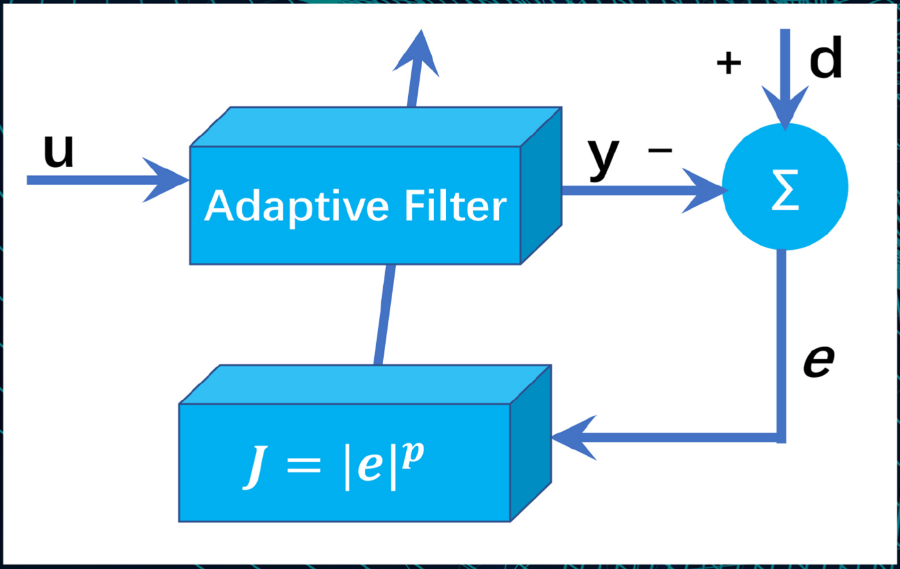
Adaptive Filtering Under Minimum Mean p-Power Error Criterion



Wentao Ma and Badong Chen

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Adaptive Filtering Under Minimum Mean p-Power Error Criterion

Adaptive filtering still receives attention in engineering as the use of the adaptive filter provides improved performance over the use of a fixed filter under the time-varying and unknown statistics environments. This application evolved communications, signal processing, seismology, mechanical design, and control engineering. The most popular optimization criterion in adaptive filtering is the well-known minimum mean square error (MMSE) criterion, which is, however, only optimal when the signals involved are Gaussian-distributed. Therefore, many "optimal solutions" under MMSE are not optimal. As an extension of the traditional MMSE, the minimum mean *p*-power error (MMPE) criterion has shown superior performance in many applications of adaptive filtering. This book aims to provide a comprehensive introduction of the MMPE and related adaptive filtering algorithms, which will become an important reference for researchers and practitioners in this application area. This book is geared to senior undergraduates with a basic understanding of linear algebra and statistics, graduate students, or practitioners with experience in adaptive signal processing.

Key Features:

- Provides a systematic description of the MMPE criterion.
- Many adaptive filtering algorithms under MMPE, including linear and nonlinear filters, will be introduced.
- Extensive illustrative examples are included to demonstrate the results.

Wentao Ma received a BSc in Mathematics and Applied Mathematics from Shaanxi University of Technology in 2003, an MSc in Computing Mathematics from Huazhong University of Science and Technology in 2008, and a Ph.D. in Information and Communication Engineering from Xi'an Jiaotong University, in 2015. Currently, he is an associate professor with the School of Electrical Engineering, Xi'an University of Technology, Xi'an, China. His research interests include statistical signal processing, machine learning, artificial intelligence, and their applications in Electrical and Computer Engineering. He has published over 50 papers in various journals and conference proceedings. He is a member of IEEE, IEEE PES, and CAA.

Badong Chen received BSc and MSc degrees in Control Theory and Engineering from Chongqing University in 1997 and 2003 respectively. He also received his Ph.D. in Computer Science and Technology from Tsinghua University in 2008. He was a postdoctoral associate at the University of Florida Computational Neuro Engineering Laboratory (CNEL) from 2010 to 2012. He was a visiting research scientist at the Nanyang Technological University (NTU), Singapore, in 2015. He also served as a senior research fellow with The Hong Kong Polytechnic University in 2017. Currently, he is a professor at the Institute of Artificial Intelligence and Robotics (IAIR), Xi'an Jiaotong University. His research interests include signal processing, machine learning, artificial intelligence, neural engineering, and robotics. He has published four books and over 300 papers in various journals and conference proceedings. He is an IEEE Senior Member and has served as a Technical Committee Member of IEEE SPS Machine Learning for Signal Processing (MLSP). He has been an associate editor of: *IEEE Transactions on Circuits and Systems for Video Technology*, *IEEE Transactions on Cognitive and Developmental Systems*, *IEEE Transactions on Neural Networks and Learning Systems*, *Journal of The Franklin Institute* and *Neural Networks*. He has also been on the editorial board of *Entropy*.

Adaptive Filtering Under Minimum Mean p-Power Error Criterion

Wentao Ma and Badong Chen

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

Designed cover image: Wentao Ma and Badong Chen

First edition published 2024
by CRC Press
2385 NW Executive Center Drive, Suite 320, Boca Raton FL 33431

and by CRC Press
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

CRC Press is an imprint of Taylor & Francis Group, LLC

ISBN: 9781032001654 (hbk)
ISBN: 9781032008622 (pbk)
ISBN: 9781003176114 (ebk)

DOI: 10.1201/9781003176114

Typeset in Palatino
by codeMantra

Contents

Symbols and Abbreviations

The main symbols and abbreviations used throughout the text are listed as follows:

k	discrete and continuous time index
t	continuous time index
$E\{\cdot\}$	the statistical-expectation operator
$\|\| \cdot \|\|$	vector or matrix norm
$I \in \mathbb{R}^{N\times N}$	the identity matrix
$(\cdot)^{\mathrm{T}}$	transpose of a vector or a matrix
$\mathrm{Tr}[\cdot]$	the trace of a matrix
$\lambda_i\{\cdot\}$	the *i*th eigenvalue of a matrix
$\mathrm{Re}(\cdot)$	the real part of a complex number
$\mathbf{u}()$	input signal vector
$\mathbf{R_{uu}}$	correlation matrix
$\mathbf{P}_{\mathbf{u}d}$	cross-correlation vector
$F_{XY}(x,y)$	joint distribution function
$\kappa(.,.)$	Mercer kernel
$\langle .,. \rangle_{\mathcal{H}}$	inner product
$\mathrm{sign}(\cdot)$	sign function
$\mathrm{diag}[\cdot]$	diagonal matrix
$\mathbf{A}^{-1}$	inverse of matrix **A**
$\mathbb{R}^n$	n-dimensional real Euclidean space
$\mathbf{w}_o$	optimal weight
$\tilde{\mathbf{w}}$	weight error power
$v(k)$	additive noise at time instant k
$\boldsymbol{\omega}$	weight vector in feature space **F**
$\tilde{\boldsymbol{\omega}}$	weight error power in feature space **F**
$\boldsymbol{\varphi}$	a nonlinear feature mapping
$\mathbf{F}$	a high-dimensional feature space
$\boldsymbol{\Psi}(\cdot)$	weighted-auto-correlation matrix
$\boldsymbol{\Phi}(\cdot)$	weighted-cross-correlation vector
γ	forgetting factor
MSE	mean square error
LAFs	linear adaptive filters
RMSE	root mean square error
EMSE	excess mean square error
MPE	mean p-power error
MFE	mean fourth error
MSD	mean square deviation
LAD	least absolute deviation
LMS	least mean square

LMF	least mean fourth
RLS	recursive least squares
FLOM	fractional lower order moment
LMP	least mean p-power
ALMP	adaptive least mean p-power
NLMP	normalized least mean p-power
SLMP	smoothed least mean p-power
VNLMP	variable normalization least mean p-power algorithm
ALMP	adaptive least mean p-power
PLMP	proportionate least mean p-power
KMPE	kernel mean p-power error
KMMPE	kernel mixture mean p-power error
LMMN	least mean mixed norm
RMN	robust mixed norm
GMN	generalized mixed norm
DGMN	diffusion generalized mixed norm
FIR	finite impulse response
KAF	kernel adaptive filtering
RKHS	reproducing kernel Hilbert spaces
KLMS	kernel least mean square
KRLS	kernel recursive least squares
KLMP	kernel least mean p-power
KRMN	kernel recursive mixed norm
RFF-EX-KRLP	random Fourier features extended KRLP
DAF	diffusion adaptive filtering
DLMS	diffusion least mean square
DRLS	diffusion recursive least squares
DLMP	diffusion least mean p-power
DNLMP	diffusion normalized least mean p-power
RDNLMP	robust diffusion normalized least mean p-power
ELM	extreme learning machine
RLS-ELM	recursive least square extreme learning machine
SRLS-ELM	sparse recursive least square extreme learning machine
LMPELM	least mean p-power extreme learning machine
RLMP-ELM	recursive least mean p-power extreme learning machine
SRLMP-ELM	sparse recursive least mean p-power extreme learning machine
BLS	broad learning system
LP-BLS	least p-power broad learning system
MN-BLS	mixed norm broad learning system
L2LP	mixed l_2-l_p adaptive algorithm
SNR	signal-to-noise ratio
GSNR	generalized SNR

