Dynamical Systems, Optimization, and Chaos

John B. Moore
Department of Systems Engineering
and Cooperative Research Centre for Robust and Adaptive Systems
Research School of Information Sciences and Engineering
Canberra ACT 0200, Australia

Abstract

Much of engineering is concerned with the topic of optimization, and at the heart of much of our optimization is dynamical systems. Dynamical systems can be thought of as either non-linear continuous-time differential equations or difference equations. Chaos occurs in dynamical systems, and frequently in engineering we seek to avoid chaos. At times chaos becomes the central fascination.

This paper first introduces a situation in signal processing for neural systems in which chaos is the perhaps unexpected phenomena and the object of study. The focus then shifts to the topic of optimization of systems via dynamical systems, where traditionally chaos is avoided as much as possible. The essential dynamical system should converge in a very smooth manner to an optimal solution to some problem of interest. Our technical approach is summarized to optimization via dynamical systems is illustrated by an application in the area of robotics.

The key questions motivating this research are: Does the human brain exploit chaos for generating intelligence? Can our computing machines and control systems enhance their intelligence by a clever introduction of chaos?

1 Introduction

A Nobel prize-winning experiment in neurophysiology extracts very faint signals from synapses. A patch of the cell membrane with a gate molecule is studied. The one molecule acting as a gate opens and closes to allow various chlorine or potassium ions to flow through the cell membrane and activate the cell. The current flow is of the order of Femto amps. At these signal levels it is not surprising that noise due to thermal agitation of the molecule can dominate the measurement process. It is important to study these very small channel currents in order to not only understand signal processing in the brain in normal behaviour, but to study the effect of drugs for anesthesia, epilepsy and other conditions. A key question of interest is: Are the underlying processes at the cell membrane and synapse level governed by chaotic equations?

This paper points to research results which suggest that the underlying processes and synapse levels are in fact chaotic, see [9], and background material [3, 4, 6, 8, 10]. It is very difficult to be absolutely sure of such a conclusion because the signals are so much buried in noise. However, through experiments and signal processing, one can assess the self-similarity of the underlying signals at different resolution scales, and indeed estimate the fractal dimension of the underlying signals. Recall, that the fractal dimension is really a measure of the ruggedness of the underlying signals.

In the case of cell channel currents, the underlying signals appear to be currents which switch between a number of levels according to some transition probability law. The transition probabilities depend in a exponential manner on the time to the last transition. The longer the time since the last transition, the less likely there will be another transition. In cell channel currents, transitions occur in pico-seconds. It is clear that the inertia of the protein molecules forming the cell channel would be very small indeed.

It appears that in the process of evolution, there has been some advantage in exploiting chaos for the underlying processes within the human brain. Usually, chaos is avoided in performing a system design or optimization. The challenge before engineers is to somehow exploit the fascinating properties of chaos to enhance their system designs, and to further their ability to optimize and control their systems.

At this stage in our understanding of optimization, we do in fact exploit dynamical system behaviour, in particular discrete-time (recursive) systems for system optimization. The dynamical vector or matrix equations may be quite elegant and with the ability to flow on a constraint manifold towards an optimal solution. In the first instance, one is content that such dynamical systems converge to an optimal solution in a well-behaved manner. Subsequently, the motivation is to enhance the convergence capabilities of such algorithms by introducing non-smooth behaviour. There may be a deliberate introduction of ill conditioning into the equation or random perturbations to ensure the final desired outcome.

In Section 2 of the paper, we review the technical approach for investigating chaos in so called hidden Markov models for discrete-state systems and its application to study cell channel currents. In Section 3 we review the technical approach for system optimization via dynamical systems with an illustration from the application area of robotics. In the final Section 4, we discuss areas for current and future research.

2 Discrete-State Systems and Chaos

Let us consider dynamical systems which switch between discrete states, denoted $S_1, S_2, \ldots, S_N$. In discrete-time, the state of the system can be indicated by an indicator-vector X_k, where $k = 0, 1, \ldots$ and usually denotes a discrete-time sequence. The state X_k belongs to a discrete-set $\{e_1, e_2, \ldots e_N\}$ where e_i is the unit vector with unity in the ith element and zero otherwise. See Figure 1.

