

Identification of Nonlinear Dynamic Systems by Combining Equilibrium and Off-Equilibrium Information

W. E. LEITHEAD^{1,2}

¹Hamilton Institute, National University of Ireland, Maynooth, Co. Kildare, Ireland,
email : w.leithead@eee.strath.ac.uk

²Dept. of Electronics & Electrical Engineering, University of Strathclyde, Glasgow G1 1QE, UK.

Abstract – An identification methodology for nonlinear dynamic systems using Gaussian process prior models is presented. It combines equilibrium and off-equilibrium information. The advantages of so doing are highlighted.

I. INTRODUCTION

Consider a nonlinear dynamic system

$$\mathbf{x}_{i+1} = \mathbf{g}(\mathbf{x}_i, \mathbf{r}_i), \quad y_i = h(\mathbf{x}_i, \mathbf{r}_i) \quad (1)$$

with state $\mathbf{x} \in \mathcal{R}^n$, input $\mathbf{r} \in \mathcal{R}^m$ and output $y \in \mathcal{R}$. The input and output are measured but the state is not. When the unmeasured state can be eliminated, the equivalent input-output representation of the system is

$$y_i = H(\mathbf{y}_{i-1}, \dots, \mathbf{y}_{i-n}, \mathbf{r}_i, \mathbf{r}_{i-1}, \dots, \mathbf{r}_{i-n}) \quad (2)$$

and the task of identifying the system from the measured data is equivalent to identifying the nonlinear map $H(\cdot)$. The objective considered here is to identify from measured input-output data the nonlinear system, (1), *i.e.* the nonlinear map $H(\cdot)$. Since measured data is generally noisy, it is natural to work within a probabilistic framework. The probabilistic description of the nonlinear map adopted is that of a Gaussian process prior model within a Bayesian probability context [1]. This description has previously been applied to the identification of non-dynamic nonlinear relationships, including time series analysis, see, for example, [1]. In a statistically rigorous investigation [2], the latter is shown to generally out-perform NN etc methods.

When tasked with identifying a nonlinear dynamic system, a number of separate input-output measurement sets are typically available. For some of these, the input-output data set is local to an equilibrium operating point. In this case, an equilibrium operating point and an associated local linear model for the dynamic system can be identified using some well-established technique. Accordingly, the data, from which the nonlinear dynamic system is to be identified, consists of a number of equilibrium operating points and their local linear models together with some off-equilibrium input-output transient data. In this paper, an identification method, based on Gaussian process prior models, for nonlinear systems that can combine equilibrium information and off-equilibrium information is presented. In particular, the benefits from combining the equilibrium and off-equilibrium information are highlighted.

II. IDENTIFICATION OF NONLINEAR MAPS

Consider a nonlinear map

$$y = f(\mathbf{z}) ; \quad \mathbf{z} \in \mathcal{R}^p, y \in \mathcal{R}. \quad (3)$$

representing some static nonlinear relationship. Suppose N measurements, $\{(\mathbf{z}_i, y_i)\}_{i=1}^N$, of the value of the function, $f(\mathbf{z})$, are known with

$$y_i = f(\mathbf{z}_i) + n_i \quad (4)$$

where n_i is independent Gaussian noise. Denote them by S . It is of interest, here, to use this data to learn the map $f(\mathbf{z})$ or, more precisely, to determine a probabilistic description of $f(\mathbf{z})$ on the domain, D , containing the data. Note, it is assumed the explanatory variable, \mathbf{z} , is noise free and that this is a regression formulation. It is natural to work within a probabilistic framework. The probabilistic description of the nonlinear map adopted is that of a Gaussian process prior model within a Bayesian probability context [1].

A. Gaussian process prior models

A brief overview of Gaussian process prior models within a Bayesian probability context is given below. For further details see [1].

The probabilistic description of the map, $f(\mathbf{z})$, is the stochastic process, $f_{\mathbf{z}}$, and the $E[f_{\mathbf{z}}]$, as \mathbf{z} varies, is interpreted to be a fit to $f(\mathbf{z})$. By necessity to define the stochastic process, $f_{\mathbf{z}}$, the probability distributions of $f_{\mathbf{z}}$ for every choice of value of $\mathbf{z} \in D$ are required together with the joint probability distributions of $f_{\mathbf{z}_i}, i = 1, \dots, k$, for every choice of finite sample, $\{\mathbf{z}_1, \dots, \mathbf{z}_k\}$, of D , for all $k > 1$. Of course, the joint probability distributions of lower dimensionality must be the marginal distributions of those of higher dimensionality. Given the joint probability distribution for $f_{\mathbf{z}_i}, i = 1..N$, and the joint probability distribution for $n_i, i = 1..N$, the joint probability distribution for $y_i, i = 1..N$, is readily obtained as their product since the measurement noise, n_i , and the $f(\mathbf{z}_i)$ (and so the $f_{\mathbf{z}_i}$) are statistically independent. S is a single event belonging to the joint probability distribution for $y_i, i = 1..N$.

