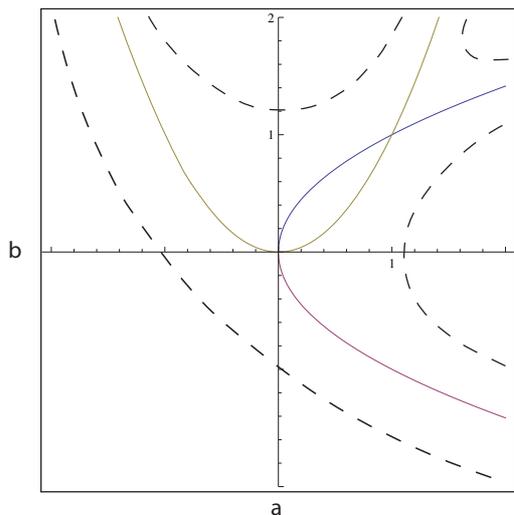


Figure 1: Basic shrimps structure: Two intersecting parabolas of superstability (full lines), extending until the derivative of the solutions exceeds 1 in absolute value (non-generically located dashed lines), where tangent or period-doubling bifurcations occur. In addition, where lines cross, we generally deal with non-ergodicity. Secondary, non-generic, system properties can complicate this fundamental structure.

for $k = 1$ we obtain from $b - (a - x^2)^2 = x$ the relation $a = \pm \sqrt{b}$. For the case $x_k = \pm \sqrt{a}$, we obtain $b = \pm \sqrt{a}$. By differentiability of f in the parameters a, b , this defines two parabolas in parameter space, which define the four legs of the main $k = 1$ -shrimp, see Fig. 1. From this equation, the above-identified two parabolas

$$a = \pm \sqrt{b}, b = \pm \sqrt{a} \quad (3)$$

emerge. By representing Feigenbaum universality in higher-dimensional parameter space, the emergence of shrimp-like structures is thus a universal nonlinear phenomenon, i.e. it must be expected to occur in any nonlinear dynamical system. The placement of the copies is, however, determined by the specific system's properties.

3. Biological Manifestations of Shrimps

To what extent such structures emerge in biological systems has mostly remained unexplored. This is a nontrivial question since there, vast areas in parameter space may not be occupied by typical real-world biological systems and processes. Here we focus on two domains where the dependence of the dynamics on system parameters is of special interest.

The first domain of examples that we consider are cDNA microarrays measurements, from three specific tissues (liver, kidney, testis) from six exchangeable mice. Therefore, the gene-specific attribute is the collection of

tissue-specific expected expression values (which are relative to a common reference mRNA pool derived from equal parts of all mRNA samples ($\mu_{liver}, \mu_{kidney}, \mu_{testis}$)) [8]. Clearly, a plot of the three expression strengths reveals the presence of a shrimps-shaped data structure Fig.2.

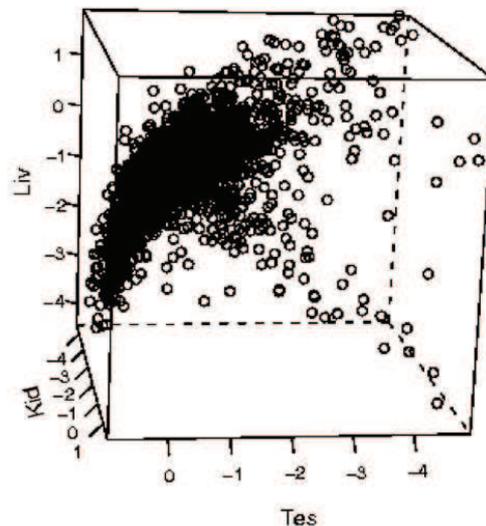


Figure 2: 3-D scatterplots of mouse genome attributes. The point cloud is concentrated around the origin, which corresponds to genes with roughly equal expression in all three tissues. Shrimps arms radiate out. From [8].