Two important properties of such indicator states X are as follows. Nonlinear functions of X are linear in X as

$$f(X) = f\left(\sum_{i=1}^{N} e_i X^{(i)}\right) = \sum_{i=1}^{N} f(e_i) X^{(i)} = FX. \tag{1}$$

where $F = [f(e_1), f(e_2) \ldots f(e_N)]$ and where $X^{(i)}$ denotes the ith element of X. This is readily checked since $X^{(i)} = 0$ for all i save some value j when $X^{(j)} = 1$ (i.e. $X = e_j$) and thus $f(X) = f(e_j)$.

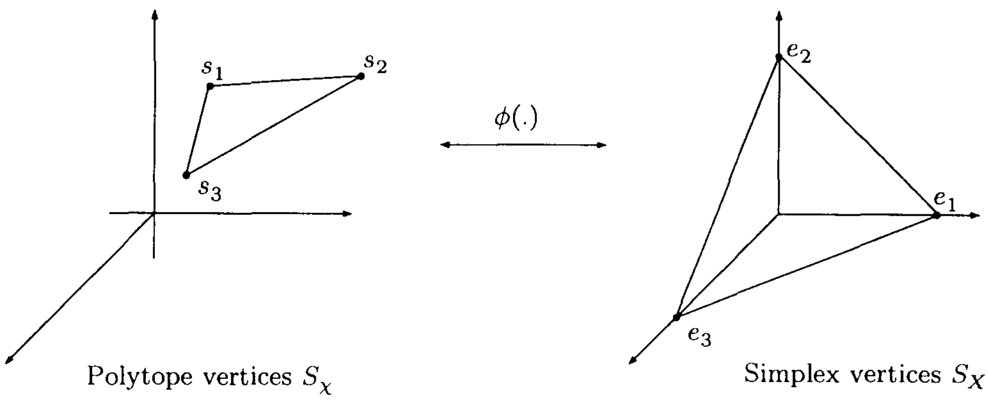


Figure 1: Depiction of state sets.

Also

$$E\left[X^{(i)}\right] = \sum_{j=1}^{N} e_j' e_i P(X = e_j) \quad = \quad P(X = e_i) \tag{2}$$

Here $E[\cdot]$ is the expectation operator and $P(\cdot)$ denotes the probability. This result is immediate since $e_j' e_i = 0$ for $c \neq j$ and $e_i' e_i = 1$.

Consider that the system switches between states according to a probability law

$$X_{k+1} = AX_k + M_{k+1}, \tag{3}$$

where A is a matrix of transition probabilities. That is, $E[X_{k+1}] = AE[X_k]$. It is immediately clear that M_{k+1} is a martingale increment process with the property

$$E[M_{k+1}|X_0, X_1, \ldots X_k] = 0 \tag{4}$$

The transition matrix A of interest to us here will also depend upon the time to the last transition, denoted τ_k. Of course, $\tau_{k+1} = \tau_k + 1$ in the event that there has been no transition, and $\tau_k = 0$ in the event that there has been a state transition. The vector consisting of X_k and τ_k is seen to be first-order Markov, in that it depends only on the previous vector, X_{k-1}, τ_{k-1}, and not on any earlier such states. Thus the augmented state model is

$$\begin{aligned} \begin{bmatrix} X_{k+1} \\ \tau_{k+1} \end{bmatrix} &= \begin{bmatrix} A(\tau_k) & 0 \\ 0 & \rho \end{bmatrix} + \begin{bmatrix} M_{k+1} \\ \rho \end{bmatrix} \\ \rho &= \begin{cases} 0 & \text{if } X_{k+1} \neq X_k \\ 1 & \text{otherwise} \end{cases} \end{aligned} \tag{5}$$

The measurement process y_k belongs to continuous range, namely $\mathbb{R}^1$. Here we take

$$y_k = CX_k + w_k, \quad C = [c_1, c_2, \ldots c_n] \tag{6}$$

where w_k is a discrete-time, identically and independently distributed noise process in a continuous range $\mathbb{R}^1$; here take w_k as zero mean, white and Gaussian with density $N\left[0,\ \sigma_w^2\right]$. In fact, we can think of the system switching between the N

states $S_i = Ce_i = c_i$ for $i = 1, 2, \ldots N$ with the measurements of the state of the system contaminated by the additive noise process w_k.

The above equations taken together denote what is termed a hidden Markov model. The word hidden refers to the fact that the states are hidden in noise. The term Markov indicates that there is an underlying state vector which summarizes all that we need to know about the past of the system in order to proceed in predicting its future. A simple situation is depicted in Figure 2.