Preface

Over the past few decades, adaptive filters have found widespread application in various scenarios, including system identification, echo cancellation, channel equalization, time series prediction, and so on. To ensure the efficient design of an adaptive filter, it is crucial to select an appropriate loss function (or criterion function) that can enhance the filter's convergence performance. The classical adaptive filtering algorithms, such as least mean square (LMS) and recursive least squares (RLS), are primarily developed based on the well-known minimum mean square error (MMSE) criterion, which performs exceptionally well when signals follow Gaussian distributions. However, the mean square error (MSE) loss function only captures the second-order statistics in the data and may result in suboptimal filtering performance in non-Gaussian situations, particularly when the underlying system is affected by noises of heavy-tailed or multimodal distributions.

To enhance the filtering performance in the presence of non-Gaussian noises, adaptive filters have been developed using various non-MSE (non-quadratic) loss functions. These include the mean p-power error (MPE) loss, Huber's loss, risk-sensitive loss, correntropy loss, error entropy loss, and others. Among these, the minimum MPE (MMPE) criterion is particularly noteworthy as it is a natural extension of MMSE and can capture higher order ($p > 2$) or lower order ($0 < p < 2$) statistics while being mathematically and computationally simple. The MMPE encompasses special cases such as the least absolute deviation (LAD) ($p = 1$), MMSE ($p = 2$), and least mean fourth (LMF) ($p = 4$). In practical applications, the MPE has demonstrated superior performance compared to the conventional MSE when used as a loss function in adaptive filtering. For a finite impulse response (FIR) filter, the MPE loss can yield a more accurate solution than the Wiener solution of MSE. In addition, by selecting an appropriate p, MPE-based adaptive filters can achieve faster and more robust convergence performance under heavy-tailed or light-tailed non-Gaussian noises.

To date, numerous adaptive filtering algorithms have been developed under the MMPE criterion. This book aims to consolidate all of these algorithms, along with their corresponding analysis and numerical results, into a single comprehensive resource. Many of the contents of this book were originally published in previous papers by the authors. This book is divided into eight chapters, with Chapter 1 providing an introduction to the background and outline of the book. Chapter 2 reviews classical adaptive filtering algorithms under the MMSE criterion, while Chapter 3 covers the basic definition and properties of MMPE as well as several extended versions of MMPE. Chapter 4 focuses primarily on gradient-based (LMS type) adaptive filtering algorithms under MMPE criterion, while Chapter 5 presents recursive

(RLS type) adaptive filtering algorithms under MMPE. Chapter 6 introduces some nonlinear adaptive filtering algorithms under MMPE criterion, and Chapter 7 focuses on adaptive filtering algorithms under mixture MMPE criterion. Finally, Chapter 8 discusses adaptive filtering under kernel MMPE criterion.

This book is a valuable resource for graduates, professionals, and researchers seeking to enhance the performance of adaptive filtering algorithms and design new adaptive algorithms under MMPE. It is also an excellent reference for those interested in adaptive system training and machine learning. In addition, this book can be used as a reference textbook for graduate or undergraduate students majoring in electronics communication, electrical or computer engineering.

The authors are grateful to the National Key R&D Program of China and National Natural Science Foundation of China, which have funded this book. We also acknowledge the support and encouragement from our colleagues and friends.

1

Introduction

1.1 Basic Knowledge of Adaptive Filtering Algorithms

Classical filters, such as the Wiener filter [1,2] and Kalman filter [3,4], require accurate estimation of the correlation coefficient and noise power of input signals for effective application. However, this is often difficult to achieve in practice, and inaccurate estimation can significantly impact filtering performance. In addition, the parameters of these filters are typically fixed and cannot be adjusted in response to changing input signals, limiting their real-time processing capabilities. To address these limitations and meet the demands of signal processing, adaptive filters (AFs) have been developed as a class of optimal filtering methods from the Wiener and Kalman filters. Unlike classical filters, AFs incorporate a feedback channel between the output and filter system, allowing for dynamic adjustment of filter coefficients based on the output and expected signal at a given time [5,6].

AFs are capable of automatically adjusting the filtering structure in digital signal processing, whereas nonadaptive filters have static filter coefficients that result in fixed transfer functions. In many applications, adaptive coefficients are required for processing due to the lack of prior knowledge of the parameters to be operated, such as the characteristics of some noise signals. In such cases, AFs are typically utilized to adjust the filter coefficients and frequency responses with feedback. With the development and maturity of adaptive filtering technologies, AFs have become widely used as effective tools in various fields, including signal processing [7–12], control [13,14], and image processing [15,16]. This is due to the stronger adaptability and better filtering performance of AFs.

AFs can be categorized into two types based on their structure: linear adaptive filters (LAFs) and nonlinear adaptive filters. Nonlinear adaptive filters (NAFs), such as Voetlrra filters [17,18], kernel filters [19,20], and neural network-based AFs [21,22], have stronger signal-processing capabilities. Owing to their low computational complexity, LAFs are still widely used in most practical applications. The LAFs built with a linear combiner are designed for sequential learning [20]. They are equipped with a mechanism that enables the filter to adjust its free parameters automatically in response

DOI: 10.1201/9781003176114-1

to statistical variations in the environment in which it operates. This capability has led to a wide range of applications of AFs in diverse fields, such as adaptive equalization in communication receivers, adaptive noise cancelation in active noise control, adaptive beamforming in radar and sonar, system identification, and adaptive control.

1.1.1 AF Framework

AFs mainly involve three elements: filter structure, cost function, and optimization algorithm. In general, AFs rely on error-correction learning for their adaptive capability. A common filtering configuration is depicted in Figure 1.1, where a tapped-delay-line (transversal) is used as the filter for adaptation. The filter has a set of adjustable parameters (weights) denoted by the vector $\mathbf{w}(k-1)$, where k denotes discrete time. An input signal vector $\mathbf{u}(k)$, applied to the filter at time k, produces the actual response $y(k)$, which is compared with an externally supplied desired response $d(k)$ to produce the error signal $e(k)$. This error signal is, in turn, used to produce an adjustment to the parameter vector $\mathbf{w}(k-1)$ of the filter by an incremental amount denoted by $\Delta\mathbf{w}(k)$. The adjustment is made to minimize the cost function $J(\mathbf{w})$, which measures the difference between the actual and desired responses. The optimization algorithm determines the incremental adjustment $\Delta\mathbf{w}(k)$ that minimizes the cost function $J(\mathbf{w})$. Accordingly, the updated parameter vector of the filter can be expressed by

$$\mathbf{w}(k)=\mathbf{w}(k-1)+\Delta\mathbf{w}(k) \tag{1.1}$$

On the next iteration at time $k+1$, $\mathbf{w}(k)$ becomes the latest value of the parameter vector to be updated. The adaptive filtering process is continually repeated in this manner until the filter reaches a condition, whereafter the parameter adjustments become small enough to stop the adaptation. As is clear, the weights here embody the hypothesis in the definition of sequential learning. Overall, the filter structure, cost function, and optimization algorithm work together to enable AFs to adapt to changing input signals and achieve optimal filtering performance.

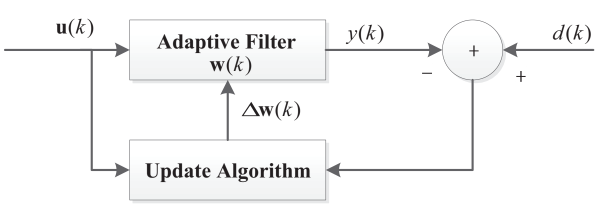


FIGURE 1.1
Block diagram of adaptive filters.

As previously mentioned, AFs utilize feedback to adjust filter coefficients and frequency responses, and the adaptive process involves update algorithms that use a cost function to determine how to change the filter coefficients to reduce the cost of the next iteration process. The adaptive algorithm generates a correction factor based on input and error signals, which is used to update the coefficients according to the defined cost function to obtain an estimation of the desired signal. The least mean square (LMS) [5] and recursive least square (RLS) [23] are two outstanding coefficient updating algorithms. The cost function is another key element and serves as the evaluation criterion for the best performance of the filter, such as the ability to reduce the noise component in the input signal.

