A Bayesian Perspective on Stochastic Neuro Control

Randa Herzallah

FET

Al-Balqa' Applied University, Jordan

Email:randa_herzal@hotmail.com

David Lowe

NCRG

Aston University, UK

Email:d.lowe@aston.ac.uk

Abstract

Control design for stochastic uncertain nonlinear systems is traditionally based on minimizing the expected value of a suitably chosen loss function. Moreover, most control methods usually assume the certainty equivalence principle to simplify the problem and make it computationally tractable. We offer an improved probabilistic framework which is not constrained by these previous assumptions, and provides a more natural framework for incorporating and dealing with uncertainty. The focus of this paper is on developing this framework to obtain an optimal control law strategy using a fully probabilistic approach for information extraction from process data which does not require detailed knowledge of system dynamics. Moreover the proposed control method framework allows handling the problem of input–dependent noise. A basic paradigm is proposed and the resulting algorithm is discussed. The proposed probabilistic control method is demonstrated theoretically on two classes of nonlinear discrete time systems: the affine and the general class. A nonlinear simulation example is also provided to validate theoretical development.

I. Introduction

In active control engineering the main objective is to ensure that systems of interest perform according to predefined specifications despite changing conditions, dynamics and noise. However, the growing complexity of control systems, accompanied by high levels of inherent uncertainty in modeling and estimation and intrinsic nonlinear dynamics involving unknown functionals, make achieving the above objective impossible except under idealized conditions. For example, several control techniques, including adaptive control, optimal control and robust control, are based on designing fixed controllers that can achieve good control results under fixed assumptions. For situations where the plant parameters are uncertain or time varying the techniques of adaptive control [2], [31] are usually performed on–line to maintain tracking performance. However, these adaptive control methods fail when the situation becomes more complex such as if the plant to be controlled is characterized by a multi–valued function or even if it exhibits a number of distinct modes of behavior during its operation. The solution to this problem has been discussed in [23]–[25] among others. Another restriction is that uncertainty modeling is usually limited to linear plant models, since propagating uncertainty through nonlinear models destroys preferred Gaussian distributions of error signals. The Pontryagin Minimum Principle, or the Hamilton-Jacobi-Bellman partial differential equation are the known methods for optimal control of continuous noise free known model processes. However, the inclusion of noise renders the approaches unmanageable except under some very special circumstances. This problem has been recently considered by Kappen [21] who derived an elegant reformulation of the HJB equations in a form similar to the quantum mechanical Schrodinger equation, thus opening up the toolbox of computational approaches from that domain. However the reformulation assumed quadratic cost functions and linear dynamics in the control parameter. Unfortunately, this formulation does not map over to the case we are considering in this paper, where the systems are not continuous, are nonlinear and generally unknown, and are not generally linear in the control parameter, nor have quadratic cost functions.

Uncertain systems however have been traditionally controlled using the theory of stochastic adaptive control. Fel'dbaum [10]–[12] discovered that control laws derived on the basis of stochastic adaptive theory generate a control signal that possess two important properties: direction and probing. Direction means that it directs the system output to follow some desired value even by taking into consideration uncertainty of the parameter estimates. While probing means that it elicits further information from the system so as to reduce further parameter uncertainty. Fel'dbaum used the term dual control to characterize control laws that possess these two properties. However dual control has not been widely used in practice because of the computational complexity associated with solving the equations required to find an optimal solution. For this reason few stochastic adaptive control schemes are based on finding a suboptimal solution that retains the desirable properties of ideal dual control while permitting implementation [4], [22], [27], [28], [33].

In the recent past, consideration of complexity in control deriving purely from the interplay of degrees of freedom through nonlinearity has previously led to the development of techniques within the framework of intelligent control theory [29], [32], [35], [39] that have enhanced the degree of adaptation, learning and autonomy. In the neuro-control field, unknown functions are usually handled by machine learning methods where extra issues of nonlinear model identification and stability under iteration have arisen [6], [7], [20], [37]. But they still tend to cling to basic assumptions of approximate knowledge of the plant and deterministic function mappings describing forward and inverse control. Some of the adaptive control schemes that attempt to deal with stochastic uncertain nonlinear control systems typically adopt a Heuristic Certainty Equivalence (HCE) [2], [14] control principle which often leads to an inadequate transient response because of the initial uncertainty involved in the unknown parameters. Most other neural network approaches try and reduce the prior uncertainty of the stochastic uncertain nonlinear control systems by performing intensive batch off–line training [5], [31], [36]. Fabri and Kadirkamanathan [9] explicitly states that the off–line training defeats the main objective of adaptive control because the off–line training phase reduces uncertainty through time existing prior to application of the control. Consequently they proposed a suboptimal dual adaptive control scheme which was proved to give superior performance to that of certainty equivalence based control methods. Other approaches for stochastic uncertain nonlinear control systems have also been developed [8], [17]–[19].

However, complexity in the control arena is also a consequence of uncertainty: uncertainty due to ignorance of underlying latent structures, incomplete specifications of the processes, and noise and randomness in data and approximate model parameters. To date these issues have been largely ignored in the active control field, with only a few notable exceptions. Despite major efforts which have been put forward in the recently developed control methods for stochastic uncertain nonlinear control systems, there is still a lack of a complete methodology guaranteeing satisfactory solutions.

In this paper we propose a novel approach based on a fully probabilistic control framework. In the proposed control method, historical process data, fully describing the forward and the inverse models of the system are processed by a semi-Bayesian algorithm. An appropriate likelihood function is then maximized instead of the usual minimization of the expected value of a data dependent loss function. The conditional distributions required by the proposed semi-Bayesian control algorithm are estimated using recent development in machine learning. Although all conditional distribution functions are assumed to be Gaussian in this paper, expected means and variances of these distributions are estimated using nonlinear neural network models. Detailed knowledge of system dynamics is not required in the underlying probabilistic extraction of information from process data, which supports the approach's generality.

The proposed probabilistic design methodology gives interesting insight into the probabilistic decision making problem. For instance, it allows us to take model uncertainty estimation into consideration when deriving the optimal control law. Moreover the explicit dependency of model uncertainty on the input data is considered in the proposed probabilistic control algorithm.

This new framework provides an alternative to standard adaptive control theory based on probabilistic control methods rather than deterministic methods. Designing a probabilistic control method allows, as will be demonstrated in the paper, taking model uncertainty into consideration when designing the near to optimal control law. Taking knowledge of uncertainty into consideration when deriving the near to optimal control gives superior control results [8], [17]–[19].

II. Preliminaries

This preparatory section recalls basic elements of modeling conditional distributions of system outputs and fully probabilistic control.

A. Probabilistic Control

In this paper we are concerned with the trajectory control problem. Here a stochastic system to be controlled acted upon by an input command u_k at time k is considered (for simplicity a discrete time system will be considered, although an equivalent continuous time formulation could be considered). Since we consider a trajectory control problem, the objective is to minimize the trajectory error, $e_t(k+\tau) = y(k+\tau) - y_d(k+\tau)$, on average such that the mean value of the system output becomes equal to the desired value, where τ is the relative degree of the plant. The resulting trajectory error, which describes the causal relationship between the input command and the system output is described by a conditional probability density function (pdf),

$$p(e_t \mid \mathbf{x}, \mathbf{u}, y_d(k+\tau)) = f(e_t(k+\tau) = e_t \mid \mathbf{x}(k) = \mathbf{x}, \mathbf{u}(k) = \mathbf{u}, y_d(k+\tau) = y_d), \qquad (1)$$

where $e_t(k+\tau)$ is the trajectory error, $\mathbf{u}(k)$ is the control input, $\mathbf{x}(k) = [y(k), \dots, y(k-q+1), \mathbf{u}(k-1), \dots, \mathbf{u}(k-p+1)]^T$ is the system state vector. In this paper all probability density functions are assumed to be unknown and will be estimated as described in Section II-B.

The randomized controller to be designed is described by the pdf

$$p(\mathbf{u} \mid \mathbf{x}, y_d) = c(\mathbf{u}(k) = \mathbf{u} \mid \mathbf{x}(k) = \mathbf{x}, y_d(k+\tau) = y_d), \qquad (2)$$

where $y_d(k+\tau)$ is the desired output at time $(k+\tau)$.

The joint distribution of $(e_t, \mathbf{u})$ will be denoted by $p(e_t, \mathbf{u})$ and the distribution of $\mathbf{u}$ conditional on e_t by $p(\mathbf{u} \mid e_t)$, with the evaluation

$$p(\mathbf{u} \mid e_t) = \frac{p(e_t, \mathbf{u})}{p(e_t)}. \qquad (3)$$

The value of the variable $\mathbf{u}$ need not be specified as a random variable jointly with e_t, but could simply be a quantity whose value parameterizes and consequently affects the distribution of e_t.

Such a variable is called a parameterizing variable, and the distribution of e_t for prescribed $\mathbf{u}$ is denoted as $p(e_t \mid ; \mathbf{u})$. On the other hand the stochastic description could be completed so that $(e_t, \mathbf{u})$ can be regarded jointly as random variables. In such a case it may or may not be true that $p(e_t \mid \mathbf{u}) = p(e_t \mid ; \mathbf{u})$ [40]. To simplify notations from now on $p(e_t \mid \mathbf{u})$ will be used throughout the paper.

