



This is to certify that the

dissertation entitled

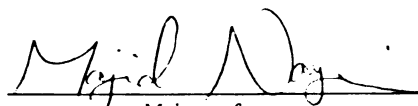
OPTIMAL BOUNDING ELLIPSOID ALGORITHMS
WITH AUTOMATIC BOUND ESTIMATION

presented by

Tsung Ming Lin

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Electrical Eng.


Major professor

Date Dec 20, 1996

LIBRARY
Michigan State
University

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.

DATE DUE	DATE DUE	DATE DUE
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____

**OPTIMAL BOUNDING ELLIPSOID ALGORITHMS
WITH
AUTOMATIC BOUND ESTIMATION**

By

Tsung-Ming Lin

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Electrical Engineering

1996

ABSTRACT

OPTIMAL BOUNDING ELLIPSOID ALGORITHMS WITH AUTOMATIC BOUND ESTIMATION

By

Tsung-Ming Lin

This research is concerned with new *Optimal Bounding Ellipsoid algorithms with Automatic Bound Estimation* (OBE-ABE) that can be applied to model parameter estimation, system identification, and related topics. To achieve convergence, conventional OBE algorithms require the *a priori* knowledge of the noise bounds which are unavailable in most real-world applications. It has been found that the exact knowledge of the noise bounds is essential to the performance of OBE algorithms. The new OBE-ABE algorithm, its computationally efficient version, Sub-OBE-ABE algorithm, and an adaptive version, Adaptive Sub-OBE-ABE algorithm, introduced in this dissertation do not require this knowledge. With the help of the automatic bound estimation (ABE) procedure, the new algorithms have excellent performance with respect to convergence, speed of convergence, computational efficiency, and tracking capability.

Another excellent feature of the new algorithms is the convergence in colored noise, or even non-stationary noise environments, which is theoretically impossible for other well-known algorithms, e.g., *Recursive Least Square* (RLS), *Least Mean Square* (LMS), and the *Kalman-Bucy Filter*. Due to these distinctive features, the new algorithms are expected to have wider applications than others.

Rigorous proofs for the *almost sure* convergence and convergence *in probability* of the new algorithms are provided for the cases of iid, mixing, ergodic, and non-stationary noises. Simulations in these noise cases are presented to support the proven convergence and to demonstrate other properties. The new algorithms are successfully applied to solve two real-world problems, the *linear prediction* analysis of speech and the blind-deconvolution problem.

**Copyright by
Tsung-Ming Lin
1996**

To SUMA CHING HAI

Acknowledgments

I am indebted to my thesis advisor, Professor Majid Nayeri, for his encouragement, patient guidance, and many hours spent in discussing with me. Thanks are also given to my co-advisor, Professor John R. Deller, Jr., for his generous literature support, valuable comments, and help in reviewing my dissertation.

I would also like to thank other members in my Ph.D. guidance committee, Professor Clifford Weil and Professor Hassan K. Khalil, for their help and encouragement in reviewing my dissertation.

I would like to express my gratitude to my mother and my wife for their love, sacrifice, and financial support.

Contents

List of Tables	ix
List of Figures	x
1 Introduction and Background	1
1.1 Introduction	1
1.2 Methods for Parameter Estimation	2
1.3 Review of OBE Algorithms	3
1.4 Motivation for New Algorithms	7
2 OBE Algorithms with Automatic Bound Estimation	11
2.1 Introduction	11
2.2 Formulation of OBE algorithms	11
2.3 OBE-ABE Algorithm	14
2.4 Sub-OBE-ABE Algorithm	14
2.5 Adaptive Sub-OBE-ABE	16
3 Convergence Analysis	20
3.1 Introduction	20
3.2 Definitions	21
3.3 Persistency of excitation condition	25
3.4 Lemmas	29
3.5 Almost Sure Convergence of OBE-ABE	35
3.6 Convergence in Probability of OBE-ABE	42
3.7 Almost Sure and Probability Convergence of Sub-OBE-ABE	50
3.8 A Necessary Condition	53
4 Simulation Studies	55
4.1 Introduction	55
4.2 IID Noise Cases	56
4.2.1 OBE-ABE vs. OBE	57
4.2.2 Sub-OBE-ABE vs. OBE-ABE	64
4.3 Colored Noise Cases	64
4.3.1 OBE-ABE vs. OBE	68
4.3.2 Sub-OBE-ABE vs. OBE-ABE	69
4.4 Tracking Performance of Adaptive Sub-OBE-ABE	74