The second domain of examples we focus on are Goldbeter biochemical reactions [9], for which corresponding experimental evidence is available [10]. Enzymatic reactions described by the reaction are known to exhibit, at certain parameter values, periodic oscillations. Although rhythmic behavior is not exclusive to enzymic reactions, but rather at all levels of biological organization, enzymic periodicities are the best described at the molecular level. On this level, the system is represented by a two-step enzymic reaction process: Substrate S is injected at constant rate v and runs through enzymic reactions comprising two positive feedback loops coupled in series. S is transformed by catalyzation by an enzyme E_1 , which is activated by its product P_1 . A second enzyme E_2 uses P_1 as substrate and is activated by its product P_2 . k_s is the first-order rate constant for the removal of P_2 . The metabolite concentrations can be described by three ordinary differential equations, cf. [9], containing, among others, parameters $k_s, v, \sigma_1, \sigma_2$.

Whereas Decroly and Goldbeter performed their analysis upon changing parameters k_s and v , we have investigated the behavior by changing σ_1 and σ_2 as shown in Fig. 3. Clearly, we find shrimp-like structures with stable periodic oscillations, starting from period 4 (dark gray) to 8 (green) to 16 (yellow) to 32 (ocher). Regions of chaotic behavior are in white. Thus, motivated by our results from

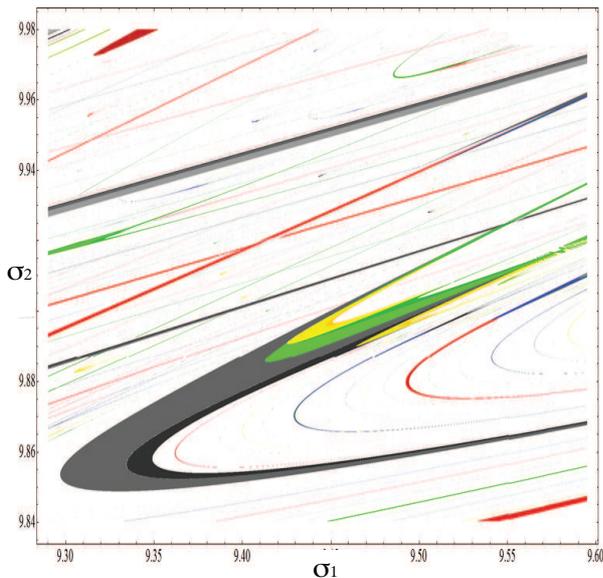


Figure 3: Shrimps in the biochemical system described in [9] (parameter space $\sigma_1 = 9.29 - 9.60$ and $\sigma_2 = 9.84 - 9.98$). Ref. [9] used $\sigma_1 = \sigma_2 = 10$.

electronic systems, we also find large shrimp-shaped domains in biological systems and microbiological reactions. The self-similarity of the shrimp areas may simplify the tuning of jumps from one to another periodic behavior, enabling in this way simple state-coding in terms of periodic signaling.

4. Feature Maps

To relate these observations to bioinformatics, it is important to note that shrimps are of course not restricted to two-dimensional parameter space, they also exist in higher dimensions. Moreover, in many applications, objects may not be directly characterized by their fundamental parameters (these are often unknown), but by easily observable features. From parameter coordinates, we arrive at the feature space by means of a feature-mapping, that most of the time is implicit. In a not too high-dimensional feature space, feature-maps of sufficient smoothness will closely reflect the situation that we have in parameter space (though a formalization of this expectation may require an advanced mathematical framework). A pictorial example is provided by the transformation $(a, b) \rightarrow (a, \text{Log}(1 + |b|), ab)$ from two- into three-dimensional space (Fig. 4).