The problem of designing a randomized controller is addressed from a stochastic adaptive perspective. This means finding an admissible control sequence that minimizes a pre-specified performance index that usually takes the form of the expected value of some cost function [2], [40]. Admissibility implies that $\mathbf{u}(k)$ must be restricted to depend upon the information state I^k, which consists of all outputs measured up to the present time, Y^k, and all previous inputs, U^{k-1}. Adherence to the admissibility condition ensures the causality of the control law derived from the optimization procedure. Following the terminology from probability theory, when $\mathbf{u}(k)$ is a function of (k, I^k) we say that $\mathbf{u}(k)$ is I^k-measurable [9].

Stochastic adaptive systems generate a control signal that possess two important properties: direction and probing. Fel'dbaum [10]–[12] used the term dual control to characterize control laws that possess these two properties. Although the performance of dual controllers is generally superior to that of non-dual systems, they have not been widely used in practice because of the computational complexity issues associated with solving the well known Bellman equations from dynamic programming [40]. For this reason most practical stochastic adaptive control algorithms seek a suboptimal solution that to a certain extent retains the desirable properties of ideal dual control whilst permitting implementations [4], [27], [28], [33], [38].

In this paper an alternative design principle to the stochastic adaptive control problem is proposed. The principle may be essentially stated as follows:

A controller of a stochastic system may affect its output distribution. The optimal controller should make the mean value of the output distribution of the estimate of the controlled system as close as possible to a desired value and at the same time minimize the variance of the output distribution. In short, the aim is to manipulate estimators of conditional pdfs.

Having accepted this rephrased control aim, it is desirable to construct the posterior distribution of the control signal which can be estimated using Bayes' rule as follows:

$$p(\mathbf{u} \mid e_t) = \frac{p(e_t \mid \mathbf{u}) p(\mathbf{u})}{p(e_t)}. \tag{4}$$

The optimization method of a controller given by Equation (4) will be discussed later in Section IV. Next we discuss the estimation process for the conditional distribution of the forward model required by the proposed randomized controller.

B. Conditional Distribution Estimation of Forward Model

The problem of forward model identification is well known and we adopt the same strategy of generic neural network modeling but augment it with the capability to quantify uncertainty. The structure of the neural network model is firstly chosen, and is assumed to be identical to that of the plant if known a priori, otherwise it is determined from the observed data as in a standard identification process. In this section the problem of estimating the conditional probability distribution of forward models will be discussed, for the following general class of nonlinear discrete time systems:

$$y(k+\tau) = f(y(k),\dots,y(k-q+1),\mathbf{u}(k),\mathbf{u}(k-1),\dots,\mathbf{u}(k-p+1)) + \eta(k+\tau). \quad (5)$$

The same discussion can be applied to other classes of systems. For the general class of discrete time nonlinear systems given in Equation (5), the neural network model is taken to have the general input-output form given by

$$\hat{y}(k+\tau) = N_f(y(k),\dots,y(k-q+1),\mathbf{u}(k),\mathbf{u}(k-1),\dots,\mathbf{u}(k-p+1)). \quad (6)$$

The parameters of the neural network model are then adjusted using an appropriate gradient–based method to optimize a performance function based on the error between the plant and the neural model output. The parameters of the forward model can be optimized either off–line or on–line. For the proposed probabilistic control scheme in this paper on–line identification is considered. Once the forward model of the plant is identified, it is possible to build a stochastic model by simply using

$$y(k+\tau) = \hat{y}(k+\tau) + \eta(k+\tau),$$

where $\eta(k+\eta)$ represents the residual error of the system output which is assumed to be Gaussian random noise of zero mean and ρ^2 variance. This term in other words, represents uncertainty of the forward model. Note, it would be possible to consider more general estimators of the residual error, but the Gaussian assumption is adequate for the moment.

The above stochastic model has been proposed in [19] to estimate model uncertainty of both the forward and inverse models of the plant. It is based on theorem 4.2.1 in [13] which states that the minimum mean squared error (MMSE) estimate of a random vector z given another random vector x is simply the conditional expectation of z given x, $\hat{z} = E(z \mid x)$.

Since the neural network models are optimized based on minimum mean squared error, this theorem is directly applicable to the output of the neural network model [19]. Let $\boldsymbol{\nu}(k) = [y(k), \ldots, y(k-q+1), \mathbf{u}(k), \mathbf{u}(k-1), \ldots, \mathbf{u}(k-p+1)]$ be the input vector to the neural network model, the optimal output of the forward model $\hat{y}(k+\tau)$ is the conditional expectation of $y(k+\tau)$ given $\boldsymbol{\nu}(k)$.

Based on this, two alternative methods have been proposed to provide an estimate for the variances of the system outputs. In [3] it has been shown that if the network mapping function is given by the conditional average of the system outputs, then the average variance of the system outputs is simply the residual error value of the sum–of–squares error function at its minimum, $\rho^2 = (1/ND) \sum_{k=1}^{D} \sum_{i=1}^{N} \parallel y(k+\tau) - \hat{y}(k+\tau) \parallel^2$. Here D is the dimensionality of the system outputs, and N is the number of samples. In the second method [15], [16] the variances of the system outputs are assumed to be input dependent. Based on the same assumption, that the optimum network outputs approximate the conditional mean of the system output, the input dependent variances of the system outputs are shown to be equal to $\rho^2 = \parallel y(k+\tau) - \hat{y}(k+\tau) \parallel^2$. To provide a prediction value for these local variances, another neural network which takes the same inputs as the one that predicts the conditional average of the system output can be trained so that the outputs of this second network approximates the conditional average of the local estimated variances, $\hat{\rho}^2 = E(\rho^2 \mid \boldsymbol{\nu}(k))$. To emphasize, this provides a pragmatic method for providing an estimate of the mean and the variance of the nonlinear forward model. It is equivalent to assuming $p(y(k+\tau) \mid \boldsymbol{\nu}(k)) \approx N[\hat{y}(k+\tau), \hat{\rho}^2]$.

For the probabilistic control method proposed in this paper, the variance of the residual error of the forward model of the system output is always assumed to be input–dependent. Estimating the variance of the forward model as a global variance has always been used in the control literature which consequently suffers from ignoring the term of the error function corresponding to that variance. In this paper we show that the standard method does not lead to an optimal control law. This is because the variance of the residual error of the forward model is input–dependent and should be included in the optimization method, as will be demonstrated shortly. Assuming an input–dependent variance allows the possibility of deriving a superior control law which minimizes the variance of the residual error of the forward model in addition to the tracking error. By doing so the optimization process can be shown to have the duality property. This means that, in addition to minimizing the tracking error, model uncertainty is also minimized at the same time.

III. Bayes' rule inference

In this section the problem of deriving the optimal control law in a neural network context on the basis of a set of training data will be addressed. The problem can be seen by a comparison with the indirect neural adaptive control problem as follows. In conventional indirect neural adaptive control, the conditional expectation of the control signals can be estimated directly by minimizing the mean square error between the system output and a predefined desired output, $e(k+\tau) = \| \mathbf{y}(k+\tau) - \mathbf{y_d}(k+\tau) \|^2$ [1], [17], [34], [39].

In this section we develop a better approach by estimating the conditional distribution of the control signals $p(\mathbf{u}(k))$, conditioned on the input parameters, rather than estimating the conditional expectation of control signals as in the standard indirect adaptive control. This new approach is based on a Bayesian method. The superiority of the proposed approach over the more common indirect adaptive control methods comes from the fact that standard indirect adaptive control simply derives the control law based on the assumption of the certainty equivalence principle. In the new proposed control method, certainty equivalence is not going to be assumed and the estimation problem will be handled completely probabilistically. This allows taking the uncertainty of the estimates into consideration in the control algorithm and consequently avoids performing intensive off–line training of the forward and inverse controller which is usually performed to reduce the prior uncertainty of the unknown parameters.

The discussion in the next sections will for simplicity of notation be limited to single-output systems, although the extension to multi-output systems is straightforward.

A. Distribution of Control Signals

The objective in the new developed Bayesian inference framework is to find the maximum aposteriori estimate of the control signals given the input values, which will be defined shortly. Therefore, a probability distribution over the control signal values is needed. Before observing the effect of the control signals on the system outputs, we assume a prior distribution denoted by, $p(\mathbf{u}(k) \mid \mathbf{y}_d(k+\tau), \mathbf{x}(k))$. Here $\mathbf{y}_d(k+\tau) = [\mathbf{y}_d^i(k+\tau)]$ and $\mathbf{x}(k) = [\mathbf{x}^i(k)]$ denote the set of training input vectors of the controller, $i = 1, \ldots, N$ is the training sample index, and N is the number of the training samples. Note here that the number of training samples, $N = 1$ if an on–line training method is assumed. Otherwise the number of training samples is specified from the available training data. Let the trajectory error resulting from the control signals in the training stage be denoted by $\mathcal{E}_t(k+\tau) = [\mathcal{E}_t^i(k+\tau)]$. Once the trajectory errors $\mathcal{E}_t(k+\tau)$ are observed, an expression for the posterior probability distribution of the control signals, which we denote by $p(\mathbf{u}(k) \mid \mathcal{E}_t(k+\tau), \mathbf{y}_d(k+\tau), \mathbf{x}(k))$ can be written using Bayes theorem:

$$p(\mathbf{u}(k) \mid \mathcal{E}_t(k+\tau), \mathbf{y}_d(k+\tau), \mathbf{x}(k)) = \frac{p(\mathcal{E}_t(k+\tau) \mid \mathbf{u}(k), \mathbf{x}(k), \mathbf{y}_d(k+\tau))p(\mathbf{u}(k) \mid \mathbf{y}_d(k+\tau), \mathbf{x}(k))}{p(\mathcal{E}_t(k+\tau) \mid \mathbf{x}(k), \mathbf{y}_d(k+\tau))}. \tag{7}$$