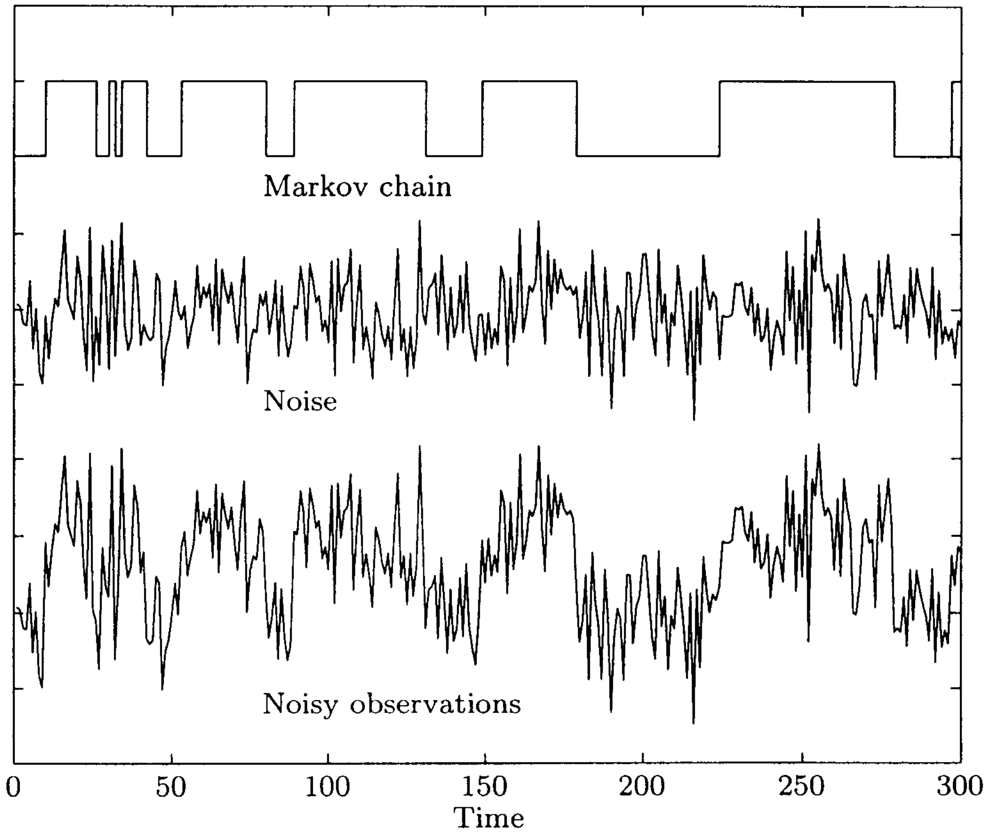


Figure 2: Binary Markov chain in noise.

A key signal processing task is to estimate the states of a hidden Markov model given the measurement data $y_0, y_1, \ldots y_k$. Ideally one would prefer a recursive signal processing scheme, which updates an estimate of the states at time $k - T$ given data up to time k, where T represents a delay in the processing so as to achieve improved estimates from future. In the first instance, signal processing algorithms assume knowledge of the transition probabilities A, the discrete-set states (parameters), namely $S_i = c_i$, and the statistics of the measurement noise process $w_k \sim N\left[0, \sigma_w^2\right]$. More sophisticated signal processing can simultaneously estimate both the parameters and the states of the model. In our situation where the transition probabilities can conceivably depend upon the time to the last transition, then one has to estimate this dependancy from the noisy data.

Of particular interest here is when elements of $A = (a_{ii})$ depend in an exponential manner on τ_k, as for example when for all i

$$a_{ii}(\tau) = a_{ii}(0) + (1 - a_{ii}(0)) \left[1 - e^{-D\tau}\right] \tag{7}$$

Here D is taken as the fractal dimension of the signal y_k. Clearly, when $D = 0$, then $a_{ii}(\tau) = a_{ii}(0)$ is independent of τ, and as D increases $a_{ii}(\tau) \to 1$ in an exponential manner. That is, as D increases there is a greater tendency for X_k to stay in the same state e_i as τ_k increases, and conversely there is more likelihood of a state transition if τ_k is small.

