Continuous-time nonlinear signal processing: A neural network based approach for gray box identification*

R. Rico-Martínez, J. S. Anderson and I. G. Kevrekidis
Department of Chemical Engineering, Princeton University,
Princeton NJ 08544

Abstract

Artificial neural networks (ANNs) are often used for short term discrete time series predictions. Continuous-time models are, however, required for qualitatively correct approximations to long-term dynamics (attractors) of nonlinear dynamical systems and their transitions (bifurcations) as system parameters are varied. In previous work we developed a black-box methodology for the characterisation of experimental time series as continuous-time models (sets of ordinary differential equations) based on a neural network platform. This methodology naturally lends itself to the identification of partially known first principles dynamic models, and here we present its extension to "gray-box" identification.

1 Introduction

Artificial Neural Networks (ANNs) have proven to be a valuable tool in nonlinear signal processing applications. Exploiting ideas common to nonlinear dynamics (attractor reconstruction) and system identification (ARMA models), methodologies for the extraction of nonlinear models from experimental time series have been developed (e.g. [1, 2]) and applied to experimental data. In previous work, we have discussed some inherent limitations of these techniques (based on *discrete-time* schemes) in characterizing the instabilities and bifurcations of nonlinear systems depending on operating parameters.

An alternative approach, resulting in *continuous-time* models (sets of Ordinary Differential Equations (ODEs)), also based on a neural network platform, was devised and implemented [3, 4, 5]. The approximations constructed in that

*This work was partially supported by ARPA/ONR, the Exxon Education Foundation and an NSF PYI award. RRM acknowledges the support of CONACyT through a fellowship.

0-7803-2026-3/94 $4.00 © 1994 IEEE

work can be described as black-box; no insight from first principles modeling of the system was incorporated in them.

In this work we extend the approach to cases where portions of the algebraic forms of the set of ODEs describing the dynamical evolution of the system are known. We attempt to capture the behavior of the overall system by "hard-wiring" the known parts and approximating the unknown parts using a neural network (gray-box identification).

In what follows we first briefly outline our black-box approach for the identification of continuous systems. This discussion naturally leads to the extension to gray-box identification. Finally, we illustrate its use through an application to the modeling of a reacting system with complicated nonlinear kinetics.

2 Black-box approach

Consider the autonomous ODE

$$\dot{\vec{X}} = \mathbf{F}(\vec{X};\vec{p}) \tag{1}$$

$$\vec{X} \in \mathcal{R}^n, \qquad \vec{p} \in \mathcal{R}^p, \qquad \mathbf{F}: \mathcal{R}^n \times \mathcal{R}^p \mapsto \mathcal{R}^n$$

where $\vec{X}$ is the vector of state variables, $\vec{p}$ is the vector of operating parameters and $\dot{\vec{X}}$ is the vector of derivatives of the state variables with respect to time. In previous work we showed a way of constructing such a set of ODEs from discrete-time experimental measurements of the state variables only ([3, 4, 5], see also [6]). We embedded the training of a neural network that approximates the function $\mathbf{F}(\vec{X};\vec{p})$ in a numerical integrator scheme. Both explicit and implicit integrators can be (and have been) used. In addition, we illustrated how the approach can be used when time series of only a single state variable are available.

Consider the simple implicit integrator (trapesoidal rule) formula for Eq. 1:

$$\vec{X}_{n+1} = \vec{X}_n + \frac{h}{2}[\mathbf{F}(\vec{X}_n;\vec{p}) + \mathbf{F}(\vec{X}_{n+1};\vec{p})] \tag{2}$$

where h is the time step of the integration, $\vec{X}_n$ is the value of the vector of states at time t and $\vec{X}_{n+1}$ is the (approximate) result of integrating the set of ODEs to time $(t+h)$. Figure 1(a) schematically depicts a neural network constructed using this numerical integrator as a template. The boxes labeled "neural network" represent the *same* neural network evaluated with two different sets of inputs for each training vector. Given the implicit nature of the integrator, the "prediction" of the integration depends on itself. Training was therefore done using standard recurrent network training ideas [7, 8]. Alternatively, a nonlinear algebraic equation solver can be used, coupled with the training, to solve exactly for the predicted value at every iteration and for every training