Compared with other types of filters, the key to the better effect of AFs is the feedback structure of AFs. The adaptive process of adaptive filtering is to adjust the coefficient of FIR or IIR filter by adaptive algorithm (update algorithm) according to a suitable cost function (or adaptive learning criterion) so that the error signal is close to zero. In the following subchapters, the classical adaptive learning criteria and corresponding update algorithms are reviewed.

1.1.2 Adaptive Criteria

An essential aspect of designing AFs is the availability of an adaptive criterion. Traditional criteria for AFs include the least square (LS) criterion [24], minimum mean square error (MMSE) criterion [25,26], least absolute deviation (LAD) [27–29], and higher order error criteria [30–32]. The LS criterion is mathematically tractable and has a closed-form solution, defined by minimizing the sum of squared errors between observed and fitted values. Usually, a regularized version of the LS solution may be preferred [33]. Many AFs have been developed using the LS criterion, such as the RLS and its variants [34–37]. The MMSE criterion is commonly used as a measure of estimation quality in statistics and signal processing, which minimizes the mean square error (MSE) between the filter output and a desired variable. In AFs, the MMSE is often used as a cost function for stochastic approximation methods, which are a family of iterative algorithms that attempt to find the extrema of functions that cannot be computed directly and only estimated through noisy observations. The LMS algorithm [38–42], proposed in 1960 by Bernard Widow and Ted Hoff, is a typical stochastic gradient descent algorithm with MMSE criterion.

The LS and MSE criteria, which rely on the assumption that the error follows a Gaussian distribution, may be inadequate for practical data due to the likelihood of non-Gaussian interferences or outliers. To address this issue, robust regression methods are necessary to mitigate the bad effects of outliers on the estimation. In addition, research has shown that modeling expression data using heavy-tailed distributions [27], like the Laplacian distribution, can lead to more accurate results. Some authors have proposed a microarray normalization method that assumes errors follow a Laplacian model [43–45], using LAD regression as an optimization technique to compute normalization coefficients and avoid the effects of outliers in the original data. The LAD algorithm, also known as the sign-error LMS or pilot LMS [46–48], calculates the sum of absolute residuals $\sum_{i=1}^{k}|e(i)|$ and searches for the minimum value, making it more robust to anomalous points with large deviations in the data compared to the LMS algorithm, which may cause relatively large fluctuations after squaring.

Moreover, several studies have indicated that AFs based on higher order moments of the error signal can outperform those using MSE in certain critical applications [49,50]. Notably, the mean fourth error (MFE) criterion has been adopted as a cost function in adaptive filtering fields due to its convexity with respect to the weight vector. By stochastic gradient method, the least mean fourth (LMF) algorithm was developed to minimize the MFE and obtain the optimal weight [51]. Research has shown that the LMF algorithm can outperform the LMS algorithm in cases involving additive non-Gaussian noise, resulting in smaller excess MSE at the same convergence speed [31].

1.1.3 Typical Algorithms

1.1.3.1 Linear Adaptive Filtering Algorithms

In recent times, a plethora of adaptive filtering algorithms (AFAs) have emerged from diverse origins, each possessing unique characteristics. Researchers are particularly interested in AFAs that exhibit fast convergence, low computational complexity, and good numerical stability. Linear adaptive filters and their corresponding algorithms are more commonly employed in practical applications compared to NAFs due to their uncomplicated structure and low computational complexity. Notable algorithms under the MMSE criterion include the LMS, RLS, affine projection [52,53], and sub-band decomposition algorithms [54–56], among others, as documented in [6].

Over the past few decades, numerous AFAs have been proposed for various applications. Despite this, the LMS and RLS algorithms remain classical algorithms that have been thoroughly examined for optimality and stability. Since their inception, they have garnered significant attention, with research focusing on convergence analysis [57–62], performance enhancement [63–67], and the development of several LMS or RLS algorithms with unique methodologies, such as sparse AFAs [68–89], diffusion AFAs [90–107], constrained AFAs [108–115], and kernel AFAs [116–131]. These algorithms have been utilized for sparse system identification, distributed estimation, nonlinear time-series prediction, and other applications. In the following section, we will review these special algorithms.

1.1.3.1.1 Sparsity-Aware AFs

Sparsity-aware AFs have been developed for sparse system identification and can be categorized into two types: sparsity constraint AFAs [68–78] and proportionate AFAs [68–78]. Sparsity constraint AFAs are designed by integrating a sparsity constraint, such as an *lp*-norm constraint, into the cost function of the classical LMS or RLS algorithm [68–78]. The zero attraction term added to the update equation of the filter tap-weight vector aims to accelerate the identification speed by attracting small coefficients toward zero. However, the steady-state performance and instantaneous behavior of these algorithms depend on the selection of the zero attractor, which should be set according to the power of the measurement noise signal to ensure good steady-state mean square performance [70,71]. An adaptive strategy is proposed to select the zero attractor in the l_0-norm constraint LMS algorithm [74]. In addition to the zero attractor LMS-aware algorithms, l_1-norm regularized RLS adaptive algorithms have also been suggested in [75–78]. The SPARLS algorithm [75] presents an expectation-maximization approach for sparse system identification. The authors of [76] propose the application of an online coordinate descent algorithm together with the least-squares cost function penalized by an l_1-norm term. Another RLS algorithm for sparse system identification is proposed in [77], where the RLS cost function is regularized by adding a weighted norm of the current system estimate. Eksioglu further considers the regularization of the RLS cost function in a manner alike to the approach as outlined in [77]; the regularization term is defined as a general convex function of the system estimate, and an update algorithm is developed for the convex regularized RLS using results from subgradient calculus [77].

By expediting the elimination of inactive taps that correspond to system sparsity, sparsity-constrained AFs can deliver substantial performance improvements compared to their conventional counterparts, particularly during the steady-state phase. Nevertheless, there are also systems that are not strictly sparse but exhibit a relatively sparse (non-uniform) structure, where a small number of taps contribute to a significant portion of the energy [79]. In situations where sparsity is a crucial factor, proportionate-type algorithms have emerged as an important class of sparsity-aware AFAs. These algorithms employ the proportionate updating mechanism to update the filter coefficients. The pioneer in this field was Duttweiler, who introduced the proportionate normalized least mean square (PNLMS) algorithm [79]. This algorithm updates the filter coefficients by assigning a gain proportional to the magnitude of the current coefficient. The PNLMS algorithm outperforms the LMS and NLMS algorithms when applied to a sparse impulse response. However, its effectiveness diminishes when the impulse response is dispersive. To address this issue, several enhanced PNLMS algorithms have been proposed in the literature [80–82] to enhance the algorithm's resilience against time-varying sparsity. A different set of algorithms was developed by searching for a condition that would lead to the quickest overall convergence when all the coefficients approach their true values simultaneously. This approach gave rise to the μ-law PNLMS (MPNLMS) [83] and its variant [84]. The MPNLMS algorithm tackles the problem of assigning excessive update gain to large coefficients, which is a common issue with PNLMS algorithms. However, the convergence rate becomes unacceptably slow when dealing with correlated input conditions, such as speech. The Wavelet domain MPNLMS (WMPNLMS) [85] algorithm effectively tackles the problem of input decorrelation while preserving the sparsity of the impulse response. It achieves this by generating the conditional probability density function of the current weight deviations based on the preceding weight deviations, using a range of proportionate-type LMS algorithms [86]. Despite extensive research on the proportionate update mechanism in the context of NLMS-based AFs, the efficient design of sparse RLS algorithms using this mechanism remains an open issue. In a previous study [87], a natural recursive least squares (NRLS) algorithm was proposed, which utilized a proportionate matrix on the input vector to exploit system sparsity. However, this approach may render NRLS more sensitive to the condition number of the input covariance matrix than the standard RLS in certain scenarios [88]. In another study [89], a proportionate recursive least squares (PRLS) algorithm was introduced, which applies a proportionate matrix on the (Kalman) gain vector of the standard RLS.