The denominator is a normalization factor which can be written

$$p(\mathcal{E}_t(k+\tau) \mid \mathbf{x}(k), \mathbf{y}_d(k+\tau)) = \int p(\mathcal{E}_t(k+\tau) \mid \mathbf{u}(k), \mathbf{x}(k), \mathbf{y}_d(k+\tau))p(\mathbf{u}(k) \mid \mathbf{y}_d(k+\tau), \mathbf{x}(k))d\mathbf{u}(k). \tag{8}$$

The problem of deriving the control law under this inference framework can be interpreted as follows. The problem starts with some prior distribution over the control signal given by $p(\mathbf{u}(k) \mid \mathbf{y}_d(k+\tau), \mathbf{x}(k))$. Once the effect of the control signal on the system output has been observed, this prior distribution can be converted to a posterior distribution using Bayes' theorem as indicated in Equation (7). The evaluation of the posterior will require expressions for the prior distribution $p(\mathbf{u}(k) \mid \mathbf{y}_d(k+\tau), \mathbf{x}(k))$ and for the likelihood function $p(\mathcal{E}_t(k+\tau) \mid \mathbf{u}(k), \mathbf{x}(k), \mathbf{y}_d(k+\tau))$.

B. Prior Distribution of Control Signals

To convert the general inference framework for this probabilistic control design into a tractable computation, we assume a Gaussian function for the prior distribution of control signals which can be written as an exponential of the form

$$p(\mathbf{u}(k) \mid \mathbf{y}_d(k+\tau), \mathbf{x}(k)) = \frac{1}{Z_u(\alpha)} \exp(-\alpha E_u), \tag{9}$$

where $Z_u(\alpha)$ is a normalization factor given by

$$Z_u(\alpha) = \int \exp(-\alpha E_u) d\mathbf{u}, \tag{10}$$

which ensures that $\int p(\mathbf{u})d\mathbf{u} = 1$, and E_u is a cost function chosen to reflect some preference in the problem. Here we take E_u to be quadratic, hence the Gaussian assumption. The motivation for this is as follows:

The discussion in standard control usually prefers small and smooth changes in values of the control signal, which consequently suggest the following simple form for E_u as a regularizer

$$E_u = \frac{1}{2} \parallel \mathbf{u}(k) \parallel^2 = \frac{1}{2} \sum_{n=1}^{N} \mathbf{u}_n^T(k)\mathbf{u}, \tag{11}$$

where U is the dimensionality of the control signals. This corresponds to the penalty term for $\mathbf{u}(k)$ which is usually added in control problems to penalize large control signals, reflecting that in practice the control amplitude needs to be constrained. However, our framework does not insist on smoothness of the control signal and other regularizers can be used. The prior distribution can then be rewritten in the following form

$$p(\mathbf{u}(k) \mid \mathbf{y}_d(k+\tau), \mathbf{x}(k)) = \frac{1}{Z_u(\alpha)} \exp(-\frac{\alpha}{2} \parallel \mathbf{u}(k) \parallel^2). \tag{12}$$

Thus when $\parallel \mathbf{u}(k) \parallel$ is large, and $p(\mathbf{u}(k) \mid \mathbf{y}_d(k+\tau), \mathbf{x}(k))$ is small, then this choice of distribution says that the control signal values are expected to be small rather than large.

The hyperparameter α will be treated as part of the learning process as will be discussed in Section VII. Since the prior distribution is Gaussian, the evaluation of the normalization coefficient $Z_u(\alpha)$ using Equation (10) is straight forward, and gives

$$Z_u(\alpha) = \left(\frac{2\pi}{\alpha}\right)^{\frac{NU}{2}}. \tag{13}$$

C. *Distribution of Trajectory Error*

In general the likelihood function in Bayes' theorem, Equation (7), can be written in the form

$$p(\mathcal{E}_t(k+\tau) \mid \mathbf{u}(k), \mathbf{x}(k), y_d(k+\tau)) = \frac{1}{Z_{\mathcal{E}_t}(\beta)} \exp(-\beta E_{\mathcal{E}_t}), \tag{14}$$

where $E_{\mathcal{E}_t}$ is an error function which will be defined shortly, and β is another hyperparameter which represents the noise variance of the trajectory error. The function $Z_{\mathcal{E}_t}(\beta)$ is a normalization factor given by

$$Z_{\mathcal{E}_t}(\beta) = \int \exp(-\beta E_{\mathcal{E}_t}) d\mathcal{E}_t(k+\tau), \tag{15}$$

where $\int d\mathcal{E}_t(k+\tau) = \int de_t^1(k+\tau) \dots de_t^N(k+\tau)$ represents an integration over the trajectory error. We remind the readers that the system outputs are assumed to be generated from a smooth function estimated by a neural network model with additive zero mean Gaussian noise. Again the assumption of Gaussian distribution on the errors of the forward model of the system output does not mean a linear Gaussian model as in the standard control theory. Here a nonlinear neural network model is used to estimate the system output. This nonlinear model is assumed to approximate the output of the system to a sufficient degree of accuracy, which leads to small error values on the system outputs and consequently the residual error of the system output is Gaussian. The Gaussian assumption on the residual error of the system output can be relaxed in a straightforward way by assuming a mixture of Gaussians on the distribution of the forward model, at the expense of a more complex formalism.

The effect of the control signals on the system output is given by the probability of observing a trajectory error value $e_t(k+\tau)$ for a given input as follows

$$p(e_t(k+\tau) \mid \mathbf{u}(k), \mathbf{x}(k), \mathbf{y}_d(k+\tau)) \propto \exp\left(-\frac{\beta}{2}\{y(k+\tau) - y_d(k+\tau)\}^2\right). \tag{16}$$

Given that the data points are drawn independently from this distribution, we have

$$\begin{aligned} p(\mathcal{E}_t(k+\tau) \mid \mathbf{u}(k), \mathbf{x}(k), \mathbf{y}_d(k+\tau)) &= \prod_{n=1}^{N} p(e_t^n(k+\tau) \mid \mathbf{u}^n(k), \mathbf{x}^n(k), \mathbf{y}_d^n(k+\tau)) \\ &= \frac{1}{Z_{\mathcal{E}_t}(\beta)} \exp\left(-\frac{\beta}{2}\sum_{n=1}^{N}\{y^n(k+\tau) - y_d^n(k+\tau)\}^2\right). \end{aligned} \tag{17}$$

The expression in Equation (15) for the normalization factor $Z_{\mathcal{E}_t}(\beta)$ is then the product of N independent Gaussian integrals which can be easily evaluated to give

$$Z_{\mathcal{E}_t}(\beta) = \left(\frac{2\pi}{\beta}\right)^{\frac{N}{2}} \tag{18}$$

The problem of determining the hyperparameter β as part of the learning process will be discussed later in Section VII.

D. Posterior Distribution of Control Signals

Once a prior distribution and an expression for the likelihood function have been chosen, Bayes' theorem in the form given in Equation (7) can be used to find the posterior distribution of the control signals. Using the prior expression given in Equation (9), and the likelihood expression given in Equation (14), the posterior distribution of control signals can be obtained in the form

$$p(\mathbf{u}(k) \mid \mathcal{E}_t(k+d), \mathbf{y}_d(k+d), \mathbf{x}(k)) = \frac{1}{Z_S}\exp(-\beta E_{\mathcal{E}_t} - \alpha E_u) = \frac{1}{Z_S}\exp(-S(\mathbf{u}(k))), \quad (19)$$

where

$$S(\mathbf{u}(k)) = \beta E_{\mathcal{E}_t} + \alpha E_u, \quad (20)$$

and

$$Z_S(\alpha, \beta) = \int \exp(-\beta E_{\mathcal{E}_t} - \alpha E_u) d\mathbf{u}(k). \quad (21)$$

The objective next is to discuss the problem of deriving the optimal control law, the Gaussian approximation to the posterior distribution, the distribution of the system output, in addition to the estimation of the hyperparameters.

IV. Derivation of the Optimal Control Law

We firstly discuss the derivation of the optimal control law for the simple class of affine nonlinear discrete time systems. The more general nonlinear class of models classified in [30] as the most complicated form of systems model will be discussed later.