In the Bayesian probability context, the prior belief is placed directly on the probability distributions describing $f_{\mathbf{z}}$ which are then conditioned on some information, S , to determine the posterior probability distributions. In particular, for the Gaussian process prior models considered here, the prior probability distributions for the $f_{\mathbf{z}}$ are chosen to be Gaussian. To complete the statistical description, requires only a definition of the mean, $\bar{f}_{\mathbf{z}}$, for all \mathbf{z} in D and

the covariance function, $C(\mathbf{f}_{z_i}, \mathbf{f}_{z_j}) = E[\mathbf{f}_{z_i}, \mathbf{f}_{z_j}]$, for all \mathbf{z}_i and \mathbf{z}_j in D . The resulting posterior probability distributions are also Gaussian. The choice of Gaussian probability distributions may seem strangely restrictive initially, but recall that this is simply a prior on the relevant stochastic process space and so places few inherent restrictions on the class of nonlinear functions that can be modelled. Indeed, it can be shown that the result is, in fact, a Bayesian form of kernel regression model [3] subsuming, amongst others, RBF, spline and many neural network models [1]. The Gaussian process prior model is non-parametric in the sense that the imposition of a specific parametric structure is avoided. This model is used to carry out inference as follows.

The measurement noise, n_i , has covariance Δ_{ij} and is statistically independent of \mathbf{f}_{z_i} . Hence, the means and covariance functions for the measured values, y_i , are simply

$$\begin{aligned} \bar{y}_i &= \bar{\mathbf{f}}_{z_i} \quad ; \quad C(y_i, y_j) = C(\mathbf{f}_{z_i}, \mathbf{f}_{z_j}) + \Delta_{ij} \\ C(y_i, \mathbf{f}_{z_j}) &= C(\mathbf{f}_{z_i}, \mathbf{f}_{z_j}) \end{aligned} \quad (5)$$

Training procedure

In general, the mean, $\bar{\mathbf{f}}_z$, and covariance function, $C(\mathbf{f}_{z_i}, \mathbf{f}_{z_j})$, are dependent on some hyperparameters, ξ . The likelihood of S is

$$L(S) = p(S | \xi) = \frac{1}{(2\pi)^{N/2} \det(\mathbf{P})} \exp(-\frac{1}{2} \mathbf{Y}^T \mathbf{P}^{-1} \mathbf{Y}) \quad (6)$$

where $\mathbf{Y} = [y_1 - \bar{y}_1, \dots, y_N - \bar{y}_N]^T$ and $\mathbf{P}_{ij} = C(y_i, y_j)$. To obtain a model given the data, S , the hyperparameters, ξ , are adapted to maximise the likelihood, $p(S | \xi)$. The model then consists of the data and the values of the hyperparameters so obtained.

Prediction procedure

Let $\bar{\mathbf{F}} = [(\mathbf{f}_{x_1} - \bar{\mathbf{f}}_{x_1}), \dots, (\mathbf{f}_{x_N} - \bar{\mathbf{f}}_{x_N})]^T$ for any finite sample, $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, of D . By Bayes theorem, $p(\bar{\mathbf{F}} | S) = p(\bar{\mathbf{F}}, S) / p(S)$, where $p(S)$ acts as a normalising constant. Hence, from the Gaussian nature of the probability distributions

$$p(\bar{\mathbf{F}} | S) \propto \exp\left[-\frac{1}{2} \begin{bmatrix} \bar{\mathbf{F}}^T & \mathbf{Y}^T \end{bmatrix} \begin{bmatrix} \Lambda_{FF} & \Lambda_{YF}^T \\ \Lambda_{YF} & \Lambda_{YY} \end{bmatrix}^{-1} \begin{bmatrix} \bar{\mathbf{F}} \\ \mathbf{Y} \end{bmatrix}\right] \quad (8)$$

where

$$(\Lambda_{FF})_{ij} = C(\mathbf{f}_{x_i}, \mathbf{f}_{x_j}) \quad ; \quad (\Lambda_{YF})_{ij} = C(\mathbf{f}_{z_i}, \mathbf{f}_{x_j}) \quad ; \quad \Lambda_{YY} = \mathbf{P}$$

Both Λ_{FF} and Λ_{YF} depend on the \mathbf{x}_i . Applying the partitioned matrix inversion lemma, it follows that

$$p(\bar{\mathbf{F}} | S) \propto \exp\left[-\frac{1}{2} (\bar{\mathbf{F}} - \hat{\mathbf{F}})^T \Lambda^{-1} (\bar{\mathbf{F}} - \hat{\mathbf{F}})\right] \quad (9)$$

with

$$\hat{\mathbf{F}} = \Lambda_{YF}^T \Lambda_{YY}^{-1} \mathbf{Y} \quad , \quad \Lambda = \Lambda_{FF} - \Lambda_{YF}^T \Lambda_{YY}^{-1} \Lambda_{YF} \quad (10)$$