5. Clustering and Parameter Estimate Consequences

Suppose that we now sample the parameter or feature space with the aim of identifying parameters that lead to a periodic system response. Whatever the sampling procedure and the test for periodicity, what will likely result is a situation where candidate systems will be from

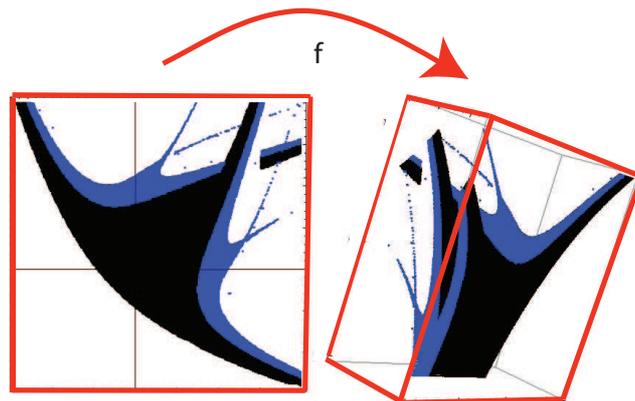


Figure 4: Artificial feature-map example: A two-dimensional shrimp is mapped into a three-dimensional feature space shrimp by means of the transformation $f : (a, b) \rightarrow (a, \text{Log}(1 + |b|), ab)$. Shrimp essentials are preserved under map f ; transformations of similar mathematical properties yield comparable results. (Left-hand side: Black area: Parameters with the same periodicity; blue area: parameters with period-doubled periodicity. Right-hand side: Corresponding features.)

primary shrimps or from lesser populated areas hosting smaller shrimps, or systems for which the data appears periodic, but is actually chaotic (unstable periodic orbits are generically embedded into chaos, and the systems' trajectories can follow such orbits for quite some time). Taking this situation as a toy example, we now proceed towards the clustering of the data into sets of similar behavior. To this end, we suppose that similar parameters generate more similar behaviors than dissimilar ones. The principle that clustering is thus based on, is the smaller the distance in space (parameter-, feature-) is, the more they are coupled and likely to be in the same cluster. The interesting observation then is that even in this case, the most prominent clustering algorithms fail in the clustering of convex-concave bounded sets such as our shrimp-like domains, since they are implicitly based on a linear separability criterion. While this is evident for the popular k-means algorithm, this also holds for hierarchical, agglomerative Wards clustering.

In a Bayesian context, standard parameter inference methods such as the standard Metropolis algorithm could be expected to fail as well in this context. To check these expectations, we performed a survey of applications of ABC methods on our parameter space, where [11] served as the references of models and methods. Our numerical experiments demonstrate that even in the context of the strongly fractionalized parameter spaces of nonlinear systems, the ABC approaches perform well. This is mainly due to the fact that they are ensemble-based. Given some observed data x and a proposed model with parameters θ , the aim of parameter inference is to find those parameter

values that give the “best fit” to the observation. To find the parameter values providing the best fit, a common choice is algorithms from the family of genetic algorithms or evolutionary strategies, that minimize the risk of getting stuck in local maxima/minima. We check here the performance of typical parameter inference methods when applied to the simplest shrimp-exhibiting system. The ABC sequential Monte Carlo ‘ABC-SMC’ method from [11] is used to infer the parameters a, b of the one-dimensional form of the Hénon map. We focus on the case $a = 1.1$ and $b = 0.3$, which is located on a shrimp of stable period-2 behavior, but closely towards its border. Following [11], starting from the initial position $x_0 = -0.2$, we generated a time series of the orbit, to which we added Gaussian noise of mean μ and variance σ^2 , to obtain the reference data x_r . As the error function, we chose the sum of squared errors $d(x, x_r) = \sum_i (x[i] - x_r[i])^2$. The prior was a uniform distribution on $a \in (0.3, 1.3)$ and $b \in (0.1, 1.0)$, sampled by invariable $N = 1000$ systems or ‘particles’. ABC-SMC finds the correct region in parameter space within relatively little computational time. The particular shape of the shrimps poses no problem, see Fig. 5c).