5	Application to Linear Prediction Analysis of Speech	84
5.1	Introduction	84
5.2	LP model of speech	84
5.3	LP analysis using OBE-ABE	86
5.4	Sub-OBE-ABE and computational complexity	94
6	Application to Blind Deconvolution	99
6.1	Introduction	99
6.2	IID input and IIR filter	101
6.3	IID input and FIR filter	102
6.4	Colored Input and IIR Filter	102
7	Conclusions	109
7.1	Concluding Remarks	109
7.2	Contributions and further work	110
	Appendix A	112
	Bibliography	116

List of Tables

2.1	The OBE-ABE algorithm.	15
2.2	The Sub-OBE-ABE algorithm.	17
2.3	The Adaptive Sub-OBE-ABE algorithm.	19

List of Figures

1.1	Block diagram of a model for a dynamic system.	2
1.2	OBE algorithm with $\gamma_n = 0.8$ (underestimated) on model (1.3). $v_n \sim U(-1, 1)$	8
1.3	OBE algorithm with $\gamma_n = 1.2$ (overestimated) on model (1.3). $v_n \sim U(-1, 1)$	9
1.4	Volumes of ellipsoids (after 100 runs) of SM-SA with $\gamma_n = 1, 1.1, 1.5$ respectively, and OBE-ABE ($\gamma_0 = 1.5$) on model (1.3). $v_n \sim U(-1, 1)$	9
4.1	Estimators of $a_{1*} = 2$. CASE 1 : $v_n \sim B(-1, 1)$ is non-zero mean for model (4.1).	59
4.2	CASE 1 : $v_n \sim B(-1, 1)$ is non-zero mean for model (4.1).	59
4.3	CASE 1 : $v_n \sim B(-1, 1)$ is non-zero mean for model (4.1).	60
4.4	Estimators of $a_{1*} = 2$. CASE 2 : $v_n \sim U(-1, 1)$ for model (4.1).	60
4.5	CASE 2 : $v_n \sim U(-1, 1)$ for model (4.1).	61
4.6	CASE 2 : $v_n \sim U(-1, 1)$ for model (4.1).	61
4.7	Estimators of $a_{1*} = 2$. CASE 3 : $v_n \sim B(-0.5, 1)$, (4.3), for model (4.1).	62
4.8	CASE 3 : $v_n \sim B(-0.5, 1)$, (4.3), for model (4.1).	62
4.9	CASE 3 : $v_n \sim B(-0.5, 1)$, (4.3), for model (4.1).	63
4.10	CASE 4 : The final ellipsoid and the trajectory of the estimator. $v_n \sim B(-1, 1)$, (4.2), for model (4.4).	63
4.11	Estimators of $a_{1*} = 2$. CASE 1 : $v_n \sim B(-1, 1)$ is non-zero mean for model (4.1).	65
4.12	CASE 1 : $v_n \sim B(-1, 1)$ is non-zero mean for model (4.1).	65
4.13	Estimators of $a_{1*} = 2$. CASE 2 : $v_n \sim U(-1, 1)$ for model (4.1).	66
4.14	CASE 2 : $v_n \sim U(-1, 1)$ for model (4.1).	66
4.15	Estimators of $a_{1*} = 2$. CASE 5 : v_n is colored in (4.5) for model (4.1).	69
4.16	CASE 5 : v_n is colored in (4.5) for model (4.1).	70
4.17	CASE 5 : v_n is colored in (4.5) for model (4.1).	70
4.18	Estimators of $a_{1*} = -0.1$. CASE 6 : v_n is colored in (4.5) for model (4.6).	71
4.19	CASE 6 : v_n is colored in (4.5) for model (4.6).	71
4.20	CASE 6 : v_n is colored in (4.5) for model (4.6).	72
4.21	Estimators of $a_{1*} = 2$. CASE 7 : Colored noise $v_n \sim B(-0.5, 1)$, (4.7), for model (4.1).	72
4.22	CASE 7 : Colored noise $v_n \sim B(-0.5, 1)$, (4.7), for model (4.1).	73
4.23	CASE 7 : Colored noise $v_n \sim B(-0.5, 1)$, (4.7), for model (4.1).	73
4.24	CASE 8 : The final ellipsoid and the trajectory of the estimator. $v_n \sim B(-1, 1)$, (4.5), for model (4.4).	74
4.25	Estimators of $a_{1*} = 2$. CASE 5 : v_n is colored in (4.5) for model (4.1).	75
4.26	CASE 5 : v_n is colored in (4.5) for model (4.1).	75
4.27	Estimators of $a_{1*} = -0.1$. CASE 6 : v_n is colored in (4.5) for model (4.6).	76