Since we are considering a stochastic system, the optimal control law corresponds to the maximum of the posterior distribution which can be extracted by minimizing the expected value of the negative logarithm of Equation (19) with respect to the control input. Since the normalization factor Z_S in Equation (21) is evaluated by integrating over the control signal, we see that the minimization process is equivalent to minimizing $< S(\mathbf{u}) >$ given by Equation (20). For the prior distribution given in Equation (12) and noise model given by Equation (17) this can be written in the form

$$
\begin{aligned}
< S(\mathbf{u}(k)) >= & \left\langle \frac{\beta}{2}\sum_{n=1}^{N}\{y^n(k+\tau)-y_d^n(k+\tau)\}^2+\frac{\alpha}{2}\sum_{n=1}^{N}\mathbf{u}_n^T(k)\mathbf{u}_n(k)\right\rangle \\
&= \left\langle \frac{\beta}{2}\sum_{n=1}^{N}\{\hat{y}^n(k+\tau)+\eta^n(k+\tau)-y_d^n(k+\tau)\}^2+\frac{\alpha}{2}\sum_{n=1}^{N}\mathbf{u}_n^T(k)\mathbf{u}_n(k)\right\rangle \\
&= \frac{\beta}{2}\sum_{n=1}^{N}\Big\{(\hat{y}^n(k+\tau)-y_d(k+\tau))^2+2(\hat{y}(k+\tau)-y_d(k+\tau)) < \eta^n(k+\tau) > \\
&\qquad + < (\eta^n(k+\tau))^2 > \Big\}+\frac{\alpha}{2}\sum_{n=1}^{N}\mathbf{u}_n^T(k)\mathbf{u}_n(k) \\
&= \frac{\beta}{2}\sum_{n=1}^{N}\left[\{\hat{y}^n(k+\tau)-y_d^n(k+\tau)\}^2+\rho^2\right]+\frac{\alpha}{2}\sum_{n=1}^{N}\mathbf{u}_n^T(k)\mathbf{u}_n(k) \qquad (22)
\end{aligned}
$$

where we substituted $y(k+\tau)=\hat{y}(k+\tau)+\eta(k+\tau)$, evaluated the expected value and used the fact that $< \eta(k+\tau) >= 0$.

As discussed in Section II-B, the variance of the residual error of the system output is input–dependent and therefore should contribute to the derivation of the optimal control law. The minimization of Equation (22) with respect to the control signal will be discussed next for the affine and the general classes of nonlinear discrete time systems.

A. Affine Class Nonlinear Discrete Time Systems

In this section the problem of deriving the optimal control law corresponding to the maximum aposterior estimate will be discussed for a stochastic multi-input single-output affine class of nonlinear discrete time systems having the general form

$$
y(k+\tau)=f[\mathbf{x}(k)]+\mathbf{g}^T[\mathbf{x}(k)]\mathbf{u}(k)+\eta(k+\tau), \qquad (23)
$$

where $y(k+\tau)$ is the output, $\mathbf{u}(k)$ is the control input vector of dimension U, $\mathbf{x}(k)=[y(k),\dots,y(k-q+1),\mathbf{u}(k-1),\dots,\mathbf{u}(k-p+1)]$ is the system state vector, $f[\mathbf{x}(k)]:\mathcal{R}^{q+Up-U}\longmapsto\mathcal{R}$ and $\mathbf{g}[\mathbf{x}(k)]:\mathcal{R}^{q+Up-U}\longmapsto\mathcal{R}^{U}$ are unknown nonlinear functions of the state and $\eta(k+\tau)$ is an additive noise signal.

Two neural networks can then be used to approximate the nonlinear functions $f[\mathbf{x}(k)]$ and $\mathbf{g}[\mathbf{x}(k)]$. The network estimations are denoted as $\hat{f}[\mathbf{x}(k)]$ and $\hat{\mathbf{g}}[\mathbf{x}(k)]$. Since in this section the hyperparameter β is taken to be input–independent, then Equation (22) will have the same form as before, repeated here:

$$< S(\mathbf{u}(k)) >= \frac{\beta}{2}\sum_{n=1}^{N}\left[\{\hat{y}^{n}(k+\tau) - y_{d}^{n}(k+\tau)\}^{2} + \rho^{2}\right] + \frac{\alpha}{2}\sum_{n=1}^{N}\mathbf{u}_{n}^{T}(k)\mathbf{u}_{n}(k). \qquad (24)$$

To observe the actual value of the noise variance, the system output at time $k+\tau$ is required. To overcome this, another neural network is used to provide an estimate for the variance of the residual error of the system output. Although two neural network models could be used to estimate ρ^2 for the affine class of nonlinear discrete time systems, we believe this complication is not necessary and therefore use one neural network denoted by $\mathbf{h}[\mathbf{x}(k)]$. This yields the following form for the variance estimate of the residual error of the system output:

$$\hat{\rho}^{2} = \mathbf{h}^{T}[\mathbf{x}(k)]\mathbf{u}(k). \qquad (25)$$

Using Equation (25) in Equation (24) yields:

$$< S(\mathbf{u}(k)) >= \frac{\beta}{2}\sum_{n=1}^{N}\left[\{\hat{y}^{n}(k+\tau) - y_{d}^{n}(k+\tau)\}^{2} + \mathbf{h}_{n}^{T}[\mathbf{x}(k)]\mathbf{u}^{n}(k)\right] + \frac{\alpha}{2}\sum_{n=1}^{N}\mathbf{u}_{n}^{T}(k)\mathbf{u}_{n}(k). \qquad (26)$$

Minimization of the explicit performance index given in Equation (26) leads to the control law specified in the following theorem:

Theorem 1: The control law minimizing the performance index in Equation (26) subject to the system of Equation (23), is given by

$$\mathbf{u}^{*}(k) = \frac{\beta\{y_{d}(k+d) - \hat{f}[\mathbf{x}(k)]\}\hat{\mathbf{g}}[\mathbf{x}(k)] - \frac{\beta}{2}\mathbf{h}[\mathbf{x}(k)] + \frac{\beta^{2}}{2\alpha}\{\hat{\mathbf{g}}^{T}[\mathbf{x}(k)]\mathbf{h}[\mathbf{x}(k)]\hat{\mathbf{g}}[\mathbf{x}(k)] - \hat{\mathbf{g}}^{T}[\mathbf{x}(k)]\hat{\mathbf{g}}[\mathbf{x}(k)]\mathbf{h}[\mathbf{x}(k)]\}}{\beta\hat{\mathbf{g}}^{T}[\mathbf{x}(k)]\hat{\mathbf{g}}[\mathbf{x}(k)] + \alpha}. \qquad (27)$$

The proof of theorem 1 can easily be carried out by taking the derivative of Equation (26) with respect to the control signal and setting the derivative equal to zero.

B. General Nonlinear Input-Output Model

Here the objective is to derive the optimal control law for a stochastic multi-input single-output, general class of nonlinear discrete time systems having the following form

$$y(k+d) = f(y(k), \ldots, y(k-q+1), \mathbf{u}(k), \mathbf{u}(k-1), \ldots, \mathbf{u}(k-p+1)) + e(k+d), \qquad (28)$$

A neural network model having the following general input-output form

$$\hat{y}(k+d) = N_f(y(k),\dots,y(k-q+1),\mathbf{u}(k),\mathbf{u}(k-1),\dots,\mathbf{u}(k-p+1)), \qquad (29)$$

can then be used to estimate the system output given in Equation (28).

Again following the discussion in Section II-B, a neural network model is then used to provide an estimate for the input–dependent variance of the residual error of the system output, which is taken for the general nonlinear class of discrete time system discussed in this section to be of the following form:

$$\hat{\rho}^2 = h(y(k),\dots,y(k-q+1),\mathbf{u}(k),\mathbf{u}(k-1),\dots,\mathbf{u}(k-p+1)). \qquad (30)$$

Using this in Equation (22) yields:

$$< S(\mathbf{u}(k)) >= \frac{\beta}{2}\Bigg[\sum_{n=1}^{N}\{\hat{y}^n(k+d) - y_d^n(k+d)\}^2$$
$$+ h^n(y(k),\dots,y(k-q+1),\mathbf{u}(k),\mathbf{u}(k-1),\dots,\mathbf{u}(k-p+1))\Bigg] + \frac{\alpha}{2}\sum_{n=1}^{N}\mathbf{u}_n^T(k)\mathbf{u}_n(k). \qquad (31)$$

The optimal control law can then be found by setting the derivative of Equation (31) with respect to the control signal to zero

$$0 = \beta[\hat{y}(k+d) - y_d(k+d)]\frac{\partial\hat{y}(k+d)}{\partial\mathbf{u}(k)} + \frac{\beta}{2}\frac{\partial\hat{\rho}}{\partial\mathbf{u}(k)} + \alpha\mathbf{u}(k). \qquad (32)$$

Since the forward model of the system output and the model of its variance are nonlinear functions of the control signal, a nonlinear optimization method is required for solving Equation (32) and deriving the optimal control law. A closed form for the optimal control law cannot be found.

V. Gaussian Approximation to the Posterior Distribution

Given the definitions of the forward model distribution and the prior of control signals, the posterior distribution defined in Equations (19) and (21) is then specified. The evaluation of the probability distribution of the forward model predictions as well as the evidence for the hyperparameters requires integration over control signals. In order to make these analytically tractable, some approximations need to be introduced. In the Bayesian work for estimating the weight parameters of neural networks, Mackay [26] used a Gaussian approximation for the posterior distribution of the weight vector. Similarly, here we use the Gaussian approximation for the posterior distribution of the control signal. This is obtained by considering the Taylor expansion of $< S(\mathbf{u}) >$ around its minimum value of $\mathbf{u}_{MP}(k)$ and retaining terms up to the second order so that

$$< S(\mathbf{u}(k)) >=< S(\mathbf{u}_{MP}(k)) > +\frac{1}{2}(\mathbf{u}(k)-\mathbf{u}_{MP}(k))A(\mathbf{u}(k)-\mathbf{u}_{MP}(k)), \tag{33}$$

where the linear term has vanished since we are expanding around a minimum of $S(\mathbf{u})$. Here A is the Hessian matrix of the total error function, with elements given by

$$A = \beta\hat{g}[\mathbf{x}(k)]\hat{g}^T[\mathbf{x}(k)] + \alpha I. \tag{34}$$

For the general nonlinear discrete time system discussed in Section IV-B, the elements of the Hessian matrix A are given by

$$A = \beta\left(\bigtriangledown^2 E_{\mathcal{Y}}^{MP} + \frac{1}{2}\bigtriangledown^2 \hat{\rho}^{MP}\right) + \alpha I, \tag{35}$$

where $E_{\mathcal{Y}}^{MP} = \frac{1}{2}[\hat{y}(k+d)|_{\mathbf{u}_{MP}} - y_d(k+d)]^2$. A variety of exact and approximate methods for evaluating the Hessian of the error function $E_{\mathcal{Y}}$ can be found in [3].