References

- [1] M.C. Romano, M. Thiel, I. Stansfield and C. Grebogi, “Queueing Phase Transition: Theory of Translation”, *Phys. Rev. Lett.*, vol. 102, pp. 198104, 2009.
- [2] D.G. Spiller., C.D. Wood, D.A. Rand and M.R.H. White, “Measurement of Single-Cell Dynamics”, *Nature*, vol. 465, pp. 736-745, 2010.
- [3] D.M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler and F. Naef, “Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics”, *Science*, vol. 332, pp. 472-474, 2011.
- [4] P. Gaspard, R. Kapral and G. Nicolis, “Bifurcation Phenomena near Homoclinic Systems: A Two-Parameter Analysis”, *J. Stat. Phys.*, vol. 35, pp. 697-727, 1984.
- [5] M. Hénon, “A Two-dimensional Mapping with a Strange Attractor”, *Comm. Math. Phys.*, vol. 50, pp. 69-77, 1976.
- [6] J.A.C. Gallas, “Structure of the Parameter Space of a Ring Cavity” *Appl. Phys. B*, vol. 60, pp. 203-213, 1995.
- [7] R. Stoop, P. Benner and Y. Uwate, “Real-world Existence and Origins of the Spiral Organization of Shrimp-shaped Domains”, *Phys. Rev. Lett.*, vol. 105, e074102, 2010.

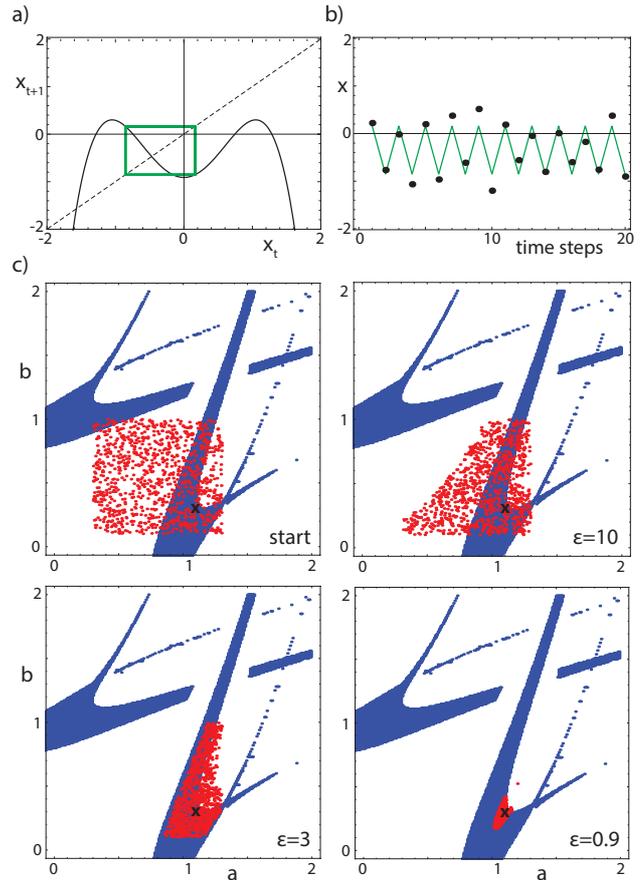


Figure 5: ABC-SMC for the parameters (a, b) of the one-dimensional version of Hénon’s map f (see text). a) Graph of $f(x)$ (black) and stable period-2 solution (green), for $a = 1.1$, $b = 0.3$ (100 initial iterations discarded). b) Noise-corrupted data (20 points of a) plus Gaussian noise ($\mu = 0$, $\sigma = 0.2$). c) ABC-SMC particles (red) and period-2 shrimps (blue) in parameter space, for different tolerances ϵ . Black cross: target values.

- [8] J. Bryan, “Problems in Gene Clustering based on Gene Expression Data”, *J. Multivariate Anal.*, vol. 90, pp. 44-66, 2004.
- [9] O. Decroly and A. Goldbeter, “Birhythmicity, chaos, and other patterns of temporal selforganization in a multiply regulated biochemical system”, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 79, pp. 6917-6921, 1982.
- [10] I.M. De la Fuente, “Diversity of Temporal Self-organized Behaviors in a Biochemical System”, *BioSystems*, vol. 50, pp. 83-97, 1999.
- [11] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M.P.H. Stumpf, “Approximate Bayesian Computation Scheme for Parameter Inference and Model Selection in Dynamical Systems”, *J. R. Soc. Interface*, vol. 6, pp. 187-202, 2009.