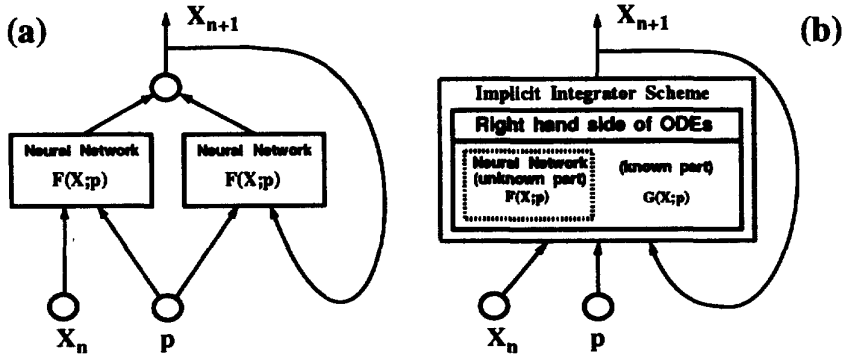

Figure 1: (a) Schematic of the evaluation of a neural network embedded in the implicit integrator (trapezoidal rule) of Eq. (2). The implicit dependence of the prediction of the state on itself results in a backward (recurrent) connection. (b) Schematic of the evaluation of a neural network for the gray-box approach. The known part of the model ($G(X;p)$) is evaluated along with the unknown part ($F(X;p)$), approximated by a neural network. In order to calculate errors (for training) the contribution of the known and unknown parts are combined using the integrator to give the state of the system at the next sampling time.

vector. Further details can be found in references [4, 5]. The use of explicit integrators is discussed in [3]. This identification procedure has been tested for experimental systems exhibiting complicated dynamics (see e.g. [3, 4]).

3 Gray-box approach

The approach discussed above can be combined with first principles modeling for cases where the full state vector is known while the understanding of the modeling of the system is only partial. Such an example is encountered in modeling reacting systems, when the kinetics of the reaction are not known *a priori* while inflow and outflow or heat transfer are well understood and easily modeled.

As in the case of black-box approximations, we embed the training of the neural network in a numerical integrator scheme. For gray-boxes, the known part of the right-hand-side of the ODEs is explicitly calculated ("hardwired") and the neural network is trained to approximate only the unknown parts.

Let us assume for the purposes of the illustration presented here that the first principles model of a given system takes the simple form:

$$\dot{\vec{X}} = \mathbf{G}(\vec{X};\vec{p}) + \mathbf{F}(\vec{X};\vec{p}) \tag{3}$$

where $\mathbf{G}(\vec{X};\vec{p})$ represents the known part of the model and $\mathbf{F}(\vec{X};\vec{p})$ is the unknown part. Note that the methodology is not restricted to models of the additive form of Eq. (3).

Figure 1(b) schematically depicts the training procedure for an implicit integrator. A "global" network is used to predict the state at the next time step. Some of the weights and nodes in this network are fixed because of the "known" part of the model; some are fixed because they pertain to the integration scheme and its constants. A neural "sub"-network is also contained in the scheme, which will upon training approximate the unknown parts of the right-hand-side of the system ODEs. We again use the implicit integrator of Eq. (2) as the basis for training this network, which – due to the implicit nature of the integrator – has recurrent connections and therefore requires multiple evaluations.