For a single value of \mathbf{z} , (8) becomes

$$\begin{aligned} p((\mathbf{f}_z - \bar{\mathbf{f}}_z) | S) \\ \propto \exp\left[-\frac{1}{2} ((\mathbf{f}_z - \bar{\mathbf{f}}_z) - \hat{\mathbf{f}}_z)^T \Lambda_z^{-1} ((\mathbf{f}_z - \bar{\mathbf{f}}_z) - \hat{\mathbf{f}}_z)\right] \end{aligned} \quad (11)$$

with

$$\hat{\mathbf{f}}_z = \Lambda_{Yz}^T \Lambda_{YY}^{-1} \mathbf{Y} \quad , \quad \Lambda_z = \Lambda_{zz} - \Lambda_{Yz}^T \Lambda_{YY}^{-1} \Lambda_{Yz} \quad (12)$$

Therefore, the prediction from this model is that the most likely value of $\mathbf{f}(\mathbf{z})$ is the mean, $(\hat{\mathbf{f}}_z + \bar{\mathbf{f}}_z)$, with variance Λ_z .

Note that $\hat{\mathbf{f}}_z$ is simply a \mathbf{z} -dependent weighted linear combination of the measured data points, \mathbf{Y} , using weights $\Lambda_{Yz}^T \Lambda_{YY}^{-1}$.

Exponential covariance

Normally the prior mean is chosen to be zero (in the absence of any evidence the value of $\mathbf{f}(\mathbf{z})$ is as likely to be positive as negative); that is,

$$\bar{\mathbf{f}}_z = 0 \quad (13)$$

Unless otherwise stated, this is the case here. A straightforward smoothness prior covariance function for \mathbf{f}_z , which ensures that the values of the map, when the explanatory variables are nearby, should have higher correlation than when widely separated, is [1]

$$\begin{aligned} C(\mathbf{f}_{z_i}, \mathbf{f}_{z_j}) \\ = K(\mathbf{z}_i, \mathbf{z}_j) = a \exp(-(\mathbf{z}_i - \mathbf{z}_j)^T M (\mathbf{z}_i - \mathbf{z}_j) / 2) \quad (14) \\ M = \text{diag}[d_1, \dots, d_p] \end{aligned}$$

The value of d_k characterises the rate of variation of the function, \mathbf{f} , in the direction of the k^{th} -axis; the smaller the magnitude of d_k , the slower the rate of variation of \mathbf{f} with respect to z_k . The matrix, M , thus indicates the degree of nonlinearity or relative smoothness in the directions of each explanatory variable. When each measurement noise is independent of the others and has the same probability distribution, then

$$\Delta_{ij} = n \delta_{ij} \quad (15)$$

With this choice of covariance function, the hyperparameters are (a, d_k, n) . Their values are restricted to be positive.

B. Derivative data and fits

In addition to the N measurements of the value of the function, $\mathbf{f}(\mathbf{z})$, suppose N_d measurements, $\{(\mathbf{z}_i, y_{d,i})\}_{i=1}^{N_d}$, of partial derivatives, $\partial \mathbf{f}(\mathbf{z}) / \partial z_m$ are known with

$$y_{d,i} = \mathbf{f}'(\mathbf{z}_i) + n_{d,i} \quad (16)$$

where $n_{d,i}$ is white Gaussian noise. It is straightforward to extend the Gaussian process prior model to include this information.

Assume that the related stochastic process, $\mathbf{f}_z^{\delta \mathbf{e}_m}$, where $\mathbf{f}_z^{\delta \mathbf{e}_m} = (\mathbf{f}_{(\mathbf{z} + \delta \mathbf{e}_m)} - \mathbf{f}_z) / \delta$ and \mathbf{e}_m is a unit basis vector, is well-defined in the limit as $\delta \rightarrow 0$, i.e. all the necessary probability distributions for a complete description exist. Denote the derivative stochastic process, i.e. the limiting

random process, by $f_{z_m}^{e_m}$. The $E[f_{z_m}^{e_m}]$ as \mathbf{z} varies is interpreted as a fit to $\frac{\partial f}{\partial z_m}(\mathbf{z})$ when the partial derivative of $f(\mathbf{z})$ in the direction \mathbf{e}_m exists. Provided the covariance $C(f_{z_i}, f_{z_j})$ is sufficiently differentiable, it is well known [4] that $f_{z_m}^{e_m}$ is itself Gaussian and that

$$E[f_{z_m}^{e_m}] = \frac{\partial}{\partial z_m} h_f(\mathbf{z}) \quad ; \quad h_f(\mathbf{z}) = E[f_{z_m}] \quad (17)$$

where z_m denotes the m^{th} element of \mathbf{z} ; that is, the expected value of the derivative stochastic process is just the derivative of the expected value of the stochastic process. Furthermore,

$$E[f_{z_i}^{e_m} f_{z_j}^{e_n}] = \nabla_m^1 \nabla_n^2 C(f_{z_i}, f_{z_j}) \quad (18)$$

where $\nabla_m^1 Q(\mathbf{z}_0, \mathbf{z}_1)$ denotes the partial derivative of $Q(\mathbf{z}_0, \mathbf{z}_1)$ with respect to the m^{th} element of its first argument, *etc.*

