PREFACE

For the first time, this book presents a comprehensive and unifying introduction to kernel adaptive filtering. Adaptive signal processing theory has been built upon three pillars: the linear model, the least-squares cost and adaptive learning algorithms. When nonlinear models are required, this simplicity evaporates and a designer has to understand function approximation, neural networks, local minima, regularization etc. Is this the only way to go beyond the linear solution? Perhaps there is an alternative, which is the focus of this book. The basic concept is to perform adaptive filtering in a linear space that is nonlinearly related to the original input space. If this is possible, then all three pillars and our intuition about linear models can still be of use, and we end up implementing nonlinear filters in the input space.

This book will draw on the theory of reproducing kernel Hilbert spaces (RKHS) to implement the nonlinear transformation of the input to a high dimensional feature space induced by a positive definite function called *reproducing kernel*. If the filtering and adaptation operations to be performed in RKHS can be expressed by inner products of projected samples, then it is possible to calculate them by kernel evaluations directly in the input space. We use this approach to introduce a family of adaptive filtering algorithms in RKHS:

- the kernel least mean square algorithm;
- the kernel affine projection algorithms;
- the kernel recursive least squares algorithm and;
- the extended kernel recursive least squares algorithm.

These kernel-learning algorithms bridge closely two important areas of adaptive filtering and neural networks, and embody beautifully two important methodologies of error-correction learning and memory-based learning. The bottlenecks of the RKHS approach to nonlinear filter design are the need for regularization, the need to select the kernel function, and the need to curtail the growth of the filter structure. This book will present in a mathematically rigorous manner the issues and the solutions to all these problems, and will illustrate with examples the gains of performing adaptive filters in RKHS. More precisely:

Chapter 1 starts with an introduction to general concepts in machine learning, linear adaptive filters and conventional nonlinear methods. Then the theory of reproducing kernel Hilbert spaces is presented as the mathematical foundation of kernel adaptive filters. We stress that kernel adaptive filters are universal function approximators, have no local minima during adaptation and require reasonable computational resources.

Chapter 2 studies the kernel least mean square algorithm, the simplest among the family of kernel adaptive filters. We develop the algorithm in a step-by-step manner and delve into all the practical aspects of selecting the kernel function, picking the step-size parameter, sparsification and regularization. Two computer experiments, one with Mackey-Glass chaotic time series prediction and the other nonlinear channel equalization, are presented.

Chapter 3 covers the kernel affine projection algorithms, a family of four similar algorithms. The mathematical equations of filtering and adaptation are thoroughly derived from first principles, and useful implementation techniques are fully discussed. Many well-known methods can be derived as special cases of the kernel affine projection algorithms. Three detailed applications are included to show their wide applicability and design flexibility.

Chapter 4 presents the kernel recursive least squares algorithm and the theory of Gaussian process regression. A new sparsification approach, approximate linear dependency, is discussed. And with the aid of the Bayesian interpretation, we also present a powerful model selection method called maximum marginal likelihood. Two computer experiments are conducted to study the performance of different sparsification schemes and the effectiveness of maximum marginal likelihood to determine the kernel parameters.

Chapter 5 discusses the extended kernel recursive least squares algorithm on the basis of the kernel recursive least squares algorithm. We systematically study the problem of estimating the state of a linear dynamic system in RKHS from a sequence of noisy observations. Several important theorems are presented with proofs to outline the significance and basic approaches. This chapter contains two examples, Rayleigh channel tracking and Lorenz time series modeling.

Chapter 6 is devoted to addressing the principal bottleneck of kernel adaptive filters, i.e. their growing structure. We introduce a subjective information measure called *surprise* and present a unifying sparsification scheme to effectively curtail the growth of kernel adaptive filters. Three interesting computer simulations are presented to illustrate the theories.

This book should appeal to engineers, computer scientists and graduate students who are interested in adaptive filtering, neural networks and kernel methods. There are a total of 12 computer-oriented experiments distributed throughout the book that have been designed to reinforce the concepts discussed in the chapters. The computer experiments are listed in Table 1. Their MATLAB® implementations can be directly downloaded from the website http://www.cnel.ufl.edu/~weifeng/publication.htm. In order to keep the codes readable, we place simplicity over performance during design and implementation. These programs are provided without any further guarantees.

We have strived to fully reflect the latest advances of this emerging area in the book. Each chapter concludes with a summary of the state of the art and potential future directions for original research. This book should be a useful guide to both those who look for nonlinear adaptive filtering methodologies to solve practical problems and those who seek inspiring research ideas in related areas.

Table 1. A listing of all computer experiments in the book. MATLAB® programs that generate the results can be downloaded by all readers from the book's website.

Computer experiment	Topic
2.1	KLMS Applied to Mackey-Glass Time Series Prediction
2.2	KLMS in Nonlinear Channel Equalization
3.1	KAPA Applied to Mackey-Glass Time Series Prediction
3.2	KAPA in Noise Cancelation
3.3	KAPA in Nonlinear Channel Equalization
4.1	KRLS Applied to Mackey-Glass Time Series Prediction
4.2	Model Selection by Maximum Marginal Likelihood
5.1	EX-KRLS in Rayleigh Channel Tracking
5.2	EX-KRLS in Lorenz Time Series Prediction
6.1	Surprise Criterion in Nonlinear Regression
6.2	Surprise Criterion in Mackey-Glass Time Series Prediction
6.3	Surprise Criterion in CO_2 Concentration Forecasting