4 An illustrative example

In order to illustrate the capabilities of the gray-box approach we will make use of simulated data from a model reacting system [9]. It consists of a well-stirred reactor in which a single irreversible reaction $A \rightarrow B$ occurs on a catalytic surface. The mass balances for species A on the catalytic surface and the gas phase take the general (dimensionless) form:

$$\begin{aligned} \frac{d\theta}{d\tau} &= K_a \Pi(1-\theta) - K_d \theta e^{-\frac{(\alpha^*\theta+\beta)}{\gamma}} - K_R \theta e^{-\frac{1}{\gamma}} \\ \frac{d\Pi}{d\tau} &= 1 - \Pi + \Pi^*[K_d \theta e^{-\frac{(\alpha^*\theta+\beta)}{\gamma}} - K_a \Pi(1-\theta)] \end{aligned} \tag{4}$$

where θ is the fractional coverage of the catalytic surface, Π is the partial pressure of the reactant in the gas phase, γ is the dimensionless temperature, τ is the dimensionless time and K_a, K_R, K_d, α^*, β and Π^* are constants. This has been suggested as one of the simplest models that can give rise to oscillations in isothermal catalytic reactions; its main characteristic is the coverage-dependent desorption activation energy (the $e^{-\frac{\alpha^*\theta}{\gamma}}$ term in Eq. (4)) caused by adsorbate-adsorbate interactions.

To illustrate the dependence of the dynamics on an operating parameter, we obtained time series from this system for several values of the dimensionless temperature γ (keeping the remaining parameters $K_a = 35$, $\alpha^* = 30$, $K_d = 350$, $\Pi^* = 0.36$, $K_R = 8.5$ and $\beta = 0.2$ constant). Depending on the value of γ, the system may evolve towards a steady state, towards oscillatory behavior, or to either of the two depending on the initial conditions. The variegation in long-term dynamics makes this example a good test of the approximating capabilities of the neural network.

Figure 2 shows the bifurcation diagram for this system: a branch of steady states undergoes a *subcritical* Hopf bifurcation to oscillatory behavior for $\gamma \approx$

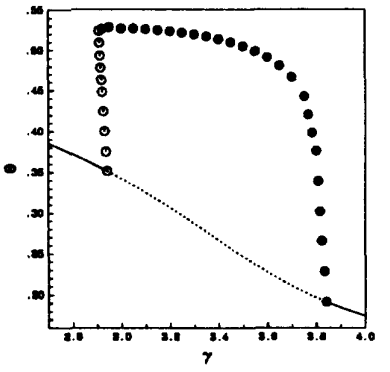

Figure 2: Bifurcation diagram for the single species surface reaction system with respect to the dimensionless temperature γ. Solid lines denote stable steady states, dashed lines unstable steady states, open circles unstable limit cycles and filled circles stable limit cycles. The maximum θ of the periodic trajectory at each value of γ is marked.

2.941. There is a small range of values of γ where a stable large amplitude oscillation coexists with a stable steady state (starting at about $\gamma \approx 2.9076$). As γ is increased the system exhibits, as its sole long-term attractor, a large amplitude limit cycle that disappears at $\gamma \approx 3.841$ via another (now *supercritical*) Hopf bifurcation.

Figure 3 shows phase portraits of the system for several values of γ in the range of the bifurcation diagram of Fig. 2.

5 Network construction and results

Using data representative of the periodic phenomena described above, we tested the neural network ODE-gray-box algorithm for identification. The training set included several time series (θ and Π vs τ) for values of γ before the subcritical Hopf (including the region of bistability), after the subcritical Hopf (limit cycle behavior), as well as after the supercritical Hopf at high values of γ.

For our illustration we assume that all terms in Eq. 4 are known except for the term representing the rate of desorption of the reactant from the catalytic surface. That is, we replace the term $K_d\theta e^{-\frac{(\alpha^*\theta+\beta)}{\gamma}}$, with an *unknown* function $f(\theta,\gamma)$ to be approximated through a neural network. The gray model we seek to construct is of the form:

$$\frac{d\theta}{d\tau} = K_a\Pi(1-\theta) - f(\theta,\gamma) - K_R\theta e^{-\frac{1}{\gamma}}$$