The measurement noise, $n_{d,i}$, has covariance $\Delta_{d,ij}$ and is statistically independent of $f_{z_i}^{e_m}$. Furthermore, given that the measurement processes for $f(\mathbf{z})$ and $\partial f(\mathbf{z})/\partial z_m$ are separate procedures, the $n_{d,i}$ are also independent of the f_{z_i} and n_i . Hence, the covariances for the measured partial derivatives, $y_{d,i}$, are

$$C(y_{d,i}, y_{d,j}) = C(f_{z_i}^{e_m}, f_{z_j}^{e_n}) + \Delta_{d,ij}$$

$$C(y_{d,i}, f_{z_j}^{e_m}) = C(f_{z_i}^{e_m}, f_{z_j}^{e_n}) \quad ; \quad C(y_{d,i}, f_{z_j}) = C(f_{z_i}^{e_m}, f_{z_j}) \quad (19)$$

The above procedure can be repeated to construct second derivative stochastic processes. The means and covariances can be determined by recursive application of (17) and (18).

Training and prediction

When each derivative measurement noise is independent of the others and has same probability distribution, then

$$\Delta_{d,ij} = n_d \delta_{ij} \quad (20)$$

The same training procedure as in Section IIA applies. The hyperparameters (a, d_k, n, n_d), whilst constrained to be positive, are now adapted to maximise the likelihood, (6), but with

$$\mathbf{Y} = [y_1, \dots, y_N, y_{d,1}, \dots, y_{d,N_d}]^T$$

$$\mathbf{P}_{ij} = \begin{cases} C(y_i, y_j) & , \quad 1 \leq i \leq N & 1 \leq j \leq N \\ C(y_i, y_{d,j}) & , \quad 1 \leq i \leq N & N_d < j \leq N + N_d \\ C(y_{d,i}, y_j) & , \quad N_d < i \leq N + N_d & 1 \leq j \leq N \\ C(y_{d,i}, y_{d,j}) & , \quad N_d < i \leq N + N_d & N_d < j \leq N + N_d \end{cases}$$

The same prediction as in Section IIA also applies. With appropriate definition of the relevant covariance matrices in (10), similar to \mathbf{P} above, predicted values for both the function and its derivative can be made.

Exponential covariance

The covariance function, (14), is sufficiently smooth for the derivative stochastic processes to be well-defined and the

relations (16) and (17) to apply [4]. Their covariances are

$$C(f_{z_i}^{e_m}, f_{z_j}^{e_n}) = (d_m \delta_{mn} - d_m(\mathbf{z}_i - \mathbf{z}_j)_m d_n(\mathbf{z}_j - \mathbf{z}_i)_n) K(\mathbf{z}_i, \mathbf{z}_j) \quad (20)$$

and

$$C(f_{z_i}^{e_m}, f_{z_j}) = -d_m(\mathbf{z}_i - \mathbf{z}_j)_m K(\mathbf{z}_i, \mathbf{z}_j) \quad (21)$$

where $(\mathbf{z}_i - \mathbf{z}_j)_m$ is the m^{th} element of $(\mathbf{z}_i - \mathbf{z}_j)$.

III. COMBINING EQUILIBRIUM AND OFF-EQUILIBRIUM DATA

When identifying the nonlinear dynamic system, (2), several different sources of information may be available. The measured data might include the following.

Equilibrium points

An equilibrium point (\mathbf{r}_0, y_0) is a stationary solution of (2) such that

$$y_0 = F(y_0, \dots, y_0, \mathbf{r}_0, \dots, \mathbf{r}_0) \quad (22)$$

These points are typically known very accurately since they can be measured over a protracted period.

Local linear models

Relative to a particular equilibrium point, (\mathbf{r}_0, y_0) , the local linear dynamics are

$$\begin{aligned} \Delta y_i = & \Delta y_{i-1}^T \nabla_{y_{i-1}} F(y_0, \dots, y_0, \mathbf{r}_0, \dots, \mathbf{r}_0) + \\ & \dots + \Delta y_{i-n}^T \nabla_{y_{i-n}} F(y_0, \dots, y_0, \mathbf{r}_0, \dots, \mathbf{r}_0) \\ & + \Delta \mathbf{r}_i^T \nabla_{\mathbf{r}_i} F(y_0, \dots, y_0, \mathbf{r}_0, \dots, \mathbf{r}_0) + \\ & \dots + \Delta \mathbf{r}_{i-n}^T \nabla_{\mathbf{r}_{i-n}} F(y_0, \dots, y_0, \mathbf{r}_0, \dots, \mathbf{r}_0) \end{aligned} \quad (23)$$

where

$$\Delta y_j = (y_j - y_0) \quad , \quad \Delta \mathbf{r}_j = (\mathbf{r}_j - \mathbf{r}_0)$$

These linear dynamics are typically known with reasonable accuracy. With a sufficiently small perturbations, $\Delta \mathbf{r}_j$, a large amount of data can be obtained with the system operating close to a particular equilibrium point. Using one of the many well-established algorithms for linear systems, the local dynamics can then be identified accurately together with the estimation error. It should be noted that the coefficients of the local linear model are the partial derivatives of the function at the equilibrium point.

