4.1 Introduction

The identification of nonlinear systems can be posed as a nonlinear functional approximation problem. From the Weierstrass Theorem (Powell, 1981) and the Kolmogorov theorem (Sprecher, 1965) in approximation theory, it is shown that the polynomial and many other approximation schemes can approximate a continuous function arbitrarily well. In recent years, a number of nonlinear system identification approaches, particularly identification using neural networks, based on the universal approximation theorem (Cybenko, 1989), are applications of a similar mathematical approach.

Using the approximation approach, two key questions concerning nonlinear system identification are important: how to judge the accuracy for the nonlinear function being approximated and how to choose nonlinear function units to guarantee the accuracy. Most nonlinear system identification approaches fix the number of nonlinear function units and use only a single performance function, e.g., $L_2$-norm of the difference between the real nonlinear system and the nonlinear model which results in the well-known least squares algorithm, to measure and judge the accuracy of the identification model and to optimise the approximation. The assumption behind choosing the $L_2$-norm is that the noise in the process and measurements has Gaussian (normal) distributions.

In nonlinear system identification there are often a number of objectives to be considered. The objectives are often conflicting and no identification which can be considered best with respect to all objectives exists. Hence, there is an inevitable trade-off between objectives, for example, the distance measurement and maximum difference measurement between the real nonlinear system and the nonlinear model. Model comparison methods, such as information criterion (Akaike, 1974), Bayesian model selection (MacKay, 1992) and minimum description length (MDL) (Rissanen, 1989), consider two such objectives, namely, Euclidean distance ($L_2$-norm) and model complexity. These procedures allow the selection of the best amongst a small number of candidate models (MacKay, 1992). In addition to the above two objectives, we consider the $L_\infty$-norm of the difference between the real nonlinear system and the nonlinear model because it represents the accuracy bound of the approximation achieved by the estimated model. These considerations lead to the study of multiobjective nonlinear system identification.
In this chapter, three multiobjective performance functions are introduced to measure the approximation accuracy and the complexity of the nonlinear model for noise with mixed distribution. Those functions are the $L_2$- and $L_\infty$-norms of the difference measurements between the real nonlinear system and the nonlinear model, and the number of nonlinear units in the nonlinear model. Genetic algorithms are used to search for a suboptimal set of nonlinear basis functions of the model to simplify model estimation. Two neural networks are applied for the model representation of the nonlinear systems. One is the Volterra polynomial basis function (VPBF) network and the other is the Gaussian radial basis function (GRBF) network. A numerical algorithm for multiobjective nonlinear model selection and identification using neural networks and genetic algorithms is also detailed. Two applications in identification of a nonlinear system and approximation of a nonlinear function with a mixed noise demonstrate the operation of the algorithm.

4.2 Multiobjective Modelling with Neural Networks

The modelling of nonlinear systems has been posed as the problem of selecting an approximate nonlinear function between the inputs and the outputs of the systems. For a single-input single-output system, it can be expressed by the nonlinear auto-regression moving average model with exogenous inputs (NARMAX) (Chen and Billings, 1989), that is,

$$ y(t) = f(y(t-1), y(t-2), ..., y(t-n_y), u(t-1), u(t-2), ..., u(t-n_u) + e(t) $$  \hspace{1cm} (4.1)

where $f(.)$ is an unknown nonlinear function, $y$ is the output, $u$ is the control input and $e$ is the noise, respectively, $n_y, n_u, n_e$ are the corresponding maximum delays. It is assumed that the noise $e(t)$ is a white noise. For the colour noise case, the modelling of the system using neural networks below needs some slight modifications, as suggested in Nerrand et al. (1994).

The nonlinear function $f(.)$ in the above NARMAX model can be approximated by a single-layer neural network, i.e., a linear combination of a set of basis functions (Billings and Chen, 1992; Liu et al., 1998a).

$$ \hat{f}(x, p) = \sum_{k=1}^{N} w_k \varphi_k(x, d_k) $$  \hspace{1cm} (4.2)

where

$$ x = [y(t-1), y(t-2), ..., y(t-n_y), u(t-1), u(t-2), ..., u(t-n_u)] $$  \hspace{1cm} (4.3)

$\varphi_k(x, d_k)$ ($k = 1, 2, ..., N$) is the basis function and $p$ is the parameter vector containing the weights $w_k$ and the basis function parameter vectors $d_k$. If the basis functions $\varphi_k(x, d_k)$ do not have the parameters $d_k$, then it is denoted by...