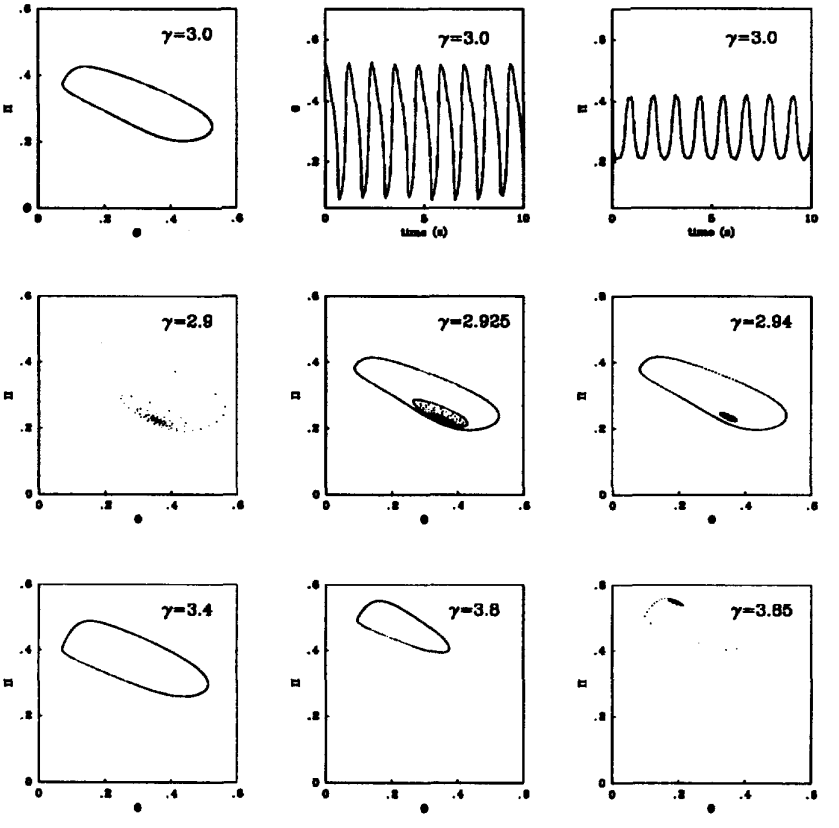

Figure 3: Long term attractors for γ values in the range of the bifurcation diagram of Fig. 2. Top row: phase portrait of the stable limit cycle at $\gamma = 3.0$ along with segments of the two corresponding time series. For the stable steady states ($\gamma = 2.9$ and $\gamma = 3.85$) phase portraits of transients approaching the steady state are shown. In the regions of bistability (γ in the range (2.925,2.94)) the unstable (and thus experimentally unobservable) limit cycle in the interior of the large amplitude stable limit cycle is also drawn.

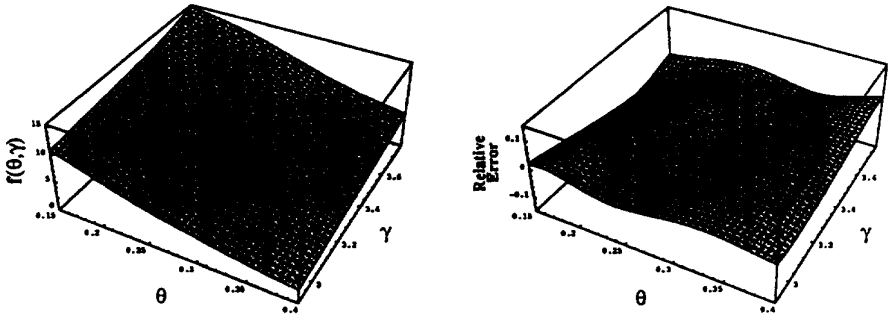

Figure 4: Predicted desorption rate as a function of surface coverage (θ) and dimensionless temperature (left) and relative prediction error (right). The plot on the right shows the difference of the predicted minus the actual desorption rate normalized by the actual rate.

$$\frac{d\Pi}{d\tau} = 1 - \Pi + \Pi^*[f(\theta,\gamma) - K_a\Pi(1-\theta)] \tag{5}$$

The (feedforward) neural sub-network, embedded in the numerical integrator of Fig. 1(b), involves two inputs (θ and γ), one output ($f(\theta,\gamma)$) and six neurons with sigmoidal (tanh-type) activation function in each of the two hidden layers. The derivatives of the error measure (energy function) with respect to network parameters needed for the training algorithm are obtained using the chain rule and (due to the recurrence) the implicit function theorem.