1.1.3.1.2 Diffusion AFs

In the field of signal processing, distributed estimation has become a fundamental problem in recent years [90]. Typically, a group of nodes distributed across a geographical area work together to estimate an unknown model parameter based on linear measurements received by all nodes. There are three main methods for distributed estimation: incremental, consensus, and diffusion strategies. Among these, the diffusion strategy has been found to offer more advantages [90]. Diffusion-based algorithms are widely employed for distributed estimation, wherein neighboring nodes diffuse their estimates and measurements to adapt and combine their estimates. Among these algorithms, diffusion LMS (DLMS) is a fundamental method that utilizes the MSE criterion and diffusion structure [90–99]. Thanks to the exponentially weighted least squares (EWLS) criterion, the diffusion RLS (DRLS) algorithm has been enhanced to achieve rapid convergence, even for colored signals [100,101]. This algorithm aims to solve the network-wide LS estimation problem in a distributed adaptive manner, approaching the optimal LS without the need to transmit or invert any matrix. Mateos et al. introduced a novel distributed RLS algorithm for solving the EWLS problem using the alternating direction method of multipliers [102]. To minimize computation and communication expenses, this algorithm was further refined by censoring observations with small innovations, resulting in several variants [103–105]. To address the issue of biased estimation due to noisy input signals, various non-cooperative bias-eliminating algorithms have been proposed that utilize a bias-compensated mechanism [106,107].

1.1.3.1.3 Constrained AFs

Linearly constrained adaptive filtering (CAF) algorithms have gained significant attention and have been effectively utilized in various applications such as system identification, interference suppression, and array signal processing [108]. The primary advantage of CAFs is their ability to prevent error accumulation resulting from error correction, making them a preferred choice in these applications. Among the linearly constrained AFAs, the constrained LMS (CLMS) stands out as a simple stochastic gradient-based adaptive algorithm [109,110]. Initially designed as an adaptive solution to a linearly constrained minimum variance filtering problem in antenna array processing, CLMS has become a popular choice in various applications. Although the CLMS algorithm is simple and computationally efficient, it suffers from low convergence speed, particularly when the input signal is correlated. To address this issue, the constrained RLS (CRLS) algorithm was introduced in [111], albeit at the cost of higher computational complexity. Further improvements to the CRLS algorithm can be found in [112,113], while several constrained affine projection (CAP) algorithms were also developed in [114,115]. However, these constrained AFs with MSE criterion tend to perform poorly, especially when the signals involve non-Gaussian noises or outliers. This is mainly because the MSE criterion only captures the second-order moment.

In addition to classical LMS and RLS family AFs, non-MSE (or non-second-order moments)-based AFs have also demonstrated exceptional performance under certain conditions. For instance, the LMF and the LAD algorithms have shown remarkable results. The MFE criterion is a convex function (and thus unimodal) of the weight vector [51,132], which can outperform AFs with MSE for non-Gaussian additive noise, such as uniform and sinusoidal noise distributions. In such cases, the LMF algorithm has been found to yield smaller excess MSE for the same convergence speed. Various stability issues, tracking behaviors, and convergence analyses of the LMF algorithm have been explored in [49,50,133–138]. The Normalized LMF (NLMF) algorithm has been found to outperform the NLMS algorithm, particularly in low SNR scenarios, resulting in better steady-state performance [138]. In recent times, there has been a surge of interest in sparse NLMF algorithms. These algorithms incorporate different sparse penalty functions, such as zero-attracting (ZA), reweighted zero-attracting (RZA), reweighted l_1-norm, l_p-norm, and l_0-norm, leading to the development of various sparse NLMF-type algorithms [139–144]. Moreover, there are proportionate-aware LMF algorithms available to estimate the parameters of a sparse system with precision [145–147]. |In addition, ref. [148] has introduced a diffusion LMF (DLMF) algorithm that utilizes the diffusion strategy to improve the performance of distributed estimation in strong, non-Gaussian noise environments. To strike a balance between fast convergence rate and low steady-state misalignment, a variable step-size method has been incorporated into the DLMF. Furthermore, a sparse diffusion LMF algorithm has been proposed for estimating sparse parameters in Gaussian mixture noise environments [149]. The behavior of the DLMF algorithm has been analyzed in terms of mean and mean square in [150].

Overall, AFAs utilizing the MSE and MFE criteria have demonstrated exceptional performance in the realm of adaptive signal processing. Nevertheless, their convergence capabilities may be compromised when confronted with measurement noise that contains impulsive interferences. Accordingly, to combat impulsive interferences [151–155], the LAD criterion, which is based on l_1-norm minimization, has been proposed. The algorithms that incorporate LAD are referred to as least absolute deviation [151,152] or sign AFAs. In recent times, there has been extensive research on the steady-state and tracking analysis of signed-aware AFAs under various assumption conditions [156–158]. Considering the robustness of the signed aware algorithms, the sparsity sign subband AF (SSAF) [159,160] and diffusion sign algorithms [161–164] have been found to be highly robust signed aware algorithms for sparse system identification and distributed estimation. These algorithms minimize the l_1-norm of the sub-band a posteriori error vector, making them effective in handling sparse data.

1.1.3.2 Nonlinear AFAs

Linear adaptive filters have gained popularity in practical applications due to their straightforward structure and low computational complexity. However, their limited signal-processing capacity has restricted their use in certain applications. Nonlinear adaptive filters, such as Volterra filters, neural network-based adaptive filters, and kernel adaptive filters (KAFs), have emerged as a promising research area in adaptive signal processing due to their robust signal-processing capabilities. This book primarily concentrates on the KAFs and neural networks with random weights (NNRW)-based nonlinear adaptive filters.

The KAF has garnered significant attention in the fields of machine learning and signal processing as a powerful tool for solving nonlinear problems [20]. By transforming input data into higher or even infinite-dimensional reproducing kernel Hilbert spaces (RKHS), KAFs based on the conventional linear framework in RKHS have been extensively researched to address a wide range of nonlinear applications, including pattern classification, system identification, time-series prediction, and channel equalization. The Kernel Recursive Least-Squares (KRLS) algorithm, which can be considered as the RLS algorithm in RKHS, was initially developed in [116]. Several variants of KRLS have been proposed in a sequential manner, including sliding-window KRLS, extended KRLS algorithms, sparse KRLS, and quantized KRLS, as documented in [117–121]. Liu et al. further developed LMS algorithm in RKHS, called kernel least-mean-square (KLMS) algorithm [122]. Moreover, its theoretical convergence behavior was analyzed and derived because of its inherent simplicity and robustness in [123] and [124]. To reduce the computational complexity of the KLMS, the quantized KLMS (QKLMS) [125] and KLMS with promoting sparsity strategy [126] were proposed by quantized and constrained growth method. Some improved versions of KLMS were presented in [127–131]. An overview of kernel adaptive filtering is referred to [20]. While traditional KAF algorithms are effective in minimizing the widely used MSE, they are primarily designed to handle Gaussian noises. Unfortunately, real-world environments often contain non-Gaussian noises, which can cause KAF algorithms to become less robust. This is because MSE only captures the second-order statistics of the error signal, leaving KAF algorithms vulnerable to the limitations of this approach.

Neural networks (NNs) have been extensively researched as effective NAFs for system identification and noise cancelation, as evidenced by numerous studies [21,165–168]. The NNRW with a non-iterative learning mechanism is a feedforward neural network that employs a random learning algorithm to select input layer parameters and obtain output layer parameters through non-iterative calculation, resulting in an exceptionally fast learning speed. According to the different network structures and the degree of randomness, the current mainstream research methods for NNRWs include Random Vector Functional Link (RVFL) networks, Extreme Learning Machine (ELM), and Broad Learning System (BLS). In this book, we focus on reviewing ELM and BLS. Both ELM and BLS share the common feature of random weight and bias from the input layer (or feature layer) to the middle layer, while the weight and bias from the middle layer to the input layer are obtained by seeking the pseudo-inverse. The key difference between BLS and ELM lies in whether the feature layer (or input layer) is connected to the output layer (BLS: yes, ELM: no) and whether the input layer directly inputs data or feature (ELM: data, BLS: feature).

The ELM [169] is a novel fast learning algorithm designed to train a single-layer feedforward network (SLFN) with hidden neuron weights that are randomly initialized and fixed. This approach differs significantly from traditional training algorithms, such as the back-propagation (BP) algorithm and its improved versions [170,171], which require the tuning of hidden neuron weights. ELM, on the other hand, offers fast learning speed [172], universal approximation capability [172,173], and a unified learning paradigm for regression and classification [173]. For online identification problems, data samples often arrive in a time-ordered sequence. To address this, Liang et al. [174] proposed the online sequential ELM (OS-ELM), which can learn data one-by-one or chunk-by-chunk with fixed or varying chunk sizes. In addition, several improvements have been proposed and successfully applied in various applications [175–179]. In ELM, OS-ELM and many variants of them, the MSE criterion is usually adopted to construct their cost functions. Since the MSE criterion only takes into account the second-order statistics, it makes sense in the signal processing with Gaussian assumption. Consequently, ELM suffers from two drawbacks: (1) MSE minimization learning can easily suffer from overfitting. The problem will be serious if the characteristics of the learned dataset can't be represented by the training data [180,181]. (2) ELM may perform poorly in the data under nonlinear and non-Gaussian situations, as it only captures the second-order statistics in the samples [182].

