

Fractional Fokker–Planck Equations and Artificial Neural Networks for Stochastic Control of Tokamak

Danilo Rastovic

Published online: 30 October 2007
© Springer Science+Business Media, LLC 2007

Abstract The general form of description of Kolmogorov–Arnold–Moser (KAM) theorem in controlled plasma fusion, is obtained via the theory of artificial fuzzy neural networks. Without of the global maximum entropy principle, the complexity function is used for the Monte Carlo simulations.

Keywords KAM theory · Tokamak · Stochastic control

Introduction

Recently, stochastic fluctuations have been discussed in relation to ELM (edge localized mode) mitigation in tokamaks. Auxilliary coils are being added to existing configurations to control transport in several tokamaks. These additional coils are new and dominating sources of stochasticity. Examples can be found on the tokamaks Tore Supra, DIII-D, and TEXTOR, and are being planned for JET, ASDEX-UPGRADE, and ITER.

Edge localized modes are periodic disturbances of the plasma periphery occuring in tokamaks with an H-mode edge transport barrier. As a result, a fraction of the plasma is transferred to the open field lines in the divertor region, ultimately appearing at the divertor target plates.

Turbulence plays a very important role in particles and energy cross-field transport to the wall in the edge plasma. Using the TCABR tokamak facility are analyzed turbulent electrostatic fluctuations in a stationary toroidal magnetoplasma, created by radio-frequency waves and confined by two different toroidal magnetic fields. Turbulence has

recurrent properties, as those observed in recurrent fully chaotic low-dimensional systems [1]. Therefore, evolution of measurements of low-dimensional dynamical systems can be used to describe the recurrence observed in the tokamak edge turbulence.

Empirically we know that we need both fuelling and heating to maintain steady state. This means that a pure heat source cannot maintain the density and a pure particle source cannot maintain the temperature. The non-Markovian mixing length has turned out to automatically give us the right level also of momentum transport.

The understanding and reduction of turbulent transport in magnetic confinement devices is not only an academic task but also a matter of practical interest, since high confinement has been chosen as the regime for ITER and possible future reactors because it reduces size and cost. Over the past decade, step-by-step new regimes of plasma operation have been identified, whereby turbulence can be externally controlled, which led to better and better confinement. Theoretical models were often only predicting the global level of turbulence as well as the scaling of this level with varying plasma parameters.

The various neutron diagnostic techniques used to determine the characteristics of the neutron fields of the plasma focus device have been developed with the following aim: identification of the neutron emission and plasma focus device operational parameters. Our algorithm recognizes the prepared scenarios and it classifies them into groups. The new feedback control of the neutron emission rate and the radiative power in the divertor has been performed. The feedback control of neutron emission rate was demonstrated with controlling the heating power.

In the aim to obtain the possibility of uniform stabilization of a control system we used the method of adaptive controllers for nonlinear transport systems. It opens the

D. Rastovic (✉)
Control Systems Group, Nehajska 62, Zagreb 10110, Croatia
e-mail: dr.rastovic@excite.com

possibilities for experimentation with the influence of real actuators on different devices in nuclear technology and radiation science.

The Kolmogorov–Arnold–Moser (KAM) theory of dynamical systems asserts that in certain circumstances most of the magnetic lines will sweep out nested flux surfaces rather than ergodic regions so that the confinement of the plasma will be adequate for fusion in power plant [2]. A generalization of Brownian motion to strongly correlated (long-memory) random processes is fractional Brownian motion (FBM). Many current models for long-memory are non-Markovian Monte Carlo methods can often be run with a limited amount of computer time and provide a solution with some estimable uncertainty.

Markov Chains

For irreducible stochastic matrix the controlled Markov chain is said to be ergodic and there exists a stationary strategy. By Ruelle–Perron–Frobenius theorem there exists a countable measure such that under certain conditions convergence toward this measure is uniform and exponential. It is called thermodynamic formalism for countable Markov shifts. The positive recurrence is necessary and sufficient condition for a Ruelle–Perron–Frobenius theorem to hold. In the case of an unbounded integrable target function, in the equation $y(k+1) = P[y(k), u(k)]$ the input function $u(k)$ should be an unbounded function. The convergence of step functions to the function $v(k)$ of the infinite fuzzy logic controller in this case cannot be uniform, i.e. the generalized fixed point theorem does not hold true.

Turbulence as a whole is generally not an equilibrium phenomenon. As a simulation approach to study thermodynamics the Monte Carlo histogram technique is usually used. In the case of infinite degrees of freedom, obtaining the histogram corresponds to considering an appropriate probabilistic distribution. In the case of nonlinearity, we must be able, by addition of another actuators with the methods of artificial intelligence and synergetics to bring the plasma regime to the linear case.