The training set consisted of a total of 2950 points allocated in the following manner: 250 points for $\gamma = 2.9$, 450 for $\gamma = 2.91$, 500 for $\gamma = 2.925$, 500 for $\gamma = 2.94$, 250 for $\gamma = 3.0$, 250 for $\gamma = 3.0$, 250 for $\gamma = 3.2$, 250 for $\gamma = 3.4$, 250 for $\gamma = 3.8$ and 250 for $\gamma = 3.85$. The time step of the integrator was 0.06 dimensionless units for all the time series used (roughly one twentieth of the period of the oscillation observed at $\gamma = 3.0$). More points are included in the region of multistability in an effort to capture accurately the hysteresis phenomena. Training was performed using a conjugate gradient algorithm with frequent restarts (see [4, 5] for a discussion). Convergence was achieved after approximately 300 network parameter updates.

The sub-network succeeds in capturing the basic form of the behavior of the

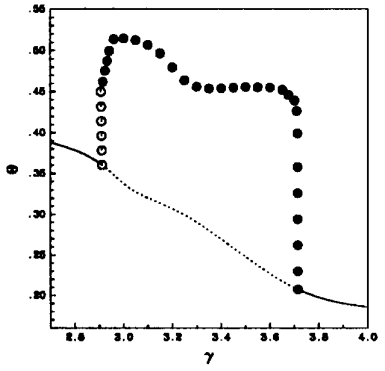

Figure 5: Predicted bifurcation diagram for the single species surface reaction system with the gray-box neural network approximation of the desorption rate.

rate of desorption with respect to θ and γ (surface coverage and temperature). Figure 4 compares the actual desorption rate (as a function of (θ, γ)) with the network predictions. More importantly, the dynamic behavior (including the *infinite-time* attractors) of the system (Eq. 5) also compares favorably with the original system (Eq. 4). Figure 5 shows the predicted bifurcation diagram using the form of the desorption rate given by the network. The network correctly predicts a subcritical Hopf bifurcation at low γ, as well as a supercritical Hopf bifurcation at higher values of γ (at a slightly lower value of γ than for the original system, Fig. 2).

The neural network gray-box approximation can be used to extract important mechanistic information pertaining to the fitted step – and thus possibly discriminate among rival candidate first principles models. For example, Fig. 6 shows that the network predicts a linear dependence of the logarithm of the desorption rate versus $\frac{1}{\gamma}$ at constant θ, in agreement with desorption being an activated process. Fig. 6 shows also that the predicted *slopes* of these plots (and thus, the activation energies) vary linearly with θ, consistent with an assumption of attractive adsorbate-adsorbate interactions (as was indeed the case).

6 Summary

We have extended a previously developed black-box neural network methodology for the characterisation of experimental systems as continuous-time models, so as to allow the identification of unknown parts of first principles models. Such modeling efforts incorporate the insight obtained from the first principles modeling (algebraic forms of the ODEs describing the dynamical evolution of the system) in a neural network framework capable of approximating (after training)