BLS [183,184] is a shallow neural network model that has emerged as a promising discriminative learning method. It has demonstrated the potential to outperform some deep neural network-based learning methods, including the multi-layer perceptron (MLP)-based method [185], deep belief networks (DBNs) [186], and stacked autoencoders (SAEs) [187]. To create a BLS, there are several essential steps that must be taken. First, the input data must be transformed into general mapped features using feature mappings. These generated mapping features are then connected by nonlinear activation functions to form the "enhancement nodes". The mapped features and the "enhancement nodes" are then sent together into the output layer, and the corresponding output weights are obtained through the use of pseudoinverse. One of the benefits of BLS is that all weights and biases of the hidden layer units can be randomly generated and remain unchanged. This means that only the weights between the hidden layer and the output layer need to be trained, which greatly simplifies the training process. Furthermore, in the event that new samples are introduced or the network requires expansion, a number of practical incremental learning algorithms have been developed to ensure that the system can be quickly remodeled without the need for a complete retraining process from the beginning [183]. As a result of these appealing features, BLS has garnered increasing attention [188–195] and has been successfully implemented in various applications, including image recognition, face recognition, and time-series prediction. In addition, several variants of BLS, such as fuzzy BLS [196], graph regularized BLS [197], recurrent BLS [198], and structured manifold BLS [199], have been developed from different perspectives.

The standard BLS algorithm employs the minimum mean square error (MMSE) criterion as its default optimization criterion for training the network output weights. However, like other MMSE-based methods mentioned earlier, it may experience a decline in performance in complex noise environments, particularly when the data are tainted by outliers.

1.2 AFAs under MMPE Criterion

While commonly used cost functions like LS and MSE are reliable in most practical situations and remain the go-to for adaptive filters, they do have limitations. For instance, they only capture second-order statistics in the data, which can be a poor approximation criterion in nonlinear and non-Gaussian scenarios, such as heavy-tail or finite-range distributions. To address this issue, researchers have explored non-MSE (nonquadratic) criteria, including mean p-power error (MPE) [200,201], maximum correntropy criterion [202–204], minimum error entropy criterion [205–207], and Huber criterion [208–210], among others. This book focuses on the MPE criterion which considers higher or lower order statistics and its application to adaptive filtering. Notably, the LS, MMSE, MAE, and MFE criteria can be viewed as special cases of the MPE.

1.2.1 MMPE Criterion

As a more general version of the MMSE approach, *lp*-norm minimization (also known as Minimum MPE or MMPE) has found widespread applications in various fields, including filter design, beamforming array, and deconvolution. In particular, when dealing with impulsive noise-contaminated signals, sinusoidal frequency estimation tends to favor *lp*-norm ($p = 1$) minimization [200,201,210]. Given the success of *lp*-norm minimization, there is growing interest in developing adaptive Finite Impulse Response (FIR) filter algorithms based on the MMPE criterion [184,194–198]. If we set $p = 2$, the generalized criterion becomes the conventional MSE criterion. However, for values of p other than 2, the MPE criterion may exhibit superior properties to the MSE criterion in certain circumstances. Notably, the MPE criterion reduces to the LAD criterion when $p = 1$, and the MFE criterion can be obtained by setting $p = 4$.

Pei and Tseng investigated the advantageous features of an adaptive FIR filter that utilizes the MMPE criterion [201]. Their findings demonstrated that the MMPE criterion outperforms the conventional MSE criterion in certain applications, provided that an appropriate value of p is selected. First, it is important to note that the optimum solution of the MPE function may outperform the Wiener solution of the MSE function. This is particularly relevant in system identification, where the MPE function may provide a solution that is closer to the true system parameters. Second, in cases where the optimal solution of the MPE function is the same as the Wiener solution of the MSE function for $p \neq 2$, the steepest descent algorithm based on the MPE criterion may exhibit superior performance, such as faster convergence speed, compared to the conventional Widrow-Hoff LMS algorithm. Third, when input signals or desired responses are corrupted by impulse noises, adaptive filters based on the MPE criterion with $p = 1$ may demonstrate stronger robustness than the LMS algorithm [156]. Furthermore, both analytical results and extensive simulations have shown that the new algorithms with $p = 3$ or $p = 4$ can perform better than the sign and LMS algorithms across a wide range of estimation scenarios.

1.2.2 MMPE Criterion based AFAs

As previously mentioned, the MMPE criterion serves as a useful cost function for designing various AFAs. This section will focus on summarizing MMPE-based AFAs, which can be broadly categorized into two types: linear and nonlinear. Figure 1.2 provides a detailed breakdown of this classification.

1.2.2.1 Linear AFAs under MMPE Criterion

1.2.2.1.1 Least Mean p-Power Error (LMP)

In [200,201], an adaptive FIR filter based on the MPE criterion is explored. This filter is a generalization of the instantaneous gradient descent algorithm for alpha-stable processes and is known as the least MPE (LMP) algorithm.

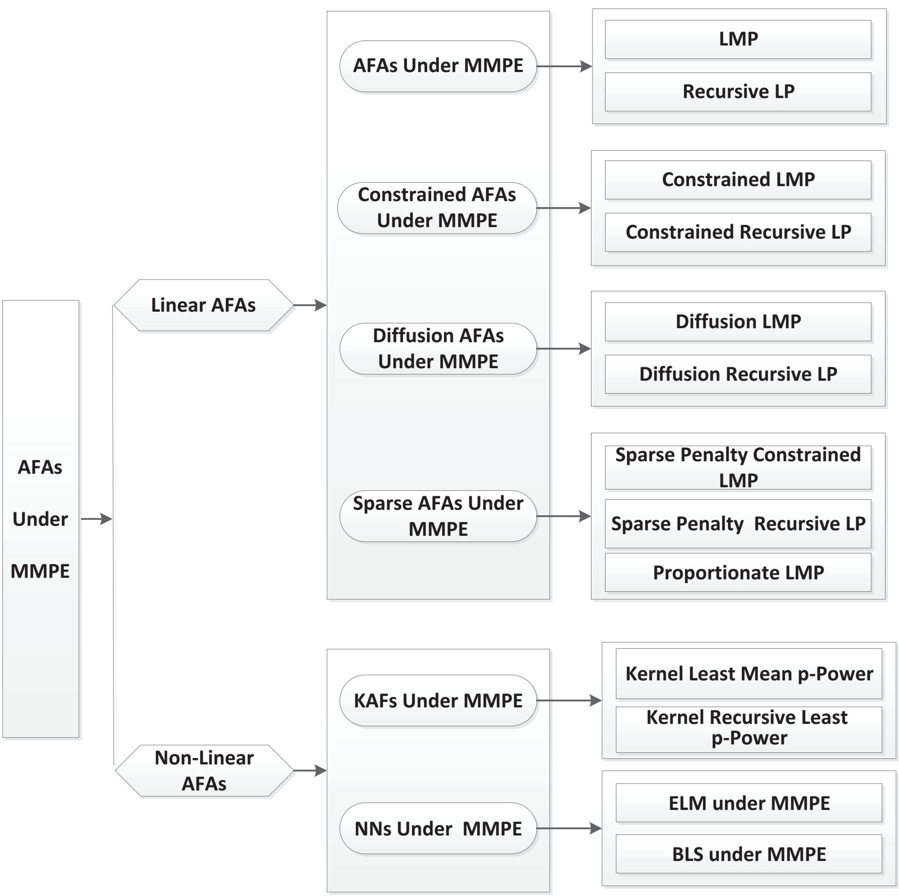


FIGURE 1.2
AFAs under MMPE.

When dealing with signals corrupted by impulsive noise, the LMP algorithm with $p = 1$ is the preferred choice. However, when the signal is affected by noise or interference, the adaptive algorithm with an appropriate selection of p may be more suitable [201]. To tackle the challenge of identifying nonlinear systems in impulsive noise environments, researchers have turned to the LMP algorithm to identify the Volterra kernels [211]. Their findings demonstrate that the cost function is convex in relation to the filter weights for $p \geq 1$. Through an approximation analysis, they have determined the convergence range for the step size of the LMP algorithm. In addition, the authors have explored the impact of p on performance and have discovered that the optimal performance is achieved when p is closest to the characteristic factor α of the alpha-stable process. Using Taylor series expansion, the steady-state mean-square error (MSE) was analyzed for both real and complex LMP algorithms [211]. In [212], the authors provided closed-form analytical expressions for the steady-state MSE, along with the corresponding restrictive conditions for step size. Inspired by the NLMS algorithm, a normalized LMP algorithm (NLMP) was developed that utilizes a normalization by dividing the update term by p-norm of the input vector [213]. A normalized LMAD algorithm can be achieved by setting $p = 1$ in the NLMP algorithm.