Off-equilibrium data

Away from the equilibrium points, the system cannot be held stationary. Hence, off-equilibrium measurements can only be obtained transiently. Consequently, this data is necessarily less accurate than the equilibrium points or local linear models.

In this section, the integration of the above sources of information is discussed. Recall in the Gaussian process prior models that it is assumed the explanatory variable, \mathbf{z} , is noise free. The recursive nature of the nonlinear dynamic system, (2), is incompatible with this requirement, since the explanatory variable, \mathbf{z} , would need to include delayed measurements, y_i , which has additive noise. Instead, the data integration is considered below for the nonlinear map

(3) with

$$\begin{aligned} f(\mathbf{z}) &= F(w_{i-1}, \dots, w_{i-n}, \mathbf{r}_i, \dots, \mathbf{r}_{i-n}) \\ \mathbf{z} &= [w_{i-1}, \dots, w_{i-n}, \mathbf{r}_i^T, \dots, \mathbf{r}_{i-n}^T]^T \end{aligned} \quad (24)$$

The link with the nonlinear dynamic system, (2), is further reinforced by choosing the sequence of explanatory variables such that

$$w_i = F(w_{i-1}, \dots, w_{i-n}, \mathbf{r}_i, \dots, \mathbf{r}_{i-n}) \quad (25)$$

In other words, the noise on the measured value of y is ignored when it is used as an element of the explanatory variable. Analogously to a nonlinear dynamic system, an equilibrium point for the nonlinear map (22) is defined as a value, $\mathbf{z}_0 = [w_0, \dots, w_0, \mathbf{r}_0^T, \dots, \mathbf{r}_0^T]^T$, of the explanatory variable such that $w_0 = f(\mathbf{z}_0)$. The definition of the local linear models then remains sensible.

A. Phase space analysis

The relationship, (24), defines a nonlinear map on ‘phase space’ (for the nonlinear dynamic system, (2), the space of delayed inputs and outputs). In the context of Gaussian process models, the values of the map, when the explanatory variables are nearby in the phase space, should have higher correlation than when widely separated.

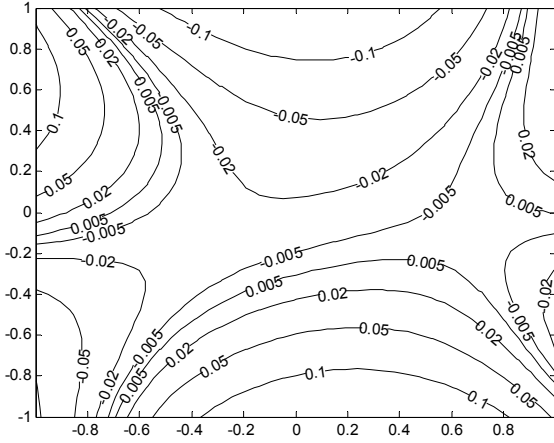


Fig.1 Constant error contours, function values only.

Example 1: $f(\mathbf{z}) = F(w_{i-1}, \mathbf{r}_{i-1}) = \sin(w_{i-1} + \mathbf{r}_{i-1})$

The training data consists of 451 off-equilibrium measurements

$$\{(\mathbf{z}_i, y_i)\}_{i=1}^{451} \quad y_i = f(\mathbf{z}_i) + n_i \quad \mathbf{z}_i = [w_{i-1}, \mathbf{r}_{i-1}]^T$$

of the function value. The sequence of explanatory variable \mathbf{r}_i is chosen to be coloured noise and the sequence of explanatory variable w_i is defined by (25). The measurement noise, n_i , for each function value is independent and has variance 0.04. On maximising the likelihood, the hyperparameters obtained (for a typical sample) with the map covariance (14) and noise covariance (15) are

$$a = 0.5307, d_1 = 0.4089, d_2 = 0.6801, n = 0.0351$$

The error between the exact and predicted values, $(f(\mathbf{z}) - \hat{f}_z)$, is depicted in Figure 1. The data lies in a horizontal band

between -0.4 and 0.4 with some clustering towards the left and right edges. It can be seen that the error increases towards the top and bottom of the plot where the data is sparse.

B. Equilibrium points and local linear models

In addition to the N measurements of the value of the function, $f(\mathbf{z})$, suppose N_0 equilibrium points, $\{(\mathbf{z}_{0,i}, y_{0,i})\}_{i=1}^{N_0}$, are known, denoting this information by S_0 . As discussed above, the equilibrium points would be known with high accuracy and they are assumed to be noise free, i.e.

$$y_{0,i} = f(\mathbf{z}_{0,i}) \quad (26)$$

It is straightforward to extend the Gaussian process prior model to include this information. Augment the off-equilibrium data with the equilibrium point data. The measurements vector, \mathbf{Y} , becomes

$$\mathbf{Y} = [y_1, \dots, y_N, y_{0,1}, \dots, y_{0,N_0}]^T \quad (27)$$

The only modification to the identification and prediction procedure required is in the noise covariance. It remains Δ_{ij} for $1 \leq i \leq N, 1 \leq j \leq N$ but, since the values at equilibrium points are noise free, it is 0, elsewhere.