The expansion in Equation (33) leads to a posterior distribution which is a Gaussian function of the control signals, given by

$$p(\mathbf{u}(k) \mid \mathcal{Y}(\mathbf{k}+\mathbf{d}), \mathbf{y}_d(k+d), \mathbf{x}(k)) = \frac{1}{Z_S^*}\exp\left(-S(\mathbf{u}_{MP}(k)) - \frac{1}{2}\triangle\mathbf{u}^T(k)A\triangle\mathbf{u}(k)\right), \tag{36}$$

where $\triangle\mathbf{u}(k) = \mathbf{u}(k) - \mathbf{u}_{MP}(k)$ and Z_S^* is the normalization constant appropriate to the Gaussian approximation.

Based on the above Gaussian approximation, the normalization factor Z_S^* can now be evaluated to give

$$Z_S^*(\alpha,\beta) = e^{-S(\mathbf{u}_{MP}(k))}(2\pi)^{U/2} \mid A \mid^{-\frac{1}{2}}. \tag{37}$$

VI. Distribution of System Output

The Bayesian formalism above is used for estimating the posterior distribution of control signals. Once the forward and inverse models of the system have been trained and converged, the distribution of control signals will affect the distribution of the trajectory error and in-turn the distribution of the system output. In addition, there will be a contribution of the trajectory error distribution arising from the assumed Gaussian noise on the error trajectory. In this section the distribution of system output values using the Gaussian approximation of the posterior distribution will be calculated.

Using the rules of probability, the distribution of the trajectory error for a given input can be written in the form

$$p(e_t(k+\tau) \mid \mathbf{x}(k), y_d(k+\tau), \mathcal{E}_t(k+\tau)) = \int p(e_t(k+\tau) \mid \mathbf{u}(k), \mathbf{x}(k), y_d(k+\tau)) \, p(\mathbf{u}(k) \mid \mathcal{E}_t(k+\tau), y_d(k+\tau), \mathbf{x}(k)) d\mathbf{u}(k), \quad (38)$$

where $p(\mathbf{u}(k) \mid \mathcal{E}_t(k+\tau), y_d(k+\tau), \mathbf{x}(k))$ is the posterior distribution of the control signal. The distribution $p(e_t(k+\tau) \mid \mathbf{u}(k), \mathbf{x}(k), y_d(k+\tau))$ is the model for the distribution of noise on the trajectory error , and is given by Equation (16).

The evaluation of this integral will require the use of the Gaussian approximation given in Equation (36) for the posterior distribution of control signals, together with the expression given in Equation (16) for the distribution of the trajectory error. This gives

$$p(e_t(k+\tau) \mid \mathbf{x}(k), y_d(k+\tau), \mathcal{E}_t(k+\tau)) \propto \int \exp\left(-\frac{\beta}{2}\{y(k+\tau) - y_d(k+\tau)\}^2\right) \exp\left(-\frac{1}{2} \triangle \mathbf{u}^T(k) A \triangle \mathbf{u}(k)\right) d \triangle \mathbf{u}(k). \quad (39)$$

The evaluation of this distribution will be carried out for the two classes of models described earlier in the paper: the affine class of nonlinear discrete time systems, and the general nonlinear class of discrete time systems.

A. Affine Class of Nonlinear Discrete Time Systems

This section is concerned with the evaluation of the distribution of the system output for the affine class of discrete time systems defined in Equation (23). Using the neural estimates of the nonlinear models of Equation (23) in Equation (39), and the fact that $y(k+\tau) = \hat{y}(k+\tau) + \eta(k+\tau)$

yields

$$
\begin{aligned}
&p(\eta(k+\tau) \mid \mathbf{x}(k), y_d(k+\tau), \mathcal{E}_t(k+\tau)) \\
&\propto \int \exp\left(-\frac{\beta}{2}\{\eta(k+\tau)+\hat{f}[\mathbf{x}(k)]+\hat{\mathbf{g}}^T[\mathbf{x}(k)]\mathbf{u}(k)-y_d(k+\tau)\}^2\right)\exp\left(-\frac{1}{2}\triangle\mathbf{u}^T(k)A\triangle\mathbf{u}(k)\right)d\triangle\mathbf{u}(k) \\
&= \int \exp\Big(-\frac{\beta}{2}\{\eta(k+\tau)+\hat{f}[\mathbf{x}(k)]+\hat{\mathbf{g}}^T[\mathbf{x}(k)][\mathbf{u}_{MP}(k)+\triangle\mathbf{u}(k)]-y_d(k+\tau)\}^2\Big) \\
&\qquad \exp\Big(-\frac{1}{2}\triangle\mathbf{u}^T(k)A\triangle\mathbf{u}(k)\Big)d\triangle\mathbf{u}(k). \qquad (40)
\end{aligned}
$$

The evaluation of the integral in Equation (40) is a Gaussian distribution of the form

$$
\begin{aligned}
&p(\eta(k+\tau) \mid \mathbf{x}(k), y_d(k+\tau), \mathcal{E}_t(k+\tau)) \\
&= \frac{1}{(2\pi\sigma_\eta^2)^{1/2}}\exp\left(-\frac{\{\eta(k+\tau)-[y_d(k+\tau)-\hat{f}[\mathbf{x}(k)]-\hat{\mathbf{g}}^T[\mathbf{x}(k)]\mathbf{u}_{MP}(k)]\}^2}{2\sigma_\eta^2}\right). \qquad (41)
\end{aligned}
$$

This distribution has a mean given by $y_d(k+\tau)-\hat{f}[\mathbf{x}(k)]-\hat{\mathbf{g}}^T[\mathbf{x}(k)]\mathbf{u}_{MP}(k)$, and a variance given by

$$
\sigma_\eta^2 = \frac{1}{\beta} + \hat{\mathbf{g}}^T[\mathbf{x}(k)]A^{-1}\hat{\mathbf{g}}[\mathbf{x}(k)]. \qquad (42)
$$

Following the discussion in Section II-B, the mean value of the residual error of the system output is zero. This means that $\hat{f}[\mathbf{x}(k)]+\hat{\mathbf{g}}^T[\mathbf{x}(k)]\mathbf{u}_{MP}(k) = y_d(k+\tau)$, i.e after training the forward and inverse models, the expected value of the system output will be equal to the desired output value.

It is clear from Equation (42) that the standard deviation of the predictive distribution of the system output has two contributions. The first contribution arises from intrinsic noise on the trajectory error which in turn is the intrinsic noise on the system output, represented by the first term of Equation (42). The second contribution arises from the width of the posterior distribution of control signals, represented by the second term of Equation (42).

B. General Nonlinear Input-Output Model

Here the evaluation of the distribution of the system output for the general class of nonlinear discrete time systems defined in Equation (28) will be discussed. Since the forward neural network model of the system output is a nonlinear function of control signals, we need to introduce some simplifying approximations. Here we shall assume that the width of the posterior distribution of control signals (determined by the Hessian matrix A) is sufficiently narrow to allow us to approximate the network function $\hat{y}(k+\tau) = N_f(y(k), \ldots, y(k-q+1), \mathbf{u}(k), \mathbf{u}(k-1), \ldots, \mathbf{u}(k-p+1))$ by its linear expansion around $\mathbf{u}_{MP}(k)$,

$$\hat{y}(k+\tau) = \hat{y}_{MP}(k+\tau) + J^T \triangle \mathbf{u}(k), \tag{43}$$

where

$$J \equiv \nabla_{\mathbf{u}(k)} \hat{y}(k+\tau) \mid_{\mathbf{u}_{MP}(k)} . \tag{44}$$

Using this and the fact that $y(k+\tau) = \hat{y}(k+\tau) + \eta(k+\tau)$ in Equation (39) yields

$$\begin{aligned} & p(\eta(k+\tau) \mid \mathbf{x}(k), y_d(k+\tau), \mathcal{E}_t(k+\tau)) \\ & \quad \propto \int \exp\left(-\frac{\beta}{2}\{\eta(k+\tau) + \hat{y}_{MP}(k+d) + J^T \triangle \mathbf{u}(k) - y_d(k+\tau)\}^2 \right) \\ & \qquad \exp\left(-\frac{1}{2} \triangle \mathbf{u}^T(k) A \triangle \mathbf{u}(k) \right) d \triangle \mathbf{u}(k), \end{aligned} \tag{45}$$

where $\hat{y}_{MP}(k+d) = \hat{y}(k+d) \mid_{\mathbf{u}_{MP}}$. The integral in Equation (45) can be evaluated as given in the Appendix to give a Gaussian distribution of the form

$$p(\eta(k+\tau) \mid \mathbf{x}(k), y_d(k+\tau), \mathcal{E}_t(k+\tau)) = \frac{1}{(2\pi\sigma_\eta^2)^{1/2}} \exp\left(-\frac{\{\eta(k+\tau) - [y_d(k+\tau) - \hat{y}_{MP}(k+d)]\}^2}{2\sigma_\eta^2} \right). \tag{46}$$

This distribution has a mean given by $y_d(k+\tau) - \hat{y}_{MP}(k+d)$ and a variance given by

$$\sigma_\eta^2 = \frac{1}{\beta} + J^T A^{-1} J. \tag{47}$$

Again following the discussion in Section II-B, the mean value of the residual error of the system output is zero. This means that $\hat{y}_{MP}(k+d) = y_d(k+\tau)$, i.e after training the forward and inverse models, the expected value of the system output will be equal to the desired output value.