To enhance the robustness of the adaptive infinite impulse response (IIR) Notch filter (ANF), a new approach was proposed in [214], which utilizes the least MPE criterion. In addition, Maha [215] conducted a steady-state analysis of the constrained ANF with MPE. The findings indicate that the ANF with $p = 1$ outperforms the LMS algorithm in canceling 60-Hz interference in electrocardiogram recordings. Furthermore, when the ANF with MPE is employed to estimate the frequency of a sinusoid embedded in white noise, it exhibits superior statistical accuracy compared to the LMS algorithm, particularly when p is set at 3. The success of MMPE has sparked interest in designing IIR filters based on the MMPE criterion [216,217]. Tseng [218] has proposed a digital IIR filter with MPE that uses the reweighted method, allowing for an arbitrarily prescribed frequency response. In addition, Xiao et al. [219] developed an adaptive algorithm based on the least MPE criterion for Fourier analysis in the presence of additive noise. Analytical results and extensive simulations have shown that the proposed algorithm for $p = 3$ or $p = 4$ generates improved discrete Fourier coefficient estimates in moderate to high SNR, with similar degrees of complexity. In [220], the filtered-x LMP algorithm (FxLMP) was proposed, which minimizes a fractional lower order moment (p-power of error) that is applicable to stable distributions. It has been demonstrated that the FxLMP algorithm with $p < a$ exhibits superior robustness to ANC of impulsive noise. To enhance the convergence performance of the FxLMP algorithm, two modified versions were proposed in [221]. The first algorithm aims to improve the robustness of the FxLMP algorithm by utilizing modified reference and error signals. The second algorithm, known as normalized FxLMP (NFxLMP), extends the concept of the NLMP to the FxLMP algorithm.

1.2.2.1.2 Constrained LMP

Constrained adaptive filters (AFAs) have a wide range of potential applications in signal-processing domain. The primary objective is to solve a constrained optimization problem explicitly. Typically, the MSE criterion is used in constrained adaptive filters, like CLMS [222], due to its attractive features of mathematical tractability, convexity, and low computational complexity. However, the CLMS is also susceptible to non-Gaussian noise interference. To address this issue, some robust constrained AFs have been developed based on the maximum correntropy criterion (MCC) and MPE criterion [223–225]. Peng et al. proposed a constrained LMP algorithm [225] by combining an equality constraint with the MPE criterion, which can achieve much better performance, especially in the presence of impulsive noises with a proper p value.

1.2.2.1.3 Diffusion LMP

The emergence of wireless sensor networks has spurred the development of distributed adaptive estimation schemes. Among these, the Diffusion LMS- [226] and RLS [100]-type algorithms have garnered significant attention. However, these algorithms rely on the MSE criterion and are therefore not well-suited for non-Gaussian noise environments. To address this issue, a diffusion LMP algorithm [227] has been proposed for distributed estimation in alpha-stable noise environments, which are commonly encountered in various settings. Despite its effectiveness, the DLMP algorithm suffers from a slow convergence rate. To overcome this limitation, a diffusion normalized LMP algorithm has been developed [228], inspired by the concept of normalized algorithms. To further enhance the performance of the DNLMP algorithm, a robust DNLMP algorithm has been introduced, which takes into account the error signal in the normalization factor and can effectively mitigate outliers' influence in impulsive noise environments. In addition, researchers have developed the diffusion LMF and LAD algorithms as special cases.

1.2.2.1.4 Sparsity-Aware LMP

Sparsity-aware AFAs have gained widespread popularity for sparse system identification. Most of these algorithms, including sparse LMS with l_0-norm constraint, proportionate LMS, and their variants, utilize the MMSE criterion as the cost function, which makes them well-suited for Gaussian noise environments. However, in practical applications, noise often exhibits non-Gaussian properties, and the MMSE criterion may result in poor performance, particularly when the noise is impulsive (e.g., alpha-stable noise). To address this issue, researchers have explored robust algorithms, such as those presented in references [162,229–232]. In [229], two frameworks, namely RLS-type and NG-type algorithms, were proposed for designing AFAs that exploit channel sparseness and achieve robust performance against impulsive noises. In addition, an improved proportionate affine projection sign algorithm (RIP-APSA) based on the p-norm of the error signal was introduced in [162]. The sparsity penalty terms play a crucial role in enabling the filters to fit well with the sparse structure of the system. Therefore, the adaptive filter and the sparsity penalty are the two main components of a sparse adaptive filter. However, finding the sparsest solution, which leads to an l_0-norm minimization problem, is a NP-hard combinatorial optimization problem. To tackle this challenging issue, the l_0-norm is often approximated by continuous functions. In recent years, the correntropy induced metric (CIM) has been proven to be an excellent approximation of the l_0-norm in [202,233], which can achieve arbitrarily close results to the l0-norm under certain conditions. To address sparse system identification in impulsive noise environments, several sparsity-aware LMP algorithms with different sparsity penalty terms (l_1-norm, reweighted l_1-norm, and CIM) have been developed in [234]. In addition, Zhang et al. proposed a proportionate LMP algorithm [235] based on the proportionate scheme, which utilizes an adaptive gain matrix to adjust the step size of each tap according to a specific rule.

1.2.2.1.5 Recursive LMP

The algorithms mentioned above that are designed to be LMP aware suffer from slow convergence when dealing with colored input signals due to the inconvenient stochastic gradient method. To address this issue and accelerate convergence in such conditions, RLS-type algorithms are typically preferred. In addition, various approaches have been proposed to improve the robustness of RLS to alpha-stable noise. For instance, a sliding window LMP algorithm has been introduced for filtering alpha-stable noise [236]. Unlike previous stochastic gradient-type algorithms, this algorithm precisely minimizes the MPE within a sliding window of fixed size, also known as the recursive LMP (RLMP) algorithm. Therefore, the RLMP algorithm exhibits similar convergence speed and computational complexity to the RLS algorithm, as opposed to stochastic gradient-based algorithms that behave like the LMS algorithm. The RLMP algorithm, proposed in [236], utilizes a reweighted least squares algorithm that converges to the minimum of reweighted MPE. While this approach benefits from a truly robust cost function, both schemes encounter practical issues. Specifically, the approach in [236] is not truly online, as it processes all samples in a window of past inputs to the filter in batch mode, requiring multiple iterations of the algorithm at every time instant. Consequently, this increases the memory and computational requirements of the filter. In this correspondence, a novel solution to the recursive least p-norm problem was proposed by utilizing a combination of adaptive filters [237]. The use of adaptive filter combinations has gained significant traction in recent times as a straightforward yet effective approach to address the various tradeoffs that impact the performance of adaptive filters. These tradeoffs include the steady-state error versus the convergence and tracking performance tradeoff. Zhang et al. proposed an enhanced RLMP algorithm [238] to further improve its tracking performance. This algorithm utilizes an adaptive gain factor in the cross-correlation vector and the input signal auto-correlation matrix. In addition, it employs the square of the estimated impulsive-free first moment of the error signal to control the updated gain factor. To address the limitations of the CLMP algorithm, a constrained AFA called the constrained recursive least p-power (CRLP) algorithm was proposed [239]. This algorithm incorporates a set of linear constraints into the MMPE criterion to directly solve a constrained optimization problem.

1.2.2.2 Nonlinear AFAs under MMPE Criterion

1.2.2.2.1 Kernel LMP

Most KAFs rely on the MSE criterion, which is chosen for mathematical simplicity and convenience. However, to enhance the performance of KAFs in the presence of non-Gaussian or impulsive noise with low probability but high amplitudes, some new KAFs based on the MMPE criterion have been developed. The kernel least mean p-power (KLMP) algorithm was designed to deal with alpha-stable distribution noise [240,241]. The KLMP algorithm is rooted in the conventional KAF framework, and it employs the MMPE criterion to mitigate the bad impact of impulsive noise on KAF. In addition, Ma [241] and Gao [242] have introduced two distinct kernel recursive least mean p-power (KRLP) algorithms that outperform KLMP in terms of convergence speed and steady-state accuracy. To enhance the convergence rate of the KRLP, a random Fourier features extended KRLP algorithm was developed [243]. This algorithm is designed to handle non-Gaussian impulsive noise and offers significant improvements in convergence rate, steady-state EMSE, and tracking ability in the presence of impulsive interference. In addition, it reduces computational complexity by replacing the calculation of kernel function with kernel approximation. Another approach to improving KAFs is the sparsified kernel adaptive filters (SKAF), which includes the projected kernel least mean p-power algorithm (PKLMP) based on the MPE criterion and vector projection method [244]. To utilize the information contained in the desired outputs, a modified PKLMP algorithm has been developed by smoothing the desired signal. In addition, Huang et al. have introduced a robust kernel conjugate gradient least mean p-power (KCGLMP) algorithm that combines the conjugate gradient optimization method with kernel trick, resulting in improved filtering accuracy and computational efficiency [245]. To address the challenges posed by nonlinear and non-Gaussian environments commonly encountered in real-world scenarios, a diffusion approximated KLMP algorithm has been developed for nonlinear distributed systems [246]. This algorithm approximates the property of shift-invariant kernel function using random Fourier features.