With knowledge of the model $\{A(\tau_k), C, \sigma_w^2\}$ then the state estimates

$$\hat{X}_{k|k} = E\left[x_k | X_0, X_1, \ldots, X_k\right] = P\left(x_k | X_0, X_1, \ldots, D_k\right)$$

which evolve as illustrated in Figure 3, are given from

$$\begin{aligned} \alpha_{k+1} &= B\left(y_{k+1}, C\right) A(\tau_k)\alpha_k, \quad \alpha_0 \\ \hat{X}_{k|k} &= \left(\underline{1}'\alpha_k\right)^{-1} \alpha_k \end{aligned} \tag{8}$$

where $\underline{1}' = [1, 1, \ldots 1]$. Here α_k is an unnormalized version of the conditional density $p(X_k | X_0, X_1, \ldots X_k)$ so that α_0 is the *a priori* density $p(X_0)$.

Maximum likelihood estimation of D involves multi-passes, both forward and backwards, through the data [9, 10].

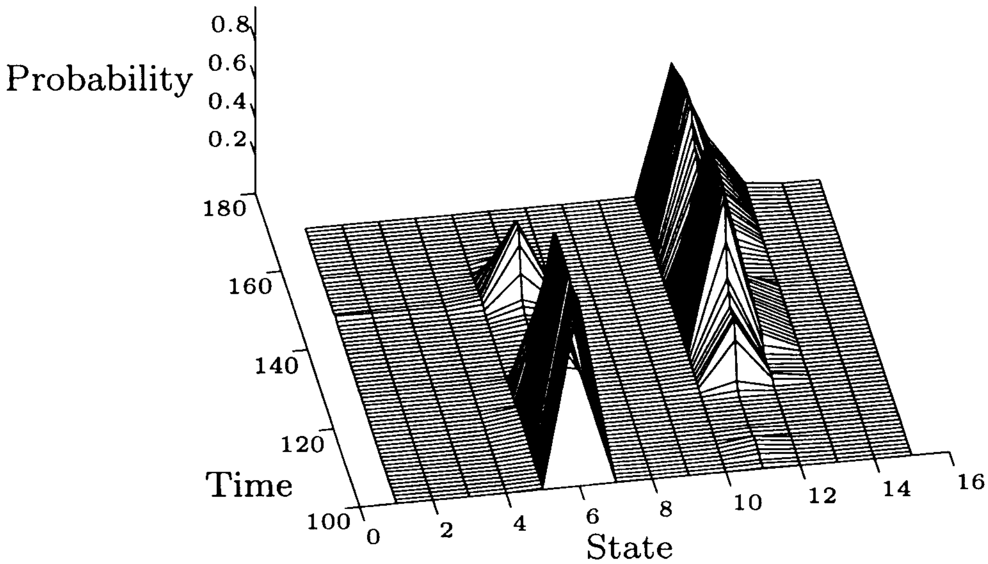


Figure 3: Evolution of state estimates.

On-line suboptimal methods based on recursive prediction error techniques can also be devised, see [6, 5]. The details are beyond the scope of this overview presentation.

In comparing how well various models and parametrizations fit the data, the key measure is simply the conditional probability p (model | data). In the absence of *a priori* information, the related probability p (data | model) is equivalent. In the neurophysiological model study, fractal models with fractal dimension in the range $D = 1.2$ to 1.6 were found to be most likely on the data tested, see [9].

3 Optimization via Dynamical Systems

In engineering applications, and in particular control applications, there is usually some underlying dynamical system description of a plant which has to be controlled by some control variable which, along with the states of the dynamical system, must satisfy certain constraints. For example, the control signals may be constrained so as not to exceed a certain magnitude. Since hardware is common to many industrial plants, the only competitive advantage of one plant over another is its control software. At the heart of this software are dynamical systems (recursive algorithms) which perform on-line optimization. These algorithms are driven by measurements from sensors placed on the process and their outputs drive the various actuators of the process. Contemporary research areas such as robotics, have brought to the fore novel control tasks. For example, in robotic dextrous hand-grasping, there are many fingers which must be co-ordinated so that in grasping and manipulating an object there is a balance of forces, excessive force is not used, and yet slipping is prevented.