The control and observation processes for many dynamical systems are often severely limited. For many systems described by partial differential equations it is usually impossible to influence or sense the state of the system at each point of the spatial domain. Indeed control and sensors are restricted to a few points or parts of the boundary. Modelling such limitations result is unbounded input and output operators. The model of collisionless plasmas, specially in controlled fusion is too idealized and collisional effects need to be incorporated. We shall see what is the role of recurrence for obtaining of good

machine behaviour. The solution of nonrecurrent processes cannot be found in the united form. The interesting problem is an analysis of limit distribution of integral functionals in a null-recurrent diffusion process. The method for obtaining of adaptive recurrence equilibriums is explained. If we have only strong stabilization of the process then in the case when there are constants appropriate from strong convergence we can get uniform convergence by the principle of uniform boundedness. Generally, we have transport equations with unbounded second-order differential collision operators. For the stabilization of such systems it is necessary introduce the unbounded input control functions.

Complexity Function

The notions of the complexity function and entropy function are introduced to describe systems with nonzero or zero Lyapunov exponents or systems that exhibit strong intermittent behaviour with flights, trappings, weak mixing etc. The important part of the new notions is the appearance of epsilon-separation of initially close trajectories. It is found that Hamiltonian chaotic dynamics possesses in many cases a kinetics that does not obey the Gaussian law process and that fluctuations of the observables can be persistent, i.e., there is not any characteristic time of the fluctuation decay. The new approach to the problem of complexity and entropy covers different limit cases, exponential and polynomial, depending on the local instability of trajectories and the way of the trajectories' dispersion.

The behaviour of systems with zero Lyapunov exponents definitely have some level of complexity and some value of entropy in physical sense, but the regular notion of the Kolmogorov–Sinai entropy of the standard definitions of complexity cannot be applied to such systems. Space-time nonuniformity suggest that vicinities of any trajectories may have very different dynamics of trajectories. It is difficult, if not impossible, to describe a trajectory finite-time behaviour on the basis of the information about the trajectory from an infinitesimal domain of phase space. A notion of complexity that is based on the verification of divergence of trajectories from fixed several ones is introduced. If consideration is restricted to a neighbourhood of one (or several) basic orbit, then fast separated pieces of orbits correspond to a mixing type of behaviour can be eliminated. It is important to consider complexity function and exit time distribution. If time t is fairly big we can select a set of points x that are almost uniformly distributed. We can consider a collection of flights and their lengths and time intervals from different domain. As a result we have semilocal flights complexity function [3].

We wish to investigate the problems of computing the intervals of possible values of the latest starting times and floats of activities in networks with uncertain durations modeled by fuzzy or interval numbers. Problem of evaluating the possible criticality of an activity—a polynomially solvable case is considered. A class of neural architectures of polynomial neural networks (PNNs) is introduced [4]. PNN is flexible neural architecture whose structure (topology) is developed through learning. In particular, the number of layers of the PNN is not fixed in advance but is generated on the fly. In this sense, PNN is a self-organizing network. Especially, the generic rules in the system assume the form “if A then $y = P(x)$ ”, where A is a fuzzy relation in the condition space while $P(x)$ is a polynomial forming a conclusion part of the rule. When the complexity of the system to be modeled increases, both experimental data and some prior domain knowledge are of importance to complete and efficient design procedure. Each neuron (node) of the network realizes a polynomial type of particular description of the mapping between input and output variables. Bernstein polynomials associated with fuzzy valued functions are employed to approximate continuous fuzzy valued function defined on a compact set. Universal approximations of continuous fuzzy valued functions by regular fuzzy neural networks are obtained [5].

Singularly perturbed control systems have been intensively studied. The multitime-scale approach is a fundamental characteristic of singular perturbation methods. In other words, the method can decompose the original systems into the fast and slow subsystems. Therefore, the controller of actual system is a compact form of the controllers for its relative fast and slow subsystems [6].

Intelligent identification is a significant approach for modeling complex, uncertain and highly nonlinear dynamic systems. The nonlinear autoregressive with exogenous input model can be used. This structure is used in most nonlinear identification methods such as neural networks and fuzzy models. A simple but effective fuzzy-rule-based models of complex systems from input-output data was developed. In recent years, the corresponding research result have been extended to multiple-input-multiple-output nonlinear systems. The basic idea of these works is to use the fuzzy logic systems to approximate the unknown nonlinear functions in systems and design adaptive fuzzy controllers by using Lyapunov stability theory. From a mathematical point of view, fuzzy logic systems can be used as practical function approximators.