1.2.2.2.2 Neural Networks with Random Weights

Neural networks provide an important approach to construct nonlinear adaptive filters. As mentioned in Section 1.1.3, neural networks with random weights (NNRW) are a type of feedforward neural networks that utilize a non-iterative learning mechanism. The ELM and BLS, as prominent examples of NNRW, rely on the MSE loss function, which is susceptible to non-Gaussian noise or outliers in the training data. To improve the robustness of these models, researchers have developed robust ELM and BLS models that employ the MPE loss function.

a. *ELM under MPE:* The neural network is a nonlinear adaptive filter, and the ELM with MSE model has gained significant attention due to its simple structure and fast learning speed [247,248]. However, traditional ELM performance may deteriorate in non-Gaussian scenarios, leading to the development of robust ELMs under MPE criterion, as reported in [249–251]. For instance, Yang et al. [249] proposed a least mean p-power ELM, which maintains the computationally simple ELM architecture while utilizing the MPE criterion to sequentially update the output weights. Real industrial processes often involve measurement samples with different statistical characteristics and are obtained one by one, making it challenging to achieve optimal learning performance for systems affected by various types of noise. To address this issue, the authors proposed an online sequential learning algorithm, known as recursive LMPELM, which is capable of designing an online ELM [250] that can provide accurate predictions of variables even in the presence of non-Gaussian noise. This approach outperforms both ELM and online sequential ELM, making it a promising solution for industrial applications. Moreover, a novel online sparse RLMP-ELM approach is introduced, which incorporates a sparsity penalty constraint on the output weights as a cost function, in addition to the MPE criterion [251].
b. *BLS under MPE:* Several alternative optimization criteria have been proposed to improve the robustness of the original BLS. These criteria combine l_1-norm with different regularization terms, resulting in a class of robust BLS (RBLS) variants [188]. By using l_1-norm, which is less sensitive to outliers, the robustness of BLS has been significantly enhanced. In addition, Chu et al. [252] introduced the weighted BLS (WBLS), which has demonstrated good robustness in a nonlinear industrial process due to its well-designed weighted penalty factor. Another notable

approach to improving the robustness of BLS is the robust manifold BLS (RM-BLS) [253]. Zheng et al. proposed a robust BLS model that replaces the l_2-norm-based optimization model in BLS with a mixed-norm-based one. This model has been used to design a powerful classifier with strong generalization capability for brain computer interface (BCI) research [254]. Furthermore, Zheng [255] has developed a least p-norm-based BLS (LP-BLS) that utilizes the p-norm of the error vector as a cost function and incorporates a fixed-point iteration strategy. The LP-BLS approach allows for flexible adjustment of the value of p ($p \geq 1$) to effectively handle interferences from various types of noise, thereby improving the modeling of unknown data. To further enhance the robustness of BLS, Zheng has also incorporated the MCC [202] to train the output weights, resulting in a correntropy-based BLS (C-BLS). The proposed C-BLS is expected to exhibit superior robustness to outliers while maintaining the original performance of the standard BLS in Gaussian or noise-free environments [256].

1.2.2.3 AFAs under KMPE Criterion

The MPE with p-th absolute moment of the error is a powerful tool for handling non-Gaussian data when a suitable p value is chosen. It is generally robust to large outliers when $p < 2$. Chen et al. introduced a novel non-second order measure, called the kernel MPE (KMPE), which is essentially the MPE in kernel space [257]. When $p = 2$, the KMPE reduces to the correntropy loss (C-Loss) [202], but with an appropriate p value, it can outperform the C-Loss when used as a cost function for robust learning. Drawing inspiration from KMPE, a novel measure called q-Gaussian kernel MPE (QKMPE) was proposed [258]. This measure is a generalization of the KMPE, defined with q-Gaussian kernel. In addition, a recursive kernel mean p-power is derived under the least QKMPE criterion for robust learning in noisy environments. This new algorithm has demonstrated superior performance against both Gaussian-type noise and non-Gaussian perturbations, particularly when the data contains large outliers. To further improve the performance of KMPE, a kernel mixture mean p-power error (KMP) criterion is proposed by combining the mixture of two Gaussian functions into the kernel function of KMPE [259]. The Nyström method is an efficient technique for controlling the growth of the network size of KAFs, and the recursive update form can enhance the tracking ability of KAFs. Finally, a recursive AFA is developed using KMPE with a forgetting factor as the cost function [260].

1.3 Outline of the Book

So far, numerous remarkable works have been accomplished on AFAs utilizing the MMPE criterion. Despite the existence of several books on AFAs designed under the MSE criterion, to our knowledge and investigation, there is still no book that presents AFAs under the MMPE criterion. Therefore, this book aims to provide a comprehensive treatment of AFAs under MMPE, with a focus on their properties, as well as linear and nonlinear AFAs. The remaining chapters of the book are organized as follows:

Chapter 2 delves into classical AFAs, such as the least mean square (LMS), recursive least square (RLS), and kernel adaptive filtering algorithms (e.g., kernel LMS and kernel RLS). This chapter serves as a foundation for readers to grasp the fundamental concepts that will be applied in subsequent chapters.

Chapter 3 presents a comprehensive overview of the minimum mean p-power error (MMPE) criterion, including its definition and properties. The chapter also delves into the relationship between MMPE and other conventional learning criteria such as MSE, MAE, and MFE. Furthermore, the chapter highlights several improved MMPE criteria, such as smoothed MMPE, adaptive MMPE, mixture MMPE, and kernel MMPE. This chapter is crucial for readers seeking to gain a deeper understanding of the fundamental principles underlying the MMPE criterion.

Chapter 4 focuses on various linear adaptive filtering algorithms that operate under the minimum mean p-power error criterion. These algorithms include the least mean p-power (LMP), adaptive LMP, smoothed LMP, sparsity-aware LMP, diffusion-aware LMP, and constrained LMP algorithms.

Chapter 5 mainly introduces the recursive AFAs under minimum mean p-power error criterion algorithms, such as recursive least mean p-power (RLP) algorithm, enhanced RLMP, sparsity RLP, diffusion RLP, and constrained RLP algorithm.

Chapter 6 presents nonlinear adaptive filtering algorithms that operate under the MMPE criterion. Specifically, we provide an overview of the kernel adaptive filtering and shallow neural network model under MMPE, including the kernel least mean p-power, kernel recursive least p-power, ELM under MMPE, and BLS under MMPE.

Chapter 7 primarily focuses on introducing the definition of the mixture MMPE criterion, along with various adaptive filtering algorithms that operate under this criterion. These include sparsity-aware AFAs, diffusion AFA, and kernel adaptive filters, all of which are designed to work effectively under mixture MMPE.

Chapter 8 provides a comprehensive overview of various adaptive filtering algorithms that are evaluated under the kernel mean p-power error criterion (KMPE). These algorithms include recursive KMPE, kernel adaptive filters (KAFs) that are based on KMPE family criteria (such as q-Gaussian KMPE and kernel mixture MPE-based KAFs), and ELM under KMPE.

2

Adaptive Filtering Algorithms under MMSE Criterion

Adaptive filtering algorithms (AFAs) have been extensively utilized in various practical applications, and the development of novel AFAs with distinct features remains a prominent research area in the field of signal processing. Nevertheless, the majority of new AFAs are based on traditional algorithms under minimum mean square error (MMSE) as their research foundation. To facilitate a better understanding of AFAs under minimum mean p-power error (MMPE), this chapter primarily focuses on reviewing some classical AFAs under MMSE, including least mean square (LMS), recursive least squares (RLS), kernel least mean square (KLMS), and kernel recursive least squares (KRLS).