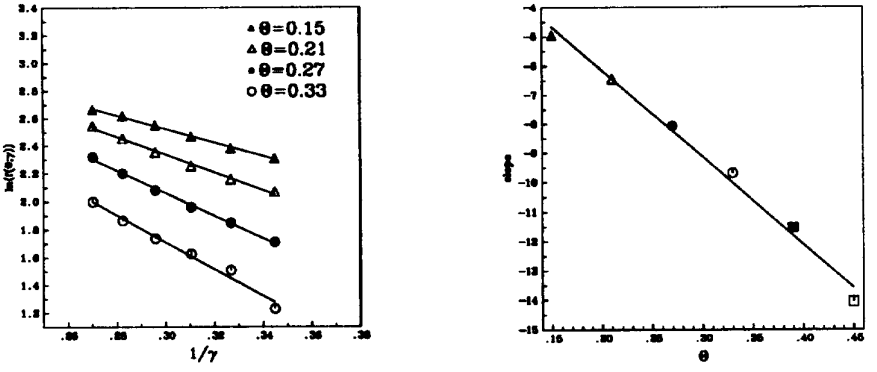

Figure 6: The linear dependence of the natural logarithm of the desorption rate with respect to $\frac{1}{\gamma}$ at constant θ is correctly captured by the neural network gray-box approximation (left); furthermore the predicted slope of the lines varies linearly with θ (right), in agreement with the assumption of adsorbate interactions used to generate the training data.

unknown parts of the model.

The capabilities of this gray-box approach were illustrated using a single species surface reaction system. In this illustration we assumed that the expression for the rate of desorption of the reactant is not known and approximated it through a neural network. Both the short- and long-term dynamic behavior of the system is well approximated by the hybrid model resulting from training. Furthermore, a study of the properties of the fitted desorption rate may yield insight in the physical mechanisms underlying it, and thus possibly assist in discriminating among rival first principles models.

Discrete-time models (based on neural networks) are trained to predict the *result* of integrating the model equations over some time period. It is difficult to "unravel" the contribution of known parts of the model to this result from the contribution of the unknown terms. When, on the other hand, the equations themselves are approximated (as opposed to the result of integrating them), the procedure naturally lends itself to incorporating processes whose modeling is established to the gray-box model.

The type of overall network presented here (with some parts of its architecture available for training, and some other parts fixed by either the known parts of the model or the integrator scheme) may prove to be a valuable tool towards understanding the dynamics of experimental systems. The particular choice of recurrent nets templated on implicit integrators presented here is motivated by the anticipated stiffness of chemical kinetic equations. Feedforward implementations based on explicit integrators are also possible. We are currently working on variants of training algorithms for recurrent nets and their implementation

on parallel computers.

References

[1] A. S. Lapedes and R. M. Farber. Nonlinear signal processing using neural networks: Prediction and system modeling. *Los Alamos Report LA-UR 87-2662* (1987).

[2] A. S. Weigend and N. A. Gershenfeld. Time series prediction: Forecasting the future and understanding the past. *Addison-Wesley* (1993).

[3] R. Rico-Martínes, K. Krischer, I. G. Kevrekidis, M. C. Kube and J. L. Hudson. Discrete- vs. continuous-time nonlinear signal processing of Cu electrodissolution data. *Chem. Eng. Comm.*, vol. 118, pp. 25–48 (1992).

[4] R. Rico-Martínes. Neural networks for the characterisation of nonlinear deterministic systems. *Ph. D. Thesis*, Department of Chemical Engineering, Princeton University (1994).

[5] R. Rico-Martínes and I. G. Kevrekidis. Continuous-time modeling of nonlinear systems: A neural network approach. *Proc. 1993 IEEE Int. Conf. Neural Networks*, IEEE Publications, vol. III, pp. 1522–1525 (1993).

[6] S. R. Chu and R. Shoureshi. A neural network approach for identification of continuous-time nonlinear dynamic systems. *Proc. of the 1991 ACC*, vol. 1, pp. 1–5 (1991).

[7] F. J. Pineda. Generalisation of back-propagation to recurrent neural networks. *Phys. Rev. Letters*, vol. 59, pp. 2229–2232 (1987).

[8] L. B. Almeida. A learning rule for asynchronous perceptrons with feedback in a combinatorial environment. *Proc. IEEE 1st Ann. Int. Conf. Neural Networks*, San Diego, CA., pp. 609–618 (1987).

[9] I. Kevrekidis, L. D. Schmidt and R. Aris. Rate multiplicity and oscillations in single species surface reactions. *Surf. Sci.*, vol. 137, pp. 151-166 (1984).