Now suppose that the local linear models at the equilibrium points are also available. As discussed above, not only will estimates of the partial derivatives be available but also the covariance of errors. The only required modification to the identification and prediction procedure, see Section IIB, is to incorporate this information into the noise covariance. Let \mathbf{Y}_d be the derivative measurements with

$$\mathbf{Y}_d = [\gamma_1^T, \dots, \gamma_{N_0}^T]^T, \quad \gamma_i^T = \nabla_{\mathbf{z}} f(\mathbf{z}_{0,i(i)}) \quad (28)$$

then the derivative noise covariance, Δ_d , is block diagonal, the individual blocks being the error covariance matrices for the particular value of \mathbf{z}_0 . Unlike Section IIB there is now no hyperparameter associated with the noise covariance.

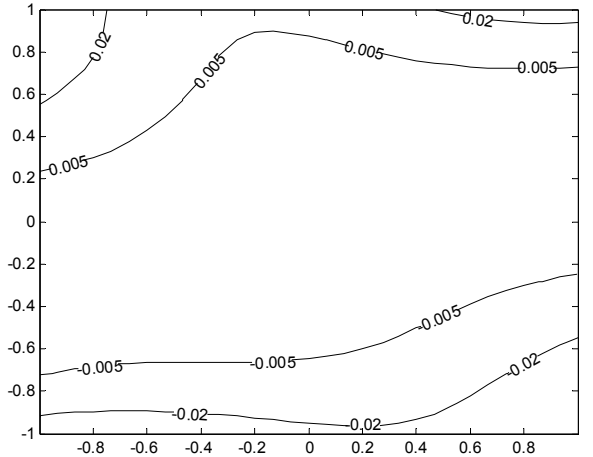


Fig.2 Constant error contours with local linear models.

Example 1(contd.): $f(\mathbf{z}) = F(\mathbf{w}_{i-1}, \mathbf{r}_{i-1}) = \sin(\mathbf{w}_{i-1} + \mathbf{r}_{i-1})$

The training data consists of the same 451 off-equilibrium measurements of the function value and 6 equilibrium points, i.e. solutions to $\mathbf{w}_0 = \sin(\mathbf{w}_0 + \mathbf{r}_0)$, with the six local linear models at the equilibrium points, i.e. the partial derivatives at the equilibrium points. The equilibrium points are listed in Table 1. The measurement noise, n_i , for each off-equilibrium function value is independent and has variance 0.04. The partial derivatives and their error covariance are identified for each equilibrium point using linear regression. These are listed in Table 2. On maximising the likelihood, the hyperparameters obtained with the map covariance (14) and noise covariance (15) are

$$a = 0.8257, d_1 = 0.2143, d_2 = 0.4215, n = 0.0351$$

The error between the exact and predicted values, $(f(\mathbf{z}) - \hat{f}_z)$, is depicted in Figure 2. In comparison to Figure 1, the region, where the confidence interval is small, has broadened considerably; in particular, it is small near the equilibrium points.

\mathbf{w}_0	\mathbf{r}_0
-0.9900	-0.4393
-0.6000	-0.0435
-0.2000	-0.0014
0.2000	0.0014
0.6000	0.0435
0.9900	0.4393

Table 1. Equilibrium values

$\mathbf{z}_0 = \begin{bmatrix} \mathbf{w}_0 \\ \mathbf{r}_0 \end{bmatrix}$	$\begin{bmatrix} \partial F(\mathbf{z}_0) / \partial \mathbf{w} \\ \partial F(\mathbf{z}_0) / \partial \mathbf{r} \end{bmatrix}$	Error covariance for partial derivatives
$\begin{bmatrix} -0.9900 \\ -0.4393 \end{bmatrix}$	$\begin{bmatrix} 0.1419 \\ 0.1398 \end{bmatrix}$	$10^{-4} \times \begin{bmatrix} 0.2350 & -0.0136 \\ -0.0136 & 0.0271 \end{bmatrix}$
$\begin{bmatrix} -0.6000 \\ -0.0435 \end{bmatrix}$	$\begin{bmatrix} 0.8001 \\ 0.7994 \end{bmatrix}$	$10^{-4} \times \begin{bmatrix} 0.6472 & 0.6455 \\ 0.6455 & 0.6501 \end{bmatrix}$
$\begin{bmatrix} -0.2000 \\ -0.0014 \end{bmatrix}$	$\begin{bmatrix} 0.9806 \\ 0.9773 \end{bmatrix}$	$10^{-4} \times \begin{bmatrix} 0.9718 & 0.9684 \\ 0.9684 & 0.9674 \end{bmatrix}$
$\begin{bmatrix} 0.2000 \\ 0.0014 \end{bmatrix}$	$\begin{bmatrix} 0.9803 \\ 0.9772 \end{bmatrix}$	$10^{-4} \times \begin{bmatrix} 0.9712 & 0.9680 \\ 0.9680 & 0.9680 \end{bmatrix}$
$\begin{bmatrix} 0.6000 \\ 0.0435 \end{bmatrix}$	$\begin{bmatrix} 0.8003 \\ 0.7995 \end{bmatrix}$	$10^{-4} \times \begin{bmatrix} 0.6476 & 0.6457 \\ 0.6457 & 0.6486 \end{bmatrix}$
$\begin{bmatrix} 0.9900 \\ 0.4393 \end{bmatrix}$	$\begin{bmatrix} 0.1422 \\ 0.1405 \end{bmatrix}$	$10^{-4} \times \begin{bmatrix} 0.2444 & -0.0087 \\ -0.0087 & 0.0260 \end{bmatrix}$