Monte Carlo Simulations

Monte Carlo methods are online simulation methods that learn from experience based on randomly generated

simulations, without the need for complete knowledge of the environment. Given a random set of experiences (or trials), with guarantee from the weak law of large numbers, the simulation result will eventually converge when each state is encountered for an infinite number of trials. The evaluative feedback is collected for each instance of example presented to the induction algorithm based on the state of the feature sets. The average feedback or weight of each feature represents the relative importance of the feature over other features. Higher weight value means that the particular feature is of higher importance.

A novel feature selection approach, the Monte Carlo evaluative selection (MCES) is proposed in the paper [7]. MCES is an objective sampling method that derives a better estimation of the relevancy measure. The algorithm is objectively designed to be applicable to both classification and nonlinear regressive tasks. Most of the filter approaches assume some type of relevancy measures to determine the relevancy of the input features are to be excluded in a feature selection system. Irrelevant features: in computational learning, accuracy is one of the efficient measures to define the relevancy. It is a relationship between the input features that are highly correlated to the some other features in the same system, then they can be considered as redundant features and should be removed. A computational learning or soft computing type of induction or inference system is generally stochastic or nondeterministic. The effectiveness of the feature selection algorithm is highly proportional to the number of trials attempted during the different learning states of the system. The goal of the feature selection algorithm is to extract the significance of each feature based on the knowledge inherently present in a system after the training process.

In recent years, support vector machines (SVMs) with linear or nonlinear kernels [8] have become one of the most promising learning algorithms for classification as well as for regression, which are two fundamental tasks in data mining. Via the use of kernel mapping, variants of SVMs have successfully incorporated effective and flexible nonlinear models. The reduced SVM mixture model via uniform random sampling minimizes the maximal model bias (deviation) between the reduced model and the full model. The uniform random subset should be done within each class and then be combined together, the stratified sampling is a “must” for multiple classification in order to reduce the variance due to the Monte Carlo sampling, especially for problems with large numbers of classes. This random subset approach can drastically cut down the model complexity, while the sampling design helps to guide the bases selection in terms of minimal model variation. The uniform design is a space-filling design and it seeks to obtain maximal model robustness.

Limitations of Markov chain Monte Carlo algorithms to sample from the posterior probability of a tree given the data are founded [9]. In particular, they design a Markov chain whose stationary distribution is the desired posterior distribution, computed using the likelihood and the priors. Hence, the running time of the algorithm depends on the convergence rate of the Markov chain to its stationary distribution. However, there is no theoretical understanding of the circumstances which the Markov chains will converge quickly or slowly.

Recently, dynamical behaviours defined on various complex networks have been extensively studied in many fields for phase transitions with long-range directed interactions. The lattice structure is replaced again after a Monte Carlo step. The system exhibits a transition from the homogenous state to a global oscillation state again and then the last self-organizing state with fraction q increasing [10]. In our simulation, we change the directed small-world structure per Monte Carlo step.

Harris recurrence is a concept introduced 50 years ago by Harris. More recently, connections between Harris recurrence and Markov chain Monte Carlo algorithms were investigated. The hierarchical structure is a common feature of many networked systems and has received a considerable amount of attention in recent years. It is shown that the hierarchical structure is related to some significant characteristics of complex systems, such as the high clustering coefficient and scale-free degree distribution.

The algorithm based on spectral partitioning for reconstructing the hierarchical structure from a network is proposed [11]. There is a large literature within computer science on spectral partitioning, in which network properties are linked to the spectrum of the graph Laplacian matrix. Many applications of Markov chain Monte Carlo (MCMC) involve very large and/or complex state spaces, and convergence rates are an important issue. A major problem in MCMC is thus to find sampling schemes whose mixing times do not grow too rapidly as the size or complexity of the space is increased.

In a multi-modal space a local chain will equilibrate rapidly within a mode, but takes a long time to move from one mode to another. Hence, the entire chain converges slowly to a target distribution. However, a small fraction of heavy-tailed proposals enables a small-world chain to move from mode to mode much more quickly.

A positive, self-similar Markov process (PSSMP) is a strong Markov process with paths which possesses a scaling properties. The main result asserts that any PSSMP may, up to its first hitting time at 0, be expressed as the exponential of a Levy process, time changed by the inverse of its exponential functional. We obtain one-to-one relation between the class of PSSMPs killed at time s and one of the Levy processes [12].