Also the variance of Equation (47) can be interpreted as before for the affine class of nonlinear models. It has two terms, the first term corresponding to the intrinsic noise on the system output, and the second term corresponding to the width of the posterior distribution of control signal.

VII. The Evidence Framework for α and β

Although we may have enough information to estimate the noise level β of the trajectory error, the Bayesian framework can in principle provide a better way for handling the hyperparameters' estimation.

The correct Bayesian treatment for the hyperparameters α and β, whose values are unknown is to integrate them out of any predictions [3]. Rewriting the posterior distribution of the control signal to include dependencies in α and β explicitly gives,

$$\begin{aligned} p(\mathbf{u}(k) \mid \mathcal{E}_t) &= \int\int p(\mathbf{u}(k), \alpha, \beta, \mid \mathcal{E}_t) d\alpha d\beta \\ &= \int\int p(\mathbf{u}(k) \mid \alpha, \beta, \mathcal{E}_t) p(\alpha, \beta \mid \mathcal{E}_t) d\alpha d\beta. \end{aligned} \tag{48}$$

Suppose that the posterior probability distribution $p(\alpha, \beta \mid \mathcal{E}_t)$ for the hyperparameters in Equation (48) is sharply peaked around their most probable values α_{MP} and β_{MP}, then Equation (48) can be written as

$$\begin{aligned} p(\mathbf{u}(k) \mid \mathcal{E}_t) &= p(\mathbf{u}(k) \mid \alpha_{MP}, \beta_{MP}, \mathcal{E}_t) \int\int p(\alpha, \beta \mid \mathcal{E}_t) d\alpha d\beta, \\ &= p(\mathbf{u}(k) \mid \alpha_{MP}, \beta_{MP}, \mathcal{E}_t). \end{aligned} \tag{49}$$

This means that values of the hyperparameters which maximize the posterior probability need to be found first, and then the remaining calculations can be performed with the hyperparameters set to these values. In order to find α_{MP} and β_{MP} the posterior distribution of α and β need to be evaluated. This is given by

$$p(\alpha, \beta \mid \mathcal{E}_t) = \frac{p(\mathcal{E}_t \mid \alpha, \beta) p(\alpha, \beta)}{p(\mathcal{E}_t)}, \tag{50}$$

which requires a choice for the hyperprior $p(\alpha, \beta)$. Since the denominator in Equation (50) is independent of α and β, the maximum posterior of these values is found by maximizing the likelihood term $p(\mathcal{E}_t \mid \alpha, \beta)$, which is called the evidence for α and β.

Making the dependencies explicit on α and β in Equation (8), it can be written in the form

$$\begin{aligned} p(\mathcal{E}_t(k+\tau) \mid \alpha, \beta) &= \int p(\mathcal{E}_t(k+\tau) \mid \mathbf{u}(k), \alpha, \beta) p(\mathbf{u}(k) \mid \alpha, \beta) d\mathbf{u}(k), \\ &= \int p(\mathcal{E}_t(k+d) \mid \mathbf{u}(k), \beta) p(\mathbf{u}(k) \mid \alpha) d\mathbf{u}(k), \end{aligned} \tag{51}$$

where we made use of the fact that the prior is independent of β and the likelihood function is independent of α. Using the exponential forms given in Equations (9) and (14) for the prior distribution of control signals and the likelihood distribution, together with Equation (20) and (21), the evidence of α and β can then be written in the form

$$p(\mathcal{E}_{\mathbf{t}}(k+\tau) \mid \alpha, \beta) = \frac{1}{Z_{\mathcal{E}_t}(\beta)} \frac{1}{Z_u(\alpha)} \int \exp -S(\mathbf{u}(k)) d\mathbf{u}(k)$$
$$= \frac{Z_S(\alpha, \beta)}{Z_{\mathcal{E}_t}(\beta) Z_u(\alpha)}. \quad (52)$$

The normalization coefficients $Z_{\mathcal{E}_t}(\beta)$ and $Z_u(\alpha)$ have already been evaluated in Equations (13) and (18) respectively. For the Gaussian approximation of the posterior distribution of control signals, $Z_S(\alpha, \beta)$ is given in Equation (37). The log of the evidence is then given by

$$\ln p(\mathcal{E}_{\mathbf{t}}(k+\tau) \mid \alpha, \beta) = -\alpha E_u^{MP} - \beta E_{\mathcal{E}_t}^{MP} - \frac{1}{2} \ln \mid A \mid + \frac{NU}{2} \ln \alpha + \frac{N}{2} \ln \beta - \frac{N}{2} \ln(2\pi), \quad (53)$$

where A defined in Equations (34) and (35) for the affine and the general nonlinear classes of discrete time systems respectively is the Hessian matrix of the error function. For input–independent noise models considered in this paper, it clearly consists of two terms as can be seen from Equations (34) and (35). The first term represents the derivative of the likelihood error function while the second term is the derivative of the prior of the control signals.

Maximization of the explicit form of the log of the evidence defined in Equation (53) leads to re-estimation equations for α and β defined in the following theorem.

Theorem 2: The hyperparameters α and β maximizing the log of the evidence defined in Equation (53) are given by

$$\alpha^{\text{new}} = \frac{\gamma}{2E_u}, \quad (54)$$
$$\beta^{\text{new}} = \frac{N-\gamma}{2E_{\mathcal{Y}}}, \quad (55)$$

where the quantity γ is defined by

$$\gamma = \sum_{i=1}^{U} \frac{\lambda_i}{\lambda_i + \alpha}, \quad (56)$$

and where λ_i denote the eigenvalues of the Hessian matrix given by the first term of Equations (34) and (35) for the affine and the general nonlinear classes of discrete time systems respectively.

The proof of theorem 2 is given in the Appendix. It can easily be carried out by taking the derivative of Equation (53) with respect to α and β and setting the derivative equal to zero.

VIII. Implementation of the Proposed Bayesian Technique

since much of the theory has been covered in our discussion of the proposed Bayesian method, we summarize here the main steps to implement this method for practical applications. Only on–line implementation will be summarized which could easily be generalized to off–line implementation.

- Choose initial values for the hyperparameters α and β. Initialize network parameters of the forward model and variance network using values drawn from the prior distribution.
- At each time instant k,
  - Calculate the desired output to be followed by the system, $y_d(k+\tau)$.
  - Solve equation (27) if affine class or equation (27) if general nonlinear class for the optimal control law.
  - Forward the optimal control law to the plant and measure the system output, $y(k+\tau)$.
  - Based on the measured value of the system output, update the network parameters of the forward model.
  - Estimate the system output from the network of forward model.
  - Calculate the variance of the output, $\rho^2 = (y(k+\tau) - \hat{y}(k+\tau))^2$.
  - Update the parameters of the variance network based on ρ^2.
  - Re-estimate values for α and β using equations (54) and (55), with γ calculated using equation (56). This requires evaluation of the Hessian matrix from equation (34) if affine class or equation (35) if general nonlinear class, and evaluation of its eigenvalue spectrum.

The flow chart of implementing the proposed Bayesian method is shown in Figure 1. Example of practical implementation of the proposed Bayesian method will be given in the next section.

IX. Simulation Example

In this section a nonlinear SISO stochastic control problem is simulated. The dynamic equation of the system is

$$\begin{aligned} y(k+1) &= \sin[y(k)] + \cos[3y(k)] \\ &\quad + \{2 + \cos[y(k)]\}u(k) + e(k+1). \end{aligned} \tag{57}$$