Table 2. Derivative values

Remark 1. Combining equilibrium points and local linear models with off-equilibrium information is straightforward. Using the equilibrium points greatly reduces the errors in their vicinity. Using the local linear models considerably increases the region of reduced errors. This issue is discussed in [5] where it is pointed out that large quantities of local data, if used to identify the local partial derivatives, can be replaced by a single accurate datum and the derivatives. The datum point (here an equilibrium point) and the derivatives (here the local linear model) essentially

summarises the information contained in the original data.

Remark 2. Combining the equilibrium information with the off-equilibrium information assists the construction of Gaussian process prior models in two ways. Firstly, an appropriate order of the explanatory variable is obtained by choosing it to be the order of the local linear models. Secondly, although data sets containing several thousand points [2] have been used to train the hyperparameters, the need to invert the matrix, \mathbf{P} , mitigates against large data sets. However, the summarising property of the local linear models, see Remark 1, can be exploited to reduce the size of the data set.

IV. IDENTIFICATION OF NONLINEAR DYNAMIC SYSTEMS

In this section, the identification from input/output data of the nonlinear dynamic system, (2), is discussed. The approach is similar to that of Section III except that the explanatory variable is chosen to be

$$\mathbf{z} = [y_{i-1}, \dots, y_{i-n}, \mathbf{r}_i^T, \dots, \mathbf{r}_{i-n}^T]^T \quad (29)$$

The noise on the measured value of y is no longer ignored. Consequently, the accuracy of the fit cannot be expected to be as good as in Section III.

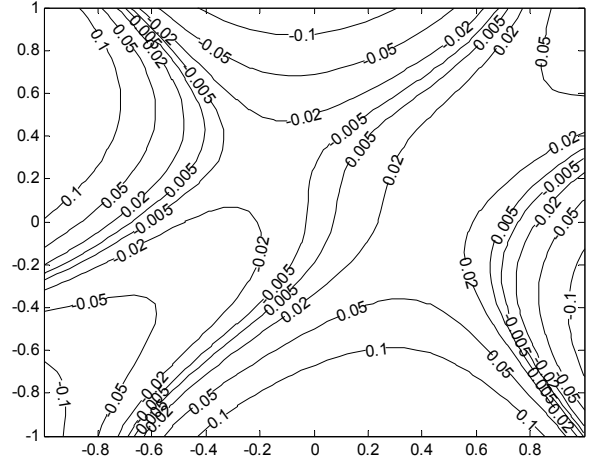


Fig.3 Constant error contours with noisy explanatory variables.

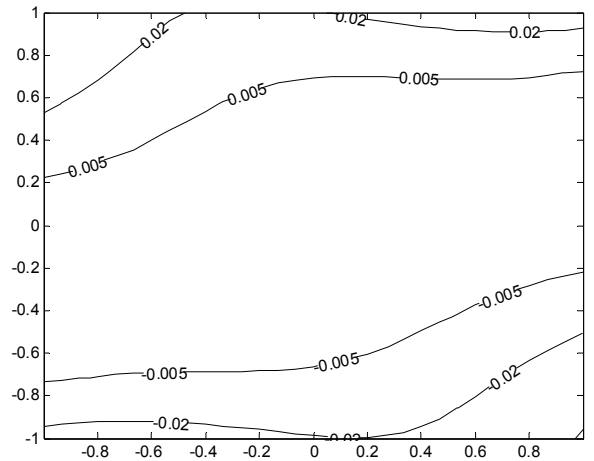


Fig.4 Constant error contours with noisy data and linear models.

Example 2: $f(\mathbf{z}) = F(w_{i-1}, r_{i-1}) = \sin(w_{i-1} + r_{i-1})$

The sequences of input r_i and measured output y_i remain the same as in Example 1 but the training data now consists of the 451 off-equilibrium measurements

$$\{(\mathbf{z}_i, y_i)\}_{i=1}^{451} \quad y_i = f(\mathbf{z}_i) + n_i \quad \mathbf{z}_i = [y_{i-1}, r_{i-1}]^T$$

of the function value. On maximising the likelihood, the hyperparameters obtained with the map covariance (14) and noise covariance (15) are

$$a = 0.4740, d_1 = 0.5911, d_2 = 0.7269, n = 0.0474$$

The error between the exact and predicted values, $(f(\mathbf{z}) - \hat{f}_z)$, is depicted in Figure 3. Comparing to Figure 1, the error is increased as expected. The training data is supplemented by the function values at the 6 equilibrium points and their local linear models. On maximising the likelihood, the hyperparameters obtained with the map covariance (14) and noise covariance (15) are

$$a = 0.8037, d_1 = 0.2225, d_2 = 0.4100, n = 0.0491$$

The error between the exact and predicted values, $(f(\mathbf{z}) - \hat{f}_z)$, is depicted in Figure 4. It is comparable to that obtained previously when noise-free values of the measured value are used in the explanatory variable, see Figure 2.