4.28	CASE 6: v_n is colored in (4.5) for model (4.6).	76
4.29	Non-adaptive OBE algorithm on the time-varying system (4.8). $\gamma_n = 1.5$. 78	78
4.30	Non-adaptive OBE (SM-SA) algorithm on the time-varying system (4.8). $\gamma_n = 1.5$	79
4.31	Adaptive Sub-OBE-ABE algorithm on the time-varying system (4.8). $\rho = 4.5\%$	79
4.32	Adaptive Sub-OBE-ABE on the time-varying system (4.8). $\rho = 3\%$. 80	80
4.33	Adaptive Sub-OBE-ABE algorithm on the time-varying system (4.8). $\rho = 5.5\%$	80
4.34	Adaptive Sub-OBE-ABE on the time-varying system (4.8). $\rho = 3.5\%$. 81	81
4.35	Adaptive Sub-OBE-ABE on the time-varying system (4.8). $\rho = 7.5\%$. 81	81
4.36	Adaptive Sub-OBE-ABE on the time-varying system (4.8). $\rho = 5.5\%$. 82	82
4.37	Adaptive Sub-OBE-ABE on the time-varying system (4.8). $\rho = 10\%$. 82	82
4.38	Adaptive Sub-OBE-ABE on the time-varying system (4.8) ($v_n \sim B(-1, 1)$). $\rho = 31\%$	83
4.39	Adaptive Sub-OBE-ABE on the time-varying system (4.8) ($v_n \sim B(-1, 1)$). $\rho = 33\%$	83
5.40	Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and autocorrelation (solid line). . .	88
5.41	Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and autocorrelation (solid line). . .	88
5.42	Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and autocorrelation (solid line). . .	89
5.43	Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and autocorrelation (solid line). . .	89
5.44	Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and autocorrelation (solid line). . .	90
5.45	Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and autocorrelation (solid line). . .	90
5.46	Spectra of voiced /l/ phoneme. Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and RLS (solid line).	91
5.47	Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and RLS (solid line).	91
5.48	Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and RLS (solid line).	92
5.49	Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and RLS (solid line).	92
5.50	Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and RLS (solid line).	93
5.51	Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and RLS (solid line).	93
5.52	Spectra of simulated voiced /u/ phoneme. Upper graph: Direct 512- point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and autocorrelation method (solid line).	95
5.53	Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on Sub-OBE-ABE (dashed line), and autocorrelation (solid line). 95	95
5.54	Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on Sub-OBE-ABE (dashed line), and autocorrelation (solid line). 96	96

