Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 50-1 (2017) 10407–10412

An Error-Entropy Minimization Algorithm for Tracking Control of Nonlinear Stochastic Systems with Non-Gaussian Variables ⋆

Yunlong Liu * **Aiping Wang** ** **Lei Guo** *** **Hong Wang** ****

* *State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819, China (e-mail: liuyunlongbuaa@163.com)*
** *Institute of Computer Sciences, Anhui University, Anhui, China*
*** *Science and Technology on Aircraft Control Laboratory, Beihang University, Beijing 100191, China (e-mail:lguo@buaa.edu.cn)*
**** *Pacific Northwest National Laboratory, 902 Battelle Blvd Richland, WA 99352 USA (e-mail: hong.wang@pnnl.gov)*

Abstract: This paper presents an error-entropy minimization tracking control algorithm for a class of dynamic stochastic systems. The system is represented by a set of time-varying discrete nonlinear equations with non-Gaussian stochastic input, where the statistical properties of stochastic input are unknown. By using Parzen windowing with Gaussian kernel to estimate the probability densities of errors, recursive algorithms are then proposed to design the controller such that the tracking error can be minimized. The performance of the error-entropy minimization is compared with the mean-square-error minimization in the simulation results.

Keywords: Minimum error entropy, information potential, non-Gaussian variables, probability density function, stochastic systems.

1. INTRODUCTION

Stochastic control is an active part of control theory which deals with data that are polluted by stochastic noise and the aim of stochastic control is to design optimal controller that performs the desired control task with minimum cost despite the existence of these disturbance. Along with the development of Pontryagin's maximum principle (MP), Bellman's dynamic programming (DP) and Kalman's linear-quadratic (LQ) control, stochastic optimal control theory has been well developed since early 1960s which mainly focus on Gaussian stochastic systems under the assumption that the noises obey Gaussian distribution (see Athans [1971]).

However, most real-life systems are governed by nonlinear models and most stochastic noises are far from being Gaussian distributed. Therefore, stochastic distribution control (see Wang [2000], Yue and Wang [2003] and references therein) for non-Gaussian stochastic system has been studied extensively in recent years in response to the increased requirements of many practical systems. The primary purpose of stochastic distribution control is to design control input such that the probability density function (PDF) of corresponding random variable is made as close as possible to a required distribution shape (see Yi et al. [2007], Yi et al. [2009]) or to minimize the entropy of corresponding random variable (see Guo and Wang [2006], Guo and Yin [2009], Liu et al. [2015]). Unfortunately, most of the existing studies focus on stochastic distribution control based on PDF functional operator mapping model that needs the priori knowledge of stochastic input, such as the PDF, which is a strong assumption and cannot always be satisfied in practice.

Since Mean square error (MSE), which only concentrates on second order statistics, is able to extract all possible information from a giving training data set under the linearity and Gaussianity assumptions, it has been well employed in the training of adaptive systems including linear filters and artificial neural networks due to the fact that Wiener [1949] established the perspective of adaptive filters as statistical function approximation. However, data densities take complex forms in many applications. When the probability distribution involved does not obey Gaussian distribution, MSE fails to capture all the information in the data. Since entropy, which was introduced by Shannon [1949], is a scalar quantity that provides a measure of the average information contained in random variable with a certain probability distribution function, minimum error entropy (MEE) is superior to MSE as an optimality criterion due to the fact that minimizing the entropy constrains all moments of the PDF. In Erdogmus and Principe [2002], the MEE criteria has been employed in adaptive systems training, and it has been proved that

⋆ This work was supported in part by the National Natural Science Foundation of China under Grant 61333007, 61621004, 61573022, the Chinese National Post-doctor Science Foundation under Grants 2015M571322 and CCSI of the Pacific Northwest National Laboratory.

10.1016/j.ifacol.2017.08.1720

minimizing the error entropy is equivalent to minimizing the distance between the joint densities of system input-output and the desired input-output pairs. Besides, the problem of optimal state estimation for stochastic systems has been considered from the view of information theory in Feng et al. [1997]. Also, in Xu et al. [2005], an adaptive Luenberger observer based on MEE has been designed to deal with the problem of nonlinear state estimation.

In this paper, an error entropy minimization algorithm is investigated in stochastic distribution control for a class of non-Gaussian stochastic systems with no priori knowledge of stochastic input. Since the error entropy minimization algorithm we proposed relied on the use of quadratic Renyi's entropy and the analytical error density distributions are not available by PDF functional operator mapping model, nonparametric estimation of the PDF of a random variable is required for the evaluation of its entropy. Parzen windowing (also called kernel density estimator) is an typical density estimation scheme, where the PDF is approximated by a sum of kernels whose centers are translated to the sample locations. A commonly used kernel function is the Gaussian, since it is continuously differentiable and it leads to continuously differentiable density estimates, which can provide a computational simplification in the gradient-based algorithm design (see Principle et al. [2000]).