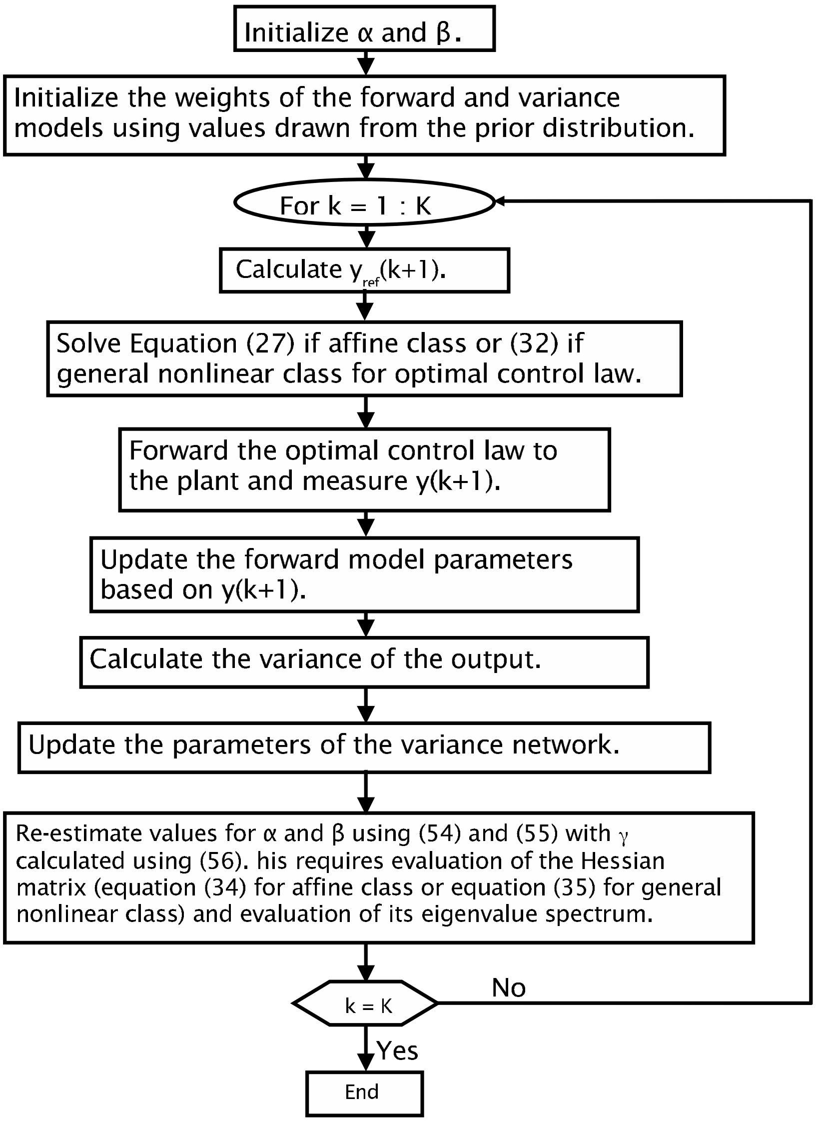

Fig. 1. On–line implementation of the proposed Bayesian method.

where $e(k+1)$ was assumed to be sampled from a Gaussian distribution, $\mathcal{N}(0, 0.2)$. This system has been used in [8], [9] to illustrate theoretical developments for suboptimal dual adaptive control.

In this paper two radial basis function neural networks with 7 and 3 Gaussian basis functions respectively are used to approximate the nonlinear functions $f(y(k)) = \sin[y(k)]+\cos[3y(k)]$ and $g(y(k)) = \{2+\cos[y(k)]\}$. The following reference model with input–output pairs $\{r(k), y_d(k+1)\}$

is chosen so that $y_d(k+1)$ represents the desired output behavior at time $k+1$

$$y_d(k+1) = r(k) + 0.0074y_d(k). \tag{58}$$

For comparison purposes, two sets of experiments were conducted to demonstrate the on–line training methods for the proposed Bayesian control algorithms and the conventional indirect adaptive control. On–line adaptation for the parameters of the forward models in the indirect adaptive control method and for the parameters of the forward models and the hyperparameters α and β in the Bayesian control method were conducted. The number of outer loops in the proposed Bayesian control algorithms is taken to be 1 in order to be consistence with the idea of online control. The same noise sequence, initial conditions, neural network structure, and reference input were used during implementation of each control method. The result is shown in Figure 2. As expected, the figure shows that the indirect adaptive control exhibits large transient overshoot because it is not taking into consideration the uncertainty of the forward and inverse models. Only after the initial period, when the parameters of the forward and inverse models converge, does the control assume good tracking. On the contrary, and although a different source of uncertainty is introduced (which is mainly coming from the uncertainty introduced from the online estimation of the hyperparameters), the proposed Bayesian method shows better characteristics in the transient response reflecting the use of knowledge about uncertainty of the forward and inverse models. The average tracking error of the Bayesian method and coventional adaptive control were 0.2223 and 2.8604 respectively.

X. CONCLUSIONS

The nonstandard formulation of the proposed probabilistic control method presented here can be viewed as an alternative method to the traditional deterministic adaptive control methods. It provides a theoretical foothold for a wider aim in probabilistic controller design.

Throughout the paper, the basic paradigm for the proposed probabilistic control method has been developed. The discussion has demonstrated the proposed method for two different classes of nonlinear uncertain systems. For the affine class of nonlinear uncertain discrete time systems a closed form for the optimal control law was found. The results were then extended to the more general class of nonlinear uncertain discrete time systems where the system equations are taken to be nonlinear functions in both the previous input and output values. Because of the

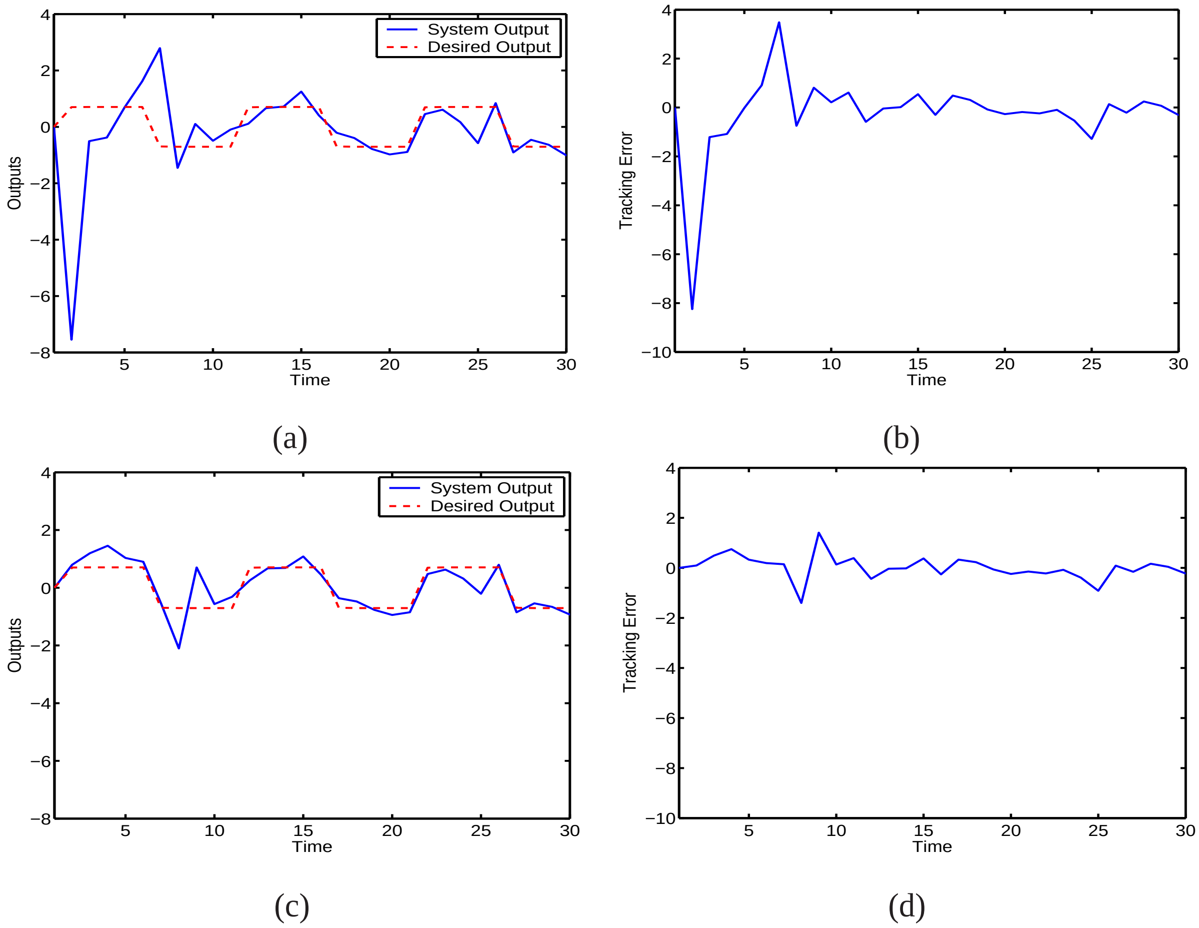

Fig. 2. Results of on–line control: output and tracking error (a) the actual and reference model outputs of standard adaptive control. (b) tracking error of standard adaptive control. (c) the actual and reference model outputs of the proposed Bayesian method. (d) tracking error of the proposed Bayesian method.

nonlinearity of the system equations a closed form for the optimal control strategy could not be found.

This more general framework for adaptive control methods has the major advantage that we can now incorporate uncertainty (in models and parameters) in a more structured framework. Different levels of uncertainty and noise models can be treated consistently using the inference machinery. Moreover, the proposed method provides an estimate for the uncertainty of the system output as shown in equations (42) and (47). Note that traditional methods which do not estimate distributions, cannot produce an estimate for the system output uncertainty. Simulation studies have validated these theoretical findings.

In the current study, the hyperparameter, β, which represents the noise variance of the trajectory error is restricted to be input–independent. This has always been assumed in Bayesian methods. In future work we will demonstrate how we can exploit the same machinery to deal with the case of an input–dependent hyperparameter.