Focusing on robotic hand-grasping, the existing optimization algorithms tend to use standard linear programming or non-linear programming techniques. Also, there are many ad-hoc devices supplied in the algorithms to achieve practical results. The challenge is to devise an on-line optimization which achieves well-conditioned optimal results, and rapid on-line calculations. For this task, we have proposed in [2, 1] that the friction constraints be viewed as the positive definiteness requirement of a certain matrix, while the force balancing constraints can be viewed as linear constraints on the elements of the positive definite matrix. The picture we have in mind then is of a cone, being the class of positive definite matrices, sliced by a hyper-plane, being the force balancing constraints, see Figure 4. The task is then to optimize the balancing of forces on this intersection of the cone and hyper-plane. Starting from an initial feasible solution at the intersection of the cone and hyper-plane, algorithms in the form of dynamical systems have been devised to converge to an optimal solution. It is important that the optimization be formulated so that there is a unique global minimum, and that the optimization is in essence a convex optimization task.

One of our first proposals for the dextrous hand-grasping problem requires a solution of a discrete-time Riccati equation modified to ensure projection of its solution into the constraint manifold. The index optimized is very similar to that which has been well studied for balancing controllability and observability in linear systems theory. It involves a term which penalises the forces at the fingertips, and a barrier function which prevents the constraints from being violated. There is a balance between these two requirements achieved in the optimization.

The optimization approach employed as expounded in [7] is to implement gradient flows of penalty functions on the smooth constraint manifolds of interest. Four key steps are formulations of the manifold, selecting cost function, choosing a Riemannian metric (or descent angle), and in discrete-time designing a step size. The "right" combination can result in elegant flow equations with linear (exponential) convergence properties to a global minimum. The "wrong" combination can result in "messy" equations which flow to local minima not the global minima.

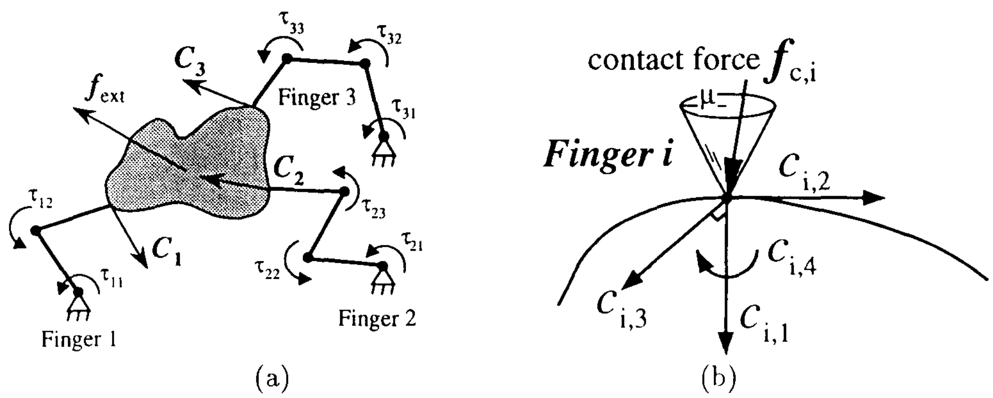


Figure 4: Object grasped by dextrous hand.

Decomposing the downhill search into two-dimensional geodesic searches, where step-size selection to achieve a minimum in the descent direction can be calculated analytically, results in highly efficient algorithms which are quadratically convergent. This is an area of current fruitful research.

To be more precise, consider the grasping situation depicted in Figure 4, with $c_{i,j}$ denoting the ith figure contact wrench (force) resolved in the j direction, then friction constraint requirements for N fingers are

$$\sqrt{c_{i,2}^2 + c_{i,3}^2} < \mu_i c_{i,1} \qquad \text{for} \quad i = 1, 2, \ldots N \tag{9}$$

Here μ_i is the coulomb fraction coefficient and the $j = 1$ direction is normal to the finger contact. These inequalities are equivalent to the positive definitiveness of $P = \text{blockdiag}\{P_1, P_2, \ldots P_N\}$ where

$$P_i = \begin{bmatrix} \mu_i c_{i,1} & 0 & c_{i,2} \\ 0 & \mu_i c_{i,1} & c_{i,3} \\ c_{i,2} & c_{i,3} & \mu_i c_{i,1} \end{bmatrix} \qquad \text{for} \quad i = 1, 2 \ldots N \tag{10}$$

The eigenvalues of P_i are

$$\lambda_i = \mu_i c_{i,1}, \quad \lambda_{2,3} = \mu_i c_{i,1} \pm \sqrt{c_{i,1}^2 + c_{i,3}^2} \tag{11}$$

The equilibrium or force balance constraint is in the form $f_{ext} = Wc$, wheref_{ext} is the extended force vector, c is the vector of forces $c_{i,j}$ and W is the grip transformation matrix describing the geometric relation between contact wrench space and object co-ordinate frame [11]. This together with the structural constraints on P, i.e. $P_{i,21} = P_{i,12} = 0$ and $P_{i,11} = P_{i,22} = P_{i,23}$, and the blockdiagonal constraint can be represented as

$$A\text{vec}(P) = q \tag{12}$$

The index selected for optimization is

$$\phi(P) = tr\left(P + \rho P^{-1}\right) \tag{13}$$

for some scalar weighing $\rho > 0$.