Artificial Neural Networks

Accurate neural network approximation for closed-loop system dynamics is achieved in a local region along a periodic state trajectory, and a learning ability is implemented during a closed-loop feedback control process. Second, based on the deterministic learning mechanism, a neural learning control scheme is proposed which can effectively recall and reuse the learned knowledge to achieve closed-loop stability and improved control performance. Most of the works in the neural control literature only require the universal approximation capability of neural networks, which is also possessed by many other function approximators such as polynomial, rational and spline functions, wavelets, and fuzzy logic systems. Adaptive control has a main feature the ability to adapt to or “learn” the unknown parameters through online adjustment of controller parameters in order to achieve a desired level for control performance. The nature of the deterministic learning mechanism has been shown to be related to the exponential stability of the closed-loop adaptive system.

When a system is trained by examples, it is important to improve its generalization ability and at the same time to reduce its complexity. By pruning some insignificant connections, this approach is also suitable for reducing the network complexity by constructing a compact one. A large initial size allows the network to learn quickly with less sensitivity to initial conditions and a lower probability to be trapped in local minima. The trimmed network favors improving generalization and reducing complexity. The neural networks can be classified as static (feedforward) and dynamic (recurrent) [13]. The output of a dynamic system is a function of past outputs and past inputs.

Recurrent neural network (RNN) is a powerful tool for sequence learning and prediction. Characterized with the recurrent connection, RNN is able to memorize the past information, therefore, it can be learn and predict dynamic properties of the sequential behaviour. The paper [14] presents incremental hierarchical discriminant regression (IHDR) which incrementally builds a decision tree or regression tree for very high-dimensional regression or decision spaces by an online, real-time learning system. The IHDR tree dynamically assigns long-term memory to avoid the loss-of-memory problem typical with a global-fitting learning algorithm for neural networks. A major challenge for an incrementally built tree is that the number of samples varies arbitrarily during the construction process. Prediction trees, also called decision trees, have been widely used in machine learning to generate a set of tree-based prediction rules for better prediction for unknown future data. The time complexity is typically not an issue for decision trees.

With the demand of online, real-time, incremental, multimodality learning with high-dimensional sensing by an autonomously learning embodied agent, we require a general purpose regression technique that satisfies all of the following challenging requirements. It must adapt to increasing complexity dynamically. It cannot have a fixed number of parameters like a traditional neural network, since the complexity of the desired regression function is unpredictable. It must be able to retain most of the information of the long-term memory without catastrophic memory loss.

For example, consider neural networks with incremental backpropagation learning. They perform incremental learning and can adapt to the latest sample with a few iterations, but they not have a systematically organized long-term memory, and, thus, early samples will be forgotten in latter training. Cascade-correlation learning architecture improves them by adding hidden units incrementally and fixing their weights to become permanent feature detectors in the network. Thus, it adds long-term memory. Major problems for them include the high-dimensional inputs and local minima.

Neural networks can be classified into static and dynamic categories. Static networks have no feedback elements and contain no delays; the output depends not only on the current input to the network, but also on the current or previous inputs, outputs, or states of the network. These dynamic networks may be recurrent networks with feedback connections or feedforward networks with imbedded tapped delay lines.

Last years, extremaly good initial results were obtained with the suppression of edge localized modes (ELMs) with stochastic layers. It remains important to understand the impact of plasma shape and plasma profiles on the stability limit. The divertor enables access to a new neutral beam injection (NBI), heated, high-density operating regime with improved confinement properties. Quasi-Monte Carlo methods are developed by using smoothing and dimension reduction of the integration domain. Specifically, we are interested in conditions under which an unbounded memory can induce qualitative changes in the distribution of the position, as compared to the Markovian case with Gaussian distribution pertaining on large space and time scale. Hence, one has the standard Markovian random walk, which, on large scales, converges to Brownian motion.

The Kolmogorov–Arnold–Moser (KAM) theorem: phase spaces of Hamiltonian systems split up in various areas when coupling between the degrees of freedom are coupled in such a way that the equations become unintegrable. These areas are either island shaped or unbroken layers (KAM tori) or chaotic [15]. Levy determined the conditions for a family of distributions to be stable. These distributions are usually called Levy stable distributions.

Unfortunately, the general form of Levy distributions is not available. Levy flight and walks are stochastic processes which provide a framework for the description and analysis of anomalous random walks in physics.

Diffusion in most plasma devices, particularly tokamaks, is higher than one would predict from understood causes. In collaboration with Ben Carreras, George Zaslavsky has undertaken to apply sophisticated methods for analyzing edge fluctuation data from DIII-D tokamak in order to discover non-Gaussian processes which may have major effects on transport. The analysis hand the start of advanced data analysis of turbulence in tokamaks [16].