5.55	Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on Sub-OBE-ABE (dashed line), and autocorrelation (solid line).	96
5.56	Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on Sub-OBE-ABE (dashed line), and autocorrelation (solid line).	97
5.57	Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on Sub-OBE-ABE (dashed line), and autocorrelation (solid line).	97
5.58	Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on Sub-OBE-ABE (dashed line), and autocorrelation (solid line).	98
6.1	Model of the received signal in the blind-deconvolution problem. . .	100
6.2	Block diagram of the blind-deconvolution problem.	103
6.3	Estimates of a_{1*} of OBE ($\gamma_n = 1$) and OBE-ABE ($\gamma_0 = 10$) for Fig. 6.2 (iid-IIR, $S/N = 42$ dB).	103
6.4	Volumes of ellipsoids of OBE ($\gamma_n = 1$) and OBE-ABE ($\gamma_0 = 10$) for Fig. 6.2 (iid-IIR, $S/N = 42$ dB).	104
6.5	First 100 samples of $\tilde{v}_n$ (top), $\hat{v}_n$ (middle), and v_n (bottom) using OBE-ABE ($\gamma_0 = 10$) in Fig. 6.2 (iid-IIR, $S/N = 42.5$ dB).	104
6.6	Estimates of a_{1*} of OBE ($\gamma_n = 1$) and OBE-ABE ($\gamma_0 = 10$) for Fig. 6.2 (iid-IIR, $S/N = 36.5$ dB).	105
6.7	First 100 samples of $\tilde{v}_n$ (top), $\hat{v}_n$ (middle), and v_n (bottom) using OBE-ABE ($\gamma_0 = 10$) in Fig. 6.2 (iid-IIR, $S/N = 36.5$ dB).	105
6.8	First 100 samples of $\tilde{v}_n$ (top), $\hat{v}_n$ (middle), and v_n (bottom) using OBE-ABE ($\gamma_0 = 10$) in Fig. 6.2 (iid-FIR, $S/N = 42.5$ dB).	106
6.9	First 100 samples of $\tilde{v}_n$ (top), $\hat{v}_n$ (middle), and v_n (bottom) using OBE-ABE ($\gamma_0 = 10$) in Fig. 6.2 (IID-FIR, $S/N = 36.5$ dB).	106
6.10	First 100 samples of $\tilde{v}_n$ (top), $\hat{v}_n$ (middle), and v_n (bottom) using OBE-ABE ($\gamma_0 = 10$) in Fig. 6.2 (Colored-IIR, $S/N = 36.5$ dB). . . .	107
6.11	First 100 samples of $\tilde{v}_n$ (top), $\hat{v}_n$ (middle), and v_n (bottom) using OBE-ABE ($\gamma_0 = 10$) in Fig. 6.2 (Colored-IIR, $S/N = 22.5$ dB). . . .	108

Chapter 1

Introduction and Background

1.1 Introduction

From engineering to social science, people use modeling to help understand and describe observed phenomena. For example, in economics, modeling is usually used to predict inflation and study the trend of stock markets. In engineering fields, a dynamic system or signal is usually modeled before further analysis or processing. In general, a model can be described as in Fig. 1.1 in which the noise (or disturbance or model-error) sequence $\{v_n\}$ is unobservable.

Modeling is usually classified into two categories: Nonparametric and parametric [42]. A nonparametric model is described by a curve, table or function. Transient analysis and correlation analysis are two well-known methods for studying nonparametric models. A parametric model is usually described by difference equations (or differential equations in continuous time) with system parameters. Typical parametric models are the *auto regressive with exogenous input* (ARX), *auto regressive moving average with exogenous input* (ARMAX) and the state-space structure.

Basic laws from physics or science are usually employed to construct a parametric model. However, in many applications, the dynamic systems are too complex to model using only physics laws.

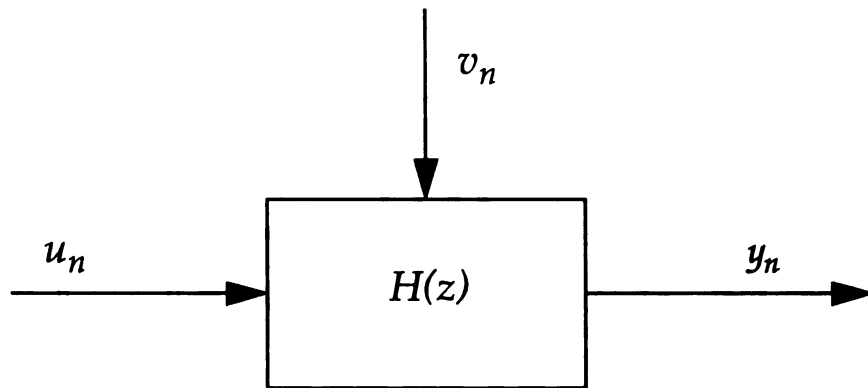


Figure 1.1: Block diagram of a model for a dynamic system.