The first term penalizes $c_{i,j}$ and the second is a barrier penalty function for the contraint (9). A very suitable Rienammenian metric for two vectors ξ_1, ξ_2 in the tangent space of the positive definite constraint manifold $P > 0$ is

$$\langle\langle \xi_1, \xi_2 \rangle\rangle = tr\left(P^{-1}\xi_1 P^{-1}\xi_2\right). \tag{14}$$

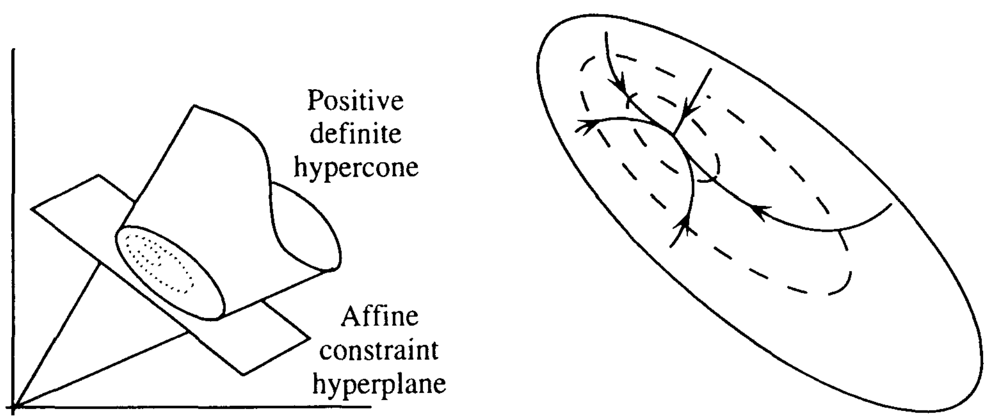
This leads to the gradient flow in the absence of the linear constraint (12) as

$$\dot{P} = \rho I - P^2 \tag{15}$$

and projecting into the hyperplane (12) we have the flow

$$\mathrm{vec}(\dot{P}) = \left[I - A'(AA')^{-1}\right] \mathrm{vec}(\rho I - P^2) \tag{16}$$

from which discrete-time flows can be derived, see [1, 2]. Figure 5 depicts the situation.



(a) Positive definite cone (b) Gradient flows

Figure 5: Positive definite hypercone with affine constraints and gradient flows on the constraint hyperplane.

These gradient flows are appealing because of their mathematical elegance, but this is also their limitation. The flows are smooth, yielding exponential convergence. Quadratic convergence and faster convergence can be achieved using recursions such as

$$\mathrm{vec}(P_{k+1}) = \left[I - A'\left(AA'\right)^{-1}\right] \mathrm{vec}\left(P_k - \alpha_k(\rho P_k^{-1} - P_k)\right)$$

for $0 < \alpha_k < 1$ which are more "violent". The α_k is selected to minimize the cost term (13) at each iteration.

In cases where the manifolds are not compact and convergence is not guaranteed, do flow equations exhibit chaos? Also, in more sophisticated optimization situations, perhaps with local minima not the global minima, must we use chaos itself to efficiently side-step a local minima?

4 Conclusions

This paper has summarized some recent research in signal processing and control in which chaos is the central fascination on the one hand and the allure to achieve improved results on the other hand. It seems clear from our studies that the human brain in its signal processing makes use of chaos for improved efficiency and performance. The challenge is for systems engineers working in the area of control applications to exploit the potential of chaos for enhanced control and on-line optimization. We have a long way to go.

References

[1] M. Buss, H. Hashimoto and J.B. Moore, "Dextrous Hand Grasping Force Optimization", *IEEE Transactions on Robotics and Automation*, to appear, see also *Proceedings of the IEEE Conference on Robotics and Automation*, pp 1034–1039, Nagoya, Japan, May 1995.