APPENDIX

A. *Proof of Theorem 2*

This section will provide the proof for Theorem 2 of Section VII. To obtain the re-estimation equation of the hyperparameter α the maximum of Equation (53) with respect to α needs to be found. In order to differentiate $\ln \mid A \mid$ with respect to α, the Hessian matrix A of the error function is firstly written as $A = H + \alpha I$ where $H = \beta\left(\bigtriangledown^2 E_{\mathcal{Y}} + \frac{1}{2} \bigtriangledown^2 \hat{\rho}^{MP} \right)$ is the Hessian matrix of the unregularized error function. If $\{\lambda_i\}$, $(i = 1, \ldots, U)$ denote the eigenvalues of H, then A has eigenvalues $\lambda_i + \alpha$ and we have

$$\begin{aligned}\frac{d}{d\alpha} \ln \mid A \mid &= \frac{d}{d\alpha} \ln \left(\prod_i (\lambda_i + \alpha) \right) \\ &= \frac{d}{d\alpha} \sum_i \ln(\lambda_i + \alpha) \\ &= \sum_i \frac{1}{\lambda_i + \alpha} = \mathrm{Tr} A^{-1}. \end{aligned} \tag{59}$$

The previous derivation has implicitly assumed the independency of the eigenvalues λ_i on α. This assumption is true for an error function $E_{\mathcal{Y}}$ which is exactly a quadratic function of the control signal as is the case for the affine class of nonlinear discrete time systems, because the Hessian matrix of the error function is independent of $\mathbf{u}$. For the general nonlinear class of discrete time systems, the Hessian matrix will be a function of $\mathbf{u}$, consequently we see that the result in Equation (59) actually neglects terms involving $d\lambda/d\alpha$.

With Equation (59), the maximization of Equation (53) with respect to α then leads to the result that,

$$2\alpha E_u^{MP} = U - \sum_{i=1}^{U} \frac{\alpha}{\lambda_i + \alpha} = \gamma, \tag{60}$$

where γ is defined in Equation (56). Now we consider the maximization of Equation (53) with respect to β. Since λ_i are the eigenvalues of $H = \beta\left(\bigtriangledown^2 E_{\mathcal{Y}} + \frac{1}{2} \bigtriangledown^2 \hat{\rho}^{MP} \right)$ it follows that λ_i is directly proportional to β and hence

$$\frac{d\lambda_i}{d\beta} = \frac{\lambda_i}{\beta}. \tag{61}$$

Thus the derivative of the log of the Hessian matrix $\ln \mid A \mid$ of the error function with respect to β is

$$\frac{d}{d\beta} \ln \mid A \mid = \frac{d}{d\beta} \sum_i \ln(\lambda_i + \alpha) \\ = \frac{1}{\beta} \sum_i \frac{\lambda_i}{\lambda_i + \alpha}. \tag{62}$$

The maximum of Equation (53) with respect to β is then be given by

$$2\beta E_{\mathcal{Y}}^{MP} = N - \sum_{i=1}^{u} \frac{\lambda_i}{\lambda_i + \alpha} = N - \gamma. \tag{63}$$

Rearranging Equations (60) and (63) yields the re-estimation equation of the hyperparameters α and β given by Equations (54) and (55) respectively.

REFERENCES

[1] J. S. Albus. A new approach to manipulator control: The cerebellar model articulation controller (CMAC). *Transactions of the ASME Journal of Dynamic Systems, Measurements, and Control*, 97:220–227, 1975.

[2] K. J. Åström and B. Wittenmark. *Adaptive Control*. Addison-Wesley, Reading, MA, U.S.A, 1989.

[3] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, New York, N.Y., 1995.

[4] S. .S. Chan and M. B. Zarrop. A suboptimal dual controller for stochastic systems with unknown parameters. *International Journal of Control*, 41(2):507–524, 1985.

[5] F. C. Chen and H. K. Khalil. Adaptive control of a class of nonlinear discrete-time systems. *IEEE Transactions on Automatic Control*, 40(5):791–801, May 1995.

[6] F. C. Chen and C. C. Liu. Adaptively controlling nonlinear continuous-time systems using multilayer neural networks. *IEEE Transactions on Automatic Control*, 39(6):1306–1310, June 1994.

[7] S. Fabri and V. Kadirkamanathan. Dynamic structure neural networks for stable adaptive control of nonlinear systems. *IEEE Transactions on Neural Networks*, 7(5):1151–1167, September 1996.

[8] S. Fabri and V. Kadirkamanathan. Dual adaptive control of nonlinear stochastic systems using neural networks. *Automatica*, 34(2):245–253, February 1998.

[9] S. G. Fabri and V. Kadirkamanathan. *Functional Adaptive Control: An Intelligent Systems Approach*. Springer-Verlag, February 2001.

[10] A. A. Fel'dbaum. Dual control theory i-ii. *Automation and Remote Control*, 21:874–880, 1033–1039, 1960.

[11] A. A. Fel'dbaum. Dual control theory iii-iv. *Automation and Remote Control*, 22:1–12, 109–121, 1961.

[12] A. A. Fel'dbaum. *Optimal Control Systems*. Academic Press, New York, 1965.

[13] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, 1992.

[14] G. C. Goodwin and K. S. Sin. *Adaptive Filtering, Prediction and Control*. Prentice-Hall, Englewood Cliffs, New Jersey, 1984.

[15] R. Herzallah and D. Lowe. Improved robust control of nonlinear stochastic systems using uncertain models. In *Controlo2002*, pages 507–512, Aveiro, Portugal, September 2002.

[16] R. Herzallah and D. Lowe. A novel approach to modelling and exploiting uncertainty in stochastic control systems. In *International Conference on Artificial Neural Networks, ICANN*, pages 801–806, Madrid, Spain, August 2002.

[17] R. Herzallah and D. Lowe. Robust control of nonlinear stochastic systems by modelling conditional distributions of control signals. *Neural Computing and Applications*, 12:98–108, 2003.

[18] R. Herzallah and D. Lowe. A Bayesian approach to modeling the conditional density of the inverse controller. In *IEEE Conference on Control Applications, CCA*, volume 2, pages 954–959, Taipei, Taiwan, September 2004.

[19] R. Herzallah and D. Lowe. A mixture density network approach to modelling and exploiting uncertainty in nonlinear control problems. *Engineering Applications of Artificial Intelligence*, 17:145–158, February 2004.

[20] S. Jagannathan and F. L. Lewis. Multilayer discrete-time neural-net controller with guaranteed performance. *IEEE Transactions on Neural Networks*, 7(1):107–130, January 1996.

[21] H.J. Kappen. A linear theory for control of non-linear stochastic systems. *arXiv:physics/0411119v1*, 2004.

[22] M. Kárný. Towards fully probabilistic control design. *Automatica*, 32(12):1719–1722, 1996.

[23] M. Kárný, J. Böhm, T. V. Guy, L. Jirsa, I. Nagy, P. Nedoma, and L. Tesař. *Optimised Bayesian Dynamic Advising: Theory and Algorithms*. Springer Wien, 2005.

[24] M. Kárny, J. Böhm, T. V. Guy, and P. Nedoma. Mixture-based adaptive probabilistic control. *International Journal of Adaptive Control and Signal Processing*, 17(2):119–132, 2003.

[25] L. Z. Liao and D. Li. Adaptive differential dynamic programming for multiobjective optimal control. *Automatica*, 38, 2002.

[26] D. J. C. Mackay. Bayesian interpolation. *Neural Computation*, 4:415–447, 1992.

[27] A. L. Maitelli and T. Yoneyama. A two-stage suboptimal controller for stochastic systems using approximate moments. *Automatica*, 30(12):1949–1954, 1994.

[28] R. Milito, C. S. Padilla, R. A. Padilla, and D. Cadorin. An innovations approach to dual control. *IEEE Transactions on Automatic Control*, 27(1):133–137, 1982.

[29] W. T. Miller, R. S. Sutton, and P. J. Werbos, editors. *Neural Networks for Control*. MIT Press, Cambridge, Massachusetts, London, England., 1990.

[30] K. S. Narendra and S. Mukhopadhyay. Adaptive control of nonlinear multivariable systems using neural networks. *Neural Networks*, 7(5):737–752, 1994.

[31] K. S. Narendra and K. Parthasarathy. Identification and control of dynamical systems using neural networks. *IEEE Transactions on Neural Networks*, 1:4–26, 1990.

[32] K. M. Passino. Intelligent control of autonomous systems. *IEEE Spectrum*, pages 55–62, June 1995.

[33] L. Pronzato, C. Kulcsar, and E. Walter. An actively adaptive control policy for linear models. *IEEE Transactions on Automatic Control*, 41(6):855–858, 1996.

[34] D. Psaltis, A. Sideris, and A. A. Yamamura. A multilayered neural network controller. *IEEE Control Systems Magazine*, 88(2):17–21, 1988.

[35] T. RayChaudhuri, L. G. C. Hamey, and R. D. Bell. From conventional control to autonomous intelligent methods. *IEEE Control Systems Magazine*, pages 78–84, October 1996.

[36] G. A. Rovithakis and M. A. Christodoulou. Adaptive control of unknown plants using dynamical neural networks. *IEEE Transactions on Systems, Man, and Cybernetics*, 24(3):4–26, 1994.

[37] R. M. Sanner and J. J. E. Slotine. Gaussian networks for direct adaptive control. *IEEE Transactions on Neural Networks*, 3(6), November 1992.

[38] E. Tse and Y. Bar-Shalom. An actively adaptive control for linear systems with random parameters via the dual control approach. *IEEE Transactions on Automatic Control*, 18(2):109–117, 1973.

[39] D. A. White and D. Sofge, editors. *Handbook of Intelligent Control*. Multiscience Press, Inc, New York, N.Y., 1992.

[40] Peter Whittle. *Optimization Over Time: Dynamic Programming and Stochastic Control*. John Wiley & Sons, Chichester. New York. Brisbane. Toronto. Singapore, 1982.