Hence, numerous system-identification or parameter-estimation techniques have been developed to estimate the unknown parameters in a model. Some well-known methods are briefly reviewed in the next section.

The main objective of this research is to devise a new algorithm, *optimal bounding ellipsoid with automatic bound estimation* (OBE-ABE) algorithm and its variants for the estimation of model parameters. The new algorithms require less restrictive *a priori* knowledge and conditions than conventional methods for the convergence (or consistency) of the estimator. In addition, the performance of the new algorithms with respect to speed of convergence, computational efficiency, and tracking capability is superior to conventional methods.

1.2 Methods for Parameter Estimation

There are varieties of methods for estimating the model parameters. Among those are *minimum mean square error* (MMSE), *maximum likelihood* (ML), and *least square error* (LSE), etc. [22, 42]. They are all *batch* methods.

Well-known *recursive* (or real-time) methods are the *recursive least square* (RLS) [22], *least mean square* (LMS) [22], *instrumental variable* (IV) [42], *Kalman-Bucy filter* [42], and *optimal bounding ellipsoid* (OBE) [8, 11, 14, 20, 35] algorithms. One of the most popular methods in engineering applications is the LSE or its recursive counterpart, RLS, since they require only the whiteness of the noise sequence $\{v_n\}$ for convergence and have simple structure with well-understood convergence performance under regular conditions [i.e., *wide-sense stationarity* (WSS) and *second-moment ergodicity* (SME)]. RLS has been modified to a *weighted* RLS (WRLS) [22, 42] using a *forgetting factor* for tracking time-varying parameters.

Since its invention fifteen years ago, OBE has attracted attention from people in real-time signal processing [11, 15], system identification [3, 10], adaptive control [24], and other fields in which computational efficiency and speed of convergence are critical, and the statistics of the inputs and noise are unknown. However, its advantages (over those of popular RLS) with regard to computational efficiency and faster speed of convergence in some noise cases (see Chap. 4) are offset by the requirement of known noise bounds which are unavailable in most applications. Without *a priori* knowledge of the exact noise bounds, the performance of OBE suffers (Chap. 4). This is why OBE has been rarely found in real-world applications.

1.3 Review of OBE Algorithms

In 1968, Schweppe introduced the first *bounding ellipsoid* algorithm (BE) [41] for estimating the states of a state space model with the assumption of bounded inputs and bounded noises. With this assumption, BE algorithm produces at each step a feasible set in state space instead of a single point which is normally produced by other well-known methods, e.g., the Kalman-Bucy filter. The BE algorithm has a set of recursive formulae similar in structure to the Kalman-Bucy filter. Without *a priori*

knowledge of the statistics of the inputs and the noise which are usually required by other methods, BE has potential for wider applications. However, as pointed out by Schweppe, the feasible set of the algorithm at each step is usually impractical to calculate. He did not provide the convergence analysis of BE algorithm either.

Thereafter, Witsenhausen [47], Bertsekas and Rodes [4] published *set-membership* (SM) algorithms based on the same bounded-input and bounded-noise assumptions. Although SM has a smaller feasible set (convex polytope, see Chap. 2) at each step, the mathematical analysis of SM is more complicated than that of BE.

In 1979, Fogel published a BE algorithm [19] for the identification of the ARX model with the implicit assumption that the noise sequence $\{v_n\}$ has an asymptotically-*known* accumulative energy bound. With this knowledge and the assumptions of the white noise $\{v_n\}$, the regular WSS, and SME, Fogel has shown, in a deterministic way, the convergence of his BE algorithm. In other words, the sequence of the ellipsoids of BE algorithm is shown to asymptotically reduce to a point (the true model parameter vector). Although the recursive formula of BE is similar in structure to RLS, the computation is less efficient than RLS.