The organization of this paper is as follows. The tracking control problem is formulated in Section 2 for a class of nonlinear stochastic systems with no priori knowledge of stochastic input. The main results and detailed derivations are given in Section 3. A numerical simulation is given in Section 4 and the performances between MSE and MEE are compared. Finally, some conclusions are given in Section 5.

2. PRELIMINARIES AND PROBLEM FORMULATION

Consider the following discrete time nonlinear stochastic system:

$$\begin{aligned} x_{k+1} &= f(x_k) + B_k u_k + H_k \omega_k, \quad x_k|_{k=0} = x_0 \\ y_k &= C_k x_k + v_k \end{aligned} \tag{1}$$

where $x_k \in R^n$, $y_k \in R$ and $u_k \in R$ are the state, measured output and control input, respectively. $f(\cdot)$ is a vector-valued nonlinear function. $x_0 \in R^n$ is the initial condition. $\omega_k \in R^s, v_k \in R$ are the additive system noise and measurement noise respectively, which are of arbitrary distribution form rather than Gaussian distribution. B_k, H_k, C_k are known parameter matrices with appropriate dimensions.

The following assumptions, which are quite general, are required to simplify the controller design procedure.

Assumption 1. $\{\omega_k\}, \{v_k\}$ $(k = 0, 1, \cdots)$ are bounded, stationary and mutually independent with no priori knowledge of statistical property.

Assumption 2. The initial state x_0 is independent of ω_k, v_k.

Assumption 3. $f(\cdot)$ is a known Borel measurable and smooth function of its arguments, and assumed to satisfy $f(0) = 0$ and

$$\|f(x_k + \delta) - f(x_k) - A_k \delta\| \leq a_1 \|\delta\| \tag{2}$$

where $A_k \in R^{n\times n}$ is a known constant matrix, $\delta \in R^n$ is a vector and a_1 is a known positive constant.

Remark 1. The nonlinear description (2), which has been adopted in Yaz and Azemi [1993], quantify the maximum possible derivations from a linear model with A_k as its system parameter matrix.

The purpose of this paper is to construct the input signal u_k such that the system state x_k can track that of the ideal deterministic model (see Athans [1971]) with certain accuracy, following

$$\begin{aligned} x_{r(k+1)} &= f(x_{rk}) + B_k r_k, \\ y_{rk} &= C_k x_{rk} \end{aligned} \tag{3}$$

where $x_{rk} \in R^n$, $r_k \in R$ and $y_{rk} \in R$ are the ideal deterministic state, bounded input, and output, respectively.

Define the tracking error as $\xi_k = x_k - x_{rk}$, it can be seen that ξ_k obeys non-Gaussian distribution since the nonlinearity of system and the non-Gaussianity of random noises. Hence, quadratic Renyi's entropy, which is given in the following equation, is employed beyond MSE (second-order statistics) to quantify the statistical property for convenience of calculation.

$$H_2(X) = -\ln \int p_X^2(x) dx \tag{4}$$

where $p_X(x)$ is the PDF of random variable X.

Since the statistical properties of the stochastic inputs, such as PDFs, are unavailable in this paper, neither the entropy nor the PDF of the tracking error can be calculated by the functional operator mapping model (see Guo and Wang [2006], Guo and Yin [2009], Liu et al. [2015]). Fortunately, Parzen windowing is an efficient way to approximate the PDF of a given sample distribution (see Devroye and Lugosi [2001]), especially in low-dimensional spaces, and it does not need prior knowledge of the system apart from the samples. For a given set of i.i.d samples $z_1, \cdots, z_n$ drawn from $q(z)$, the Parzen windowing estimate for the distribution, assuming a fixed-size kernel function $K_\sigma(\cdot)$ for simplicity, is given by

$$p(z) = \frac{1}{n} \sum_{i=1}^{n} K_\sigma(z - z_i) \tag{5}$$

where σ is the window width, and can be optimized in accordance with the least-square cross-validation, likelihood cross-validation, the test graph method or other rules-of-thumb method (see Devroye and Lugosi [2001], Principe and Erdogmus [2000]).

And Gaussian function, which is given in the following equation, is chosen as kernel function in this paper as it is continuously differentiable and can simplify the computation in the algorithm design (see Principe et al. [2000]).

$$K_\sigma(x) = G(x - \mu, \Sigma) = \frac{1}{(\sqrt{2\pi})^n |\Sigma|^{\frac{1}{2}}} e^{-\frac{(x-\mu)^T |\Sigma|^{-1} (x-\mu)}{2}} \tag{6}$$

where μ and Σ are the mean and covariance matrix, respectively. Since entropy is invariant to the mean of the sample data, μ is chosen as 0 and Σ is selected as