Remark 3. The accuracy of the fit obtained for the mapping is degraded by the noise on the measured values when they are incorporated into the explanatory variable. However, this degradation is much reduced when combining the equilibrium information with the off-equilibrium information. In this manner, a nonlinear dynamic system can be identified using Gaussian process prior models with some reasonable accuracy.

The response of the system, $\{\hat{y}_i\}$, to any input, $\{r_i\}$, can be predicted by recursive application of (12); that is, from

$$\hat{y}_i = \Lambda_{YZ}^T \Lambda_{YY}^{-1} \mathbf{Y} \quad \mathbf{z} = [\hat{y}_{i-1}, \dots, \hat{y}_{i-n}, \mathbf{r}_i^T, \dots, \mathbf{r}_{i-n}^T]^T \quad (30)$$

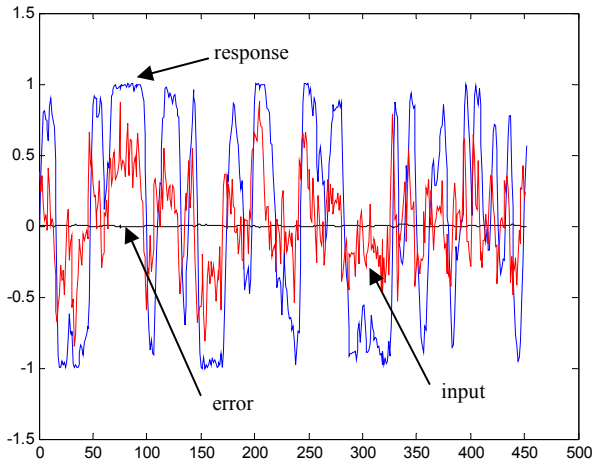


Fig.5 Input, response and error for training data.

Example 2(contd.): $f(\mathbf{z}) = F(w_{i-1}, r_{i-1}) = \sin(w_{i-1} + r_{i-1})$

Using the model of the nonlinear system identified above, the response to the training input sequence is determined.

The training input, its response and error are shown in Figure 5. The response and error for a further input is shown in Figure 6. In both cases the error is small.

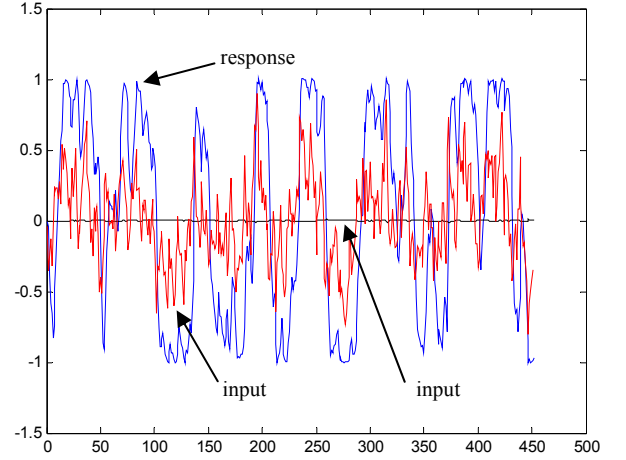


Fig.6 Input, response and error for validation data.

V. CONCLUSIONS

An identification methodology for nonlinear dynamic systems based on Gaussian process prior models is described that explicitly utilises both equilibrium knowledge and off-equilibrium knowledge. The equilibrium knowledge consists of known equilibrium points and their local linear models. The use of both improves the accuracy of the identified models is substantially, thereby, mitigating the effects of noise included in the explanatory variable through the measured values. Furthermore, the construction of the Gaussian process prior model is assisted. The order of the explanatory variable is obtained from the local linear models and the computational burden is eased by data compression.

VI. ACKNOWLEDGEMENTS

This work was supported by Science Foundation Ireland, grant 00/PI.1/C067, and by EPSRC, grant GR/M76379.

VII. REFERENCES

- [1] C. K.I Williams, "Prediction with Gaussian Processes: From linear regression to linear prediction and beyond", *Learning and Inference in Graphical Models*, M. I. Jordan, 1998, Kluwer.
- [2] C. E. Rasmussen, *Evaluation of Gaussian processes and other methods for nonlinear regression*, Ph.D. thesis, 1996, University of Toronto.
- [3] P. J. Green, B. W. Silverman, *Nonparametric Regression and Generalised Linear Models*, 1994, Chapman & Hall, London.
- [4] A. O'Hagan, "On curve fitting and optimal design for regression", 1978, *J. Royal Stat Soc. B*, **40**, 1-42.
- [5] E. Solak, R. Murray-Smith, W. E. Leithead, D. J. Leith, C. E. Rasmussen, "Derivative observations in Gaussian process models of dynamic systems", 2003, *Advances in Neural Information Processing Systems*, **15**, 1033-1040.