In 1982, Fogel and Huang published the first OBE algorithm (F/H OBE) [20] which featured *selective updating* to ignore “redundant” data (in the OBE sense). This innovative feature improves the computational efficiency of Fogel’s BE algorithm. However, F/H OBE algorithm still has $\mathcal{O}(m^2)$ computational complexity due to the check for selective updating at each step (Chap. 2). Selective updating is achieved by variably weighting the incoming data for an optimal (maximum reduction in size) ellipsoid at each step. The optimization is based on the assumption that the noise bounds are known pointwise. If incoming data set does not help shrink the ellipsoid, the weighting is set to zero, i.e. the incoming data set is “redundant” and hence discarded. In contrast to OBE, Fogel’s BE algorithm has constant data weights

(= 1). Detailed formulation of OBE algorithms can be found in the next chapter.

Fogel and Huang also provided a proof of convergence of OBE based on the assumption that the noise sequence is white and its pointwise upper bounds are known in addition to the regular WSS and SME. However, the correctness of the proof is arguable (also see [13, p.1913]) because the authors use in the proof the questionable assertion that convergence of Fogel's BE implies convergence of OBE. In fact, the whiteness of the noise is neither sufficient nor necessary for the convergence of the OBE algorithm. Theorem 3.4 and 3.6 in Chap. 3 validate this assertion.

In 1987, Dasgupta and Huang published another version of OBE (D/H OBE) [8] and showed the “convergence” of its estimator to a *non*-infinitesimal region of the true parameter vector under the same assumptions. An interesting and practically important feature of the algorithm is that it has an efficient $\mathcal{O}(m)$ check for innovation instead of $\mathcal{O}(m^2)$ in F/H OBE. However, comments are found in literature [14, 37] on the lack of interpretability of its optimization criterion [14], i.e., it minimizes, at each step, a quantity which is not directly related to the “size” of the ellipsoids.

Thereafter, other versions of OBE algorithm, e.g. *set-membership weighted recursive least square* (SM-WRLS) [11, 15, 37], and *set-membership stochastic-approximation* (SM-SA) [11, 27, 35], were developed by Nayeri, Deller, and their students. SM-WRLS is the first algorithm to clearly relate the OBE philosophy and conventional WRLS (see [11]). They also showed that all versions of OBE algorithm based on the same optimization criterion of minimizing the ellipsoid at each step, are basically the same algorithm [14]. Specifically, F/H OBE, SM-WRLS, and SM-SA, though different in their weighting policies and computational complexities, have identical estimators, ellipsoids, selected data, and convergence performance. A suboptimal check for innovation is introduced to SM-WRLS that results in a computationally efficient version of “interpretable” OBE algorithm [13, 37]. Further, using SM-SA, the first proof

of convergence *in probability* (p. convergence) [17] for OBE algorithms were accomplished under the assumptions of *independent identically distributed* (iid) noise and *symmetric* bounds.

The last OBE-like algorithm seeking to improve the performance of OBE in this review is *optimal volume ellipsoid* (OVE) algorithm proposed by Cheung *et al.* [7] in 1991. The algorithm uses an *affine transformation* [21] to get the smallest ellipsoid at each step. The simulations in [7] show that the ellipsoid at each step is smaller than those of OBE. However, the estimator of OVE, at the expense of computational cost, does not show significant improvement over those of OBE's (also see [27, p.6]). A proof of "convergence" of OVE similar to that of D/H OBE was provided in [7].

In 1991, Veres and Norton published a paper [44] presenting the first p. convergence (under very general conditions) of the *exact polytope algorithm* (EPA) which is an SM algorithm with a polytope feasible set at each step. Although they also provided a proof of *almost sure* (a.s.) convergence (or *strong* convergence in their notation with equivalent definition) for the EPA, the justification is arguable. What they have proved for a.s. convergence [on a probability space $(\Omega, \mathcal{F}, P)$] is:

$$P(\cup_{n=1}^{\infty} \{\omega : \|\theta_n - \theta_*\| < \epsilon\}) = P(\lim_{m \rightarrow \infty} (\cup_{n=1}^m \{\omega : \|\theta_n - \theta_*\| < \epsilon\})) = 1, \forall \epsilon > 0 \quad (1.1)$$