[2] M. Buss, L. Faybusovich, and J.B. Moore, "Dikin-Type Semidefinite Programming Algorithms for Dextrous Grasping Force Optimization", *International Journal of Robust and Nonlinear Control*, submitted.

[3] S.H. Chung, V. Krishnamurthy and J.B. Moore, "Adaptive Processing Techniques based on Hidden Markov Models for Characterizing Very Small Channel Currents Buried in Noise and Deterministic Interferences", *Philosophical Transactions of Royal Society*, Vol. 334, 1991, pp. 357-384.

[4] S.H. Chung, J.B. Moore, and L. Xia, P. Gage, L.S. Premkumar, "Characterization of Single Channel Currents using Digital Signal Processing Techniques based on Hidden Markov Models", *Phil. Trans. of Royal Society*, London B, Vol. 329, September 1990, pp. 265-285, see also *Proc. of Australian Physiological Society*, Newcastle, September 1989.

[5] I.B. Collings, V. Krishnamurthy and J.B. Moore, "On-line Identification of Hidden Markov Models via Recursive Prediction Error Techniques", *IEEE Transactions on Signal Processing*, vol 42, no 12, pp 3535-3539, see also *Proc. of IFAC World Congress*, Sydney, July 1993, vol V, 423-426.

[6] R.E. Elliott, L. Aggoun and J.B. Moore, *Hidden Markov Models: Estimation and Control*, Springer-Verlag, 1995, (361 pages).

[7] U. Helmke and J.B. Moore, *Optimization and Dynamical Systems*, Springer-Verlag, 1993, (390 pages).

[8] V. Krishnamurthy, J.B. Moore and S.H. Chung, "Hidden Markov Model Signal Processing in Presence of Unknown Deterministic Interferences", *IEEE Trans. on Automatic Control*, Vol 38, Jan 1993, pp.146-152, see also *Conf. on Dec. and Control*, Brighton UK, December 1991, pp. 662-667.

[9] V. Krishnamurthy, J.B. Moore and S.H. Chung, "On Hidden Fractal Model Signal Processing", Vol. 24, Issue No. 2, August 1991, pp. 177-191, see also *Proc. of Int. Symp. on Information Theory and its Applications*, Hawaii, December 1990, pp. 243-246.

[10] V. Krishnamurthy, J.B. Moore and S.H. Chung, "Signal Processing of semi-Markov Models with Exponentially Decaying States", *Conference on Dec. and Control*, Brighton UK, December 1991, Vol 3, pp. 2744-2749.

[11] M. Mason and J.Salisbury, *Robot Hands and the Mechanics of Manipulation*, Cambridge, Massachusetts: MIT Press, 1995.

Commentary by K. L. Teo

This paper first considers discrete dynamical systems which switch between states. The states are hidden in noise and there is an underlying state vector which summarizes all that are to know about the past of the system in order to proceed in predicting its future. This situation is termed Hidden Markov model. The technical approach is reviewed for investigating chaos in a hidden Markov model for discrete-state systems and its application to study cell channel currents.

The focus of the paper then shifts to the topic of optimization of system via dynamical systems. Usually, chaos is avoided in performing a system design or optimization. For an optimization problem in the area of robotic is used for illustration to a technical approach summarized in the paper.

The paper is both informative and interesting.

Commentary by T. Vincent

The author raises an interesting question in the abstract, "Can our computing machines and control systems enhance their intelligence by clever introduction of chaos?" One computing machine of interest to the author is the brain and he points to some results which suggest that the underlying processes and synapse levels are in fact chaotic. Another area of interest discussed by the author is optimization. While presenting an optimization approach, which uses gradient flows of penalty functions on smooth constraint manifolds, the question is raised if it is possible to use chaos to side-step local minima? Using chaos in this way to solve global optimization problems is a very attractive idea. No doubt we can look forward to progess in this area.

Six papers in this volume do address the authors question in regard to control systems. My own contribution uses the ergotic nature to chaos to eventually bring the system to a point in state space where some control action will be effective. The targeting methods discussed in the papers by Glass, Kostelich, and Ott represents a more 'clever' way to use chaos to accomplish the same objective. Mareels observes that chaotic behavior could be beneficial in achieving control objectives in adaptive control systems. Grantham shows that suppressing a single frequency can convert chaotic motion into a desired periodic one. I think the answer to Moore's question is a definite yes, but I also agree with his assessment, "We have a long way to go."