2.1 LMS Algorithm

The traditional supervised adaptive filters rely on error-correction learning for their adaptive capability. To show the learning progress, the filtering structure depicted in Figure 2.1 is considered. The filter embodies a set of adjustable parameters (weights), which is denoted by the vector $\mathbf{w}(k-1)$, where k denotes discrete time instant, $\mathbf{u}(k)=[u(k),u(k-1),\ldots,u(k-M)]^T$ is an input signal vector applied to the filter at time k to produce the actual response $y(k)=\mathbf{u}^T(k)\mathbf{w}(k-1)$. This actual response is compared with an externally supplied desired response $d(k)$ to produce the error signal $e(k)=d(k)-y(k)$. This error signal is, in turn, used to produce an adjustment to the parameter vector $\mathbf{w}(k-1)$ of the filter by an incremental amount denoted by $\Delta\mathbf{w}(k)$. Accordingly, the updated parameter vector of the filter can be expressed by [261]

$$\mathbf{w}(k)=\mathbf{w}(k-1)-\Delta\mathbf{w}(k) \tag{2.1}$$

On the next iteration at time k, $\mathbf{w}(k)$ becomes the latest value of the parameter vector to be updated. The adaptive filtering process is continually repeated in this manner until the filter reaches a condition, whereafter the parameter

DOI: 10.1201/9781003176114-2

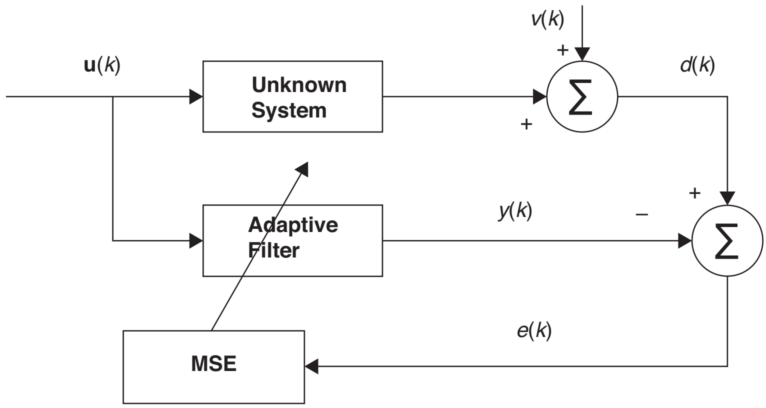


FIGURE 2.1
Basic structure of an adaptive filter.

adjustments become small enough to stop the adaptation. As is clear, the weights here embody the hypothesis in the definition of sequential learning.

Starting from some initial conditions denoted by $\mathbf{w}(0)$, the ensemble-averaged square error can be defined as:

$$J = E[e^2(k)], \qquad k = 1,2,\ldots \tag{2.2}$$

where $E[]$ is the expectation operator. The (2.2) is carried out for an ensemble of different training sets, which can trace the learning curve of the adaptive filtering process. J is a quadratic function of the weight $\mathbf{w}$, i.e.

$$J = \mathbf{w}\mathbf{R}_{\mathbf{uu}}\mathbf{w} - 2\mathbf{P}_{\mathbf{u}d}^T\mathbf{w} + \sigma_d^2 \tag{2.3}$$

where $\mathbf{R}_{\mathbf{uu}} = E(\mathbf{u}(k)\mathbf{u}^T(k))$ is the correlation matrix, $\mathbf{P}_{\mathbf{u}d} = E(\mathbf{u}(k)d(k))$ denotes the cross-correlation vector, and $\sigma_d^2 = E(d^2(k))$ is the variance of desired signals. Hence, the following expression from gradient vector can be obtained

$$\nabla_2 = 2\mathbf{R}_{\mathbf{uu}}\mathbf{w} - 2\mathbf{P}_{\mathbf{u}d} \tag{2.4}$$

Let $\nabla_2 = 0$, we get the unique optimum weight vector as

$$\mathbf{w}_o = \mathbf{R}_{\mathbf{uu}}^{-1}\mathbf{P}_{\mathbf{u}d} \tag{2.5}$$

Equation (2.5) is the Wiener solution.

In the design of adaptive filters, a crucial consideration is to ensure that the learning curve converges as the number of iterations increases. This requires defining the speed of adaptation, such that the ensemble-averaged square error reaches a relatively stable value, indicating that the adaptive filter has converged in the mean square error (MSE) sense.

The LMS algorithm is the most widely used and straightforward form of AFAs. Essentially, it operates by minimizing the instantaneous MSE cost function as

$$J(k) = \frac{1}{2} e^2(k) \tag{2.6}$$

where the factor 1/2 is introduced to simplify the mathematical formulation. Given that the parameter vector of the filter is $\mathbf{w}(k-1)$, the error signal $e(k)$ is defined by

$$e(k) = d(k) - \mathbf{w}(k-1)^T \mathbf{u}(k) \tag{2.7}$$

Correspondingly, the instantaneous gradient vector can be calculated by

$$\frac{\partial}{\partial \mathbf{w}(k-1)} J(k) = -e(k)\mathbf{u}(k) \tag{2.8}$$

Following the instantaneous version of the method of gradient descent, the adjustment $\Delta\mathbf{w}(k)$ applied to the algorithm at time k is defined by

$$\Delta\mathbf{w}(k) = \mu e(k)\mathbf{u}(k) \tag{2.9}$$

where μ is the step size parameter which controls the convergence speed of the LMS algorithm. Thus, using Eq. (2.9) in Eq. (2.1) yields the following update rule for the filter's parameter vector:

$$\mathbf{w}(k) = \mathbf{w}(k-1) + \mu e(k)\mathbf{u}(k) \tag{2.10}$$

The LMS algorithm is also known as the stochastic gradient algorithm, and its simplicity is highlighted in Table 2.1. To initialize the algorithm, it is common practice to set the weight vector's initial value to zero.

TABLE 2.1

Least Mean Square Algorithm

Initialization: w(0) = 0, μ
For k = 1, 2,…Do
1. $y(k) = \mathbf{u}^T(k)\mathbf{w}(k-1)$
2. $e(k) = d(k) - y(k)$
3. $\mathbf{w}(k) = \mathbf{w}(k-1) + \mu e(k)\mathbf{u}(k)$
End

Upon examining the computations described above, it becomes clear that the LMS algorithm is fundamentally simple. Despite its simplicity, this algorithm can deliver effective performance, provided that the step size parameter μ is appropriately selected. One of the most significant advantages of the LMS algorithm is its model independence, as it imposes no structural restrictions on how the training data were generated. As a result, the LMS algorithm is renowned for its robustness. To achieve optimal performance, it is recommended to assign a relatively small value to the step size parameter μ. However, from a practical standpoint, this approach has a significant drawback: a small step size causes the LMS algorithm to converge slowly.

The benefits of utilizing the LMS can be succinctly summarized as follows: (i) it boasts a low computational complexity; (ii) it is straightforward to implement; (iii) it enables real-time operation; and (iv) it does not require any statistical knowledge of signals, such as $\mathbf{R}_{\mathbf{uu}}$ and $\mathbf{P}_{\mathbf{ud}}$.

The convergence of the LMS adaptive filter is dependent on the auto-correlation matrix $\mathbf{R}_{\mathbf{uu}}$. To ensure that the system converges in the mean, two conditions must generally be met:

1. The auto-correlation matrix, $\mathbf{R}_{\mathbf{uu}}$, is positive definite.
2. $0 < \mu < 1/\lambda_{\max}$, where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{R}_{\mathbf{uu}}$.

Here a brief analysis of the convergence condition presented in (2) is only performed. For ease of analysis, it is assumed that $\mathbf{w}(k)$ is independent of $\mathbf{u}(k)$. Taking expectation on both sides of (2.10), we have

$$\begin{aligned} E[\mathbf{w}(k)] &= E[\mathbf{w}(k-1)] + \mu E[e(k)\mathbf{u}(k)] \\ &= E[\mathbf{w}(k-1)] + \mu E[d(k)\mathbf{u}(k) - \mathbf{u}(k)(\mathbf{u}^T(k)\mathbf{w}(k-1))] \\ &= E[\mathbf{w}(k-1)] + \mu \mathbf{P}_{\mathbf{u}d} - \mu \mathbf{R}_{\mathbf{uu}} E[\mathbf{w}(k-1)] \\ &= (I - \mu \mathbf{R}_{\mathbf{uu}}) E[\mathbf{w}(k-1)] + \mu \mathbf{P}_{\mathbf{u}d} \end{aligned} \tag{2.11}$$

Following the previous derivation, it will converge to the Wiener filter weights in the mean sense if

$$\begin{bmatrix} \lim\limits_{n\to\infty}(1-\mu\lambda_1)^n & 0 & \cdots & 0 \\ 0 & \lim\limits_{n\to\infty}(1-\mu\lambda_2)^n & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & & \lim\limits_{n\to\infty}(1-\mu\lambda_L)^n \end{bmatrix} = 0$$