The non-Markovian character of the dynamics is expressed in the fact that the evolution equation is different for each initial position. Now, the problem is to describe non-Markovian processes. The next theorem holds:

Theorem Let us have plasma behaviour which is described by KAM theory. Then, it can be obtained as follows: (a) by a generalized fixed point method with unbounded input and output functions in the deterministic case (H mode), and (b) by artificial neural networks with delays in the stochastic case (L–H transitions and ELMs).

Proof (a) The proof is given in Ref. [17]. It all follows according to the contraction mapping principle with infinite fuzzy logic controllers.

(b) Instead of the maximum entropy principle we must consider learning and simulations of complexity by artificial neural networks to obtain recurrent behaviour of the system.

Conclusion

Self-similarity and long-range correlations are present in the plasma edge of fusion devices. Since the turbulence is self-similar, the probability distribution functions (PDF) may also be expected to display self-similarity, i.e. when averaging the signal over time, the PDF changes amplitude but not shape.

A new method is presented to derive kinetic equations for systems undergoing nonlinear transport in the presence of memory effects. In the framework of mesoscopic non-equilibrium thermodynamics is obtained a generalized Fokker–Planck equation incorporating memory effects through time-dependent coefficients. The nonMarkovian dynamics of anomalous diffusion is discussed [18].

There are many different areas where fractional equations describe real processes. The physical reasons for the appearance of fractional equations are intermittancy, dissipation, wave propagation in complex media, long memory and others. The equilibrium of the fractal turbulent medium exists for the magnetic field with the power law relation

[19]. Magnetohydrodynamics equations for the fractal distribution of charged particles are suggested. The fractional integrals are considered as approximations of integrals on fractals. The generalization of the Fokker–Planck equation can be used to describe kinetics in fractals media. Some results hold true for fractional Vlasov–Poisson–Fokker–Planck equation. The physical values on fractals can be “averaged”, and the distribution of the values on fractal can be replaced by some continuous distribution [20]. In the absence of any external source, turbulent flows decay because of dissipation. The uniformly sheared flow is one of the simplest ways of producing and maintaining turbulence [21, 22].

References

1. A.A. Ferreira, M.V.A.P. Heller, M.S. Baptista, I.L. Caldas, Braz. J. Phys. **32**(1), 85–88 (2002)
2. P.R. Garabedian, Comm. Pure Appl. Math. **27**, 281–292 (1994)
3. V. Afraimovich, G.M. Zaslavsky, Chaos **13**(2), 519–546 (2003)
4. S.K. Oh, W. Pedrycz, Fuzzy Sets Syst. **142**, 163–198 (2004)
5. P. Liu, Fuzzy Sets Syst. **119**, 313–320 (2001)
6. T.H.S. Li, K.J. Lin, IEEE Trans. Fuzzy Syst. **15**(2), 176–187 (2007)
7. K.H. Quah, C. Quek, IEEE Trans. Neural Networks **18**(2), 431–448 (2007)
8. Z. Shi, M. Han, IEEE Trans. Neural Networks **18**(2), 359–372 (2007)
9. E. Mossel, E. Vigoda, Ann. Appl. Prob. **16**(4), 2215–2234 (2006)
10. C.Y. Ying, D.Y. Hua, L.Y. Wang, J. Phys. A: Math. Theor. **40**, 4477–4482 (2007)
11. F. Chen, Z. Chen, Z. Liu, L. Xiang, Z. Yuan, J. Phys. A: Math. Theor. **40**, 5013–5023 (2007)
12. M.E. Caballero, L. Chaumont, J. Appl. Prob. **43**, 967–983 (2006)
13. A. Savran, IEEE Trans. Neural Networks **18**(2), 373–384 (2007)
14. J.A. Starzyk, H. He, IEEE Trans. Neural Networks **18**(2), 344–358 (2007)
15. D. Rastovic, Nucl. Technol. Radiat. Protect. **21**(2), 14–20 (2006)
16. B.Ph. Milligen et al., in *Proceedings of the 26th EPS Conf. on Contr. Fusion Plasma Phys.*, vol. 23 (ECA, 1999), pp. 49–52
17. D. Rastovic, Stab. Contr. Theory Appl. **2**, 53–58 (1999)
18. V.E. Tarasov, G.M. Zaslavsky, Comm. Nonlin. Sci. Num. Simul. **11**, 885–898 (2006)
19. I. Santamaria-Holek, J.M. Rubi, Physica A: Stat. Mech. Appl. **326**, 384–399 (2003)
20. V.E. Tarasov, Phys. Plasmas **13**, 052107 (2006)
21. D. Rastovic, J. Fusion Energ., Online (Dec. 2007)
22. D. Rastovic, Chaos, Solitons Fract. **32**, 676–681 (2007)