or, equivalently, after suppressing ω for brevity,

$$P(\cap_{n=1}^{\infty} (\|\theta_n - \theta_*\| > \epsilon)) = P(\lim_{m \rightarrow \infty} (\cap_{n=1}^m (\|\theta_n - \theta_*\| > \epsilon))) = 0, \forall \epsilon > 0 \quad (1.2)$$

which are not sufficient to imply a.s. convergence. This becomes evident upon comparing (1.1) with (3.3), or (1.2) with (3.6).

1.4 Motivation for New Algorithms

Despite their superiority over RLS in computational efficiency and speed of convergence in many noise cases, OBE algorithms have been rarely found in real-world applications. The main reason is that the assumption of the known noise bounds, while less restrictive (without the knowledge of statistics of noise, e.g., mean, variance, probabilistic distribution, etc.), is often impractical. Lack of proof of convergence for OBE in more realistic noise cases (than symmetric iid noise) is another (minor) reason.

Since the noise sequence $\{v_n\}$ is usually unobservable in practice, exact noise bounds are not available in most applications. Without the knowledge of exact noise bounds, the performance of convergence and the speed of convergence of OBE suffers. If one or more bounds are underestimated, i.e., $v_n^2 < \gamma_n$ at one or more n , then the algorithm is no longer theoretically valid. Simulations reveal that the underestimation of the noise bounds results in an *inconsistent* [17, 42] or diverging estimator.

A conservative sequence of bounds $\{\gamma_n\}$ (overestimated bounds) will assure a meaningful ellipsoid at each n . However, recent work of Nayeri *et al.* [27, 35] has demonstrated that the estimator may be very imprecise, even asymptotically, if the bounds are too “loose.”

To illustrate the effects of incorrectly estimated error bounds, an AR(3) model for Fig. 1.1 is constructed as follows:

$$y_n = \theta_*^T \mathbf{x}_n + v_n \quad (1.3)$$

in which $\mathbf{x}_n = [y_{n-1}, y_{n-2}, y_{n-3}]^T$ is a sequence of observable vectors, $\theta_* = [2, -1.48, 0.34]^T$ is the unknown parameter vector to be identified, and the unobservable noise sequence $\{v_n\}$ is uniformly distributed and iid with constant bound $\gamma_* = 1$. Figure 1.2 shows

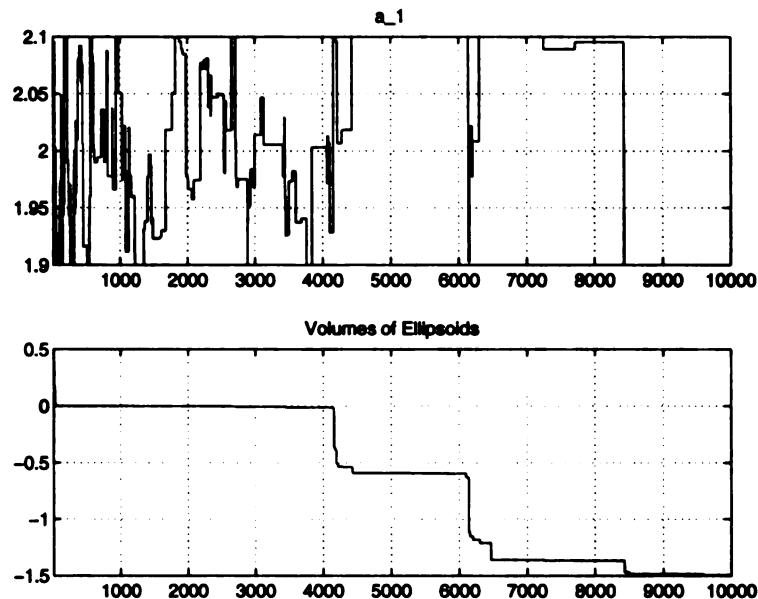


Figure 1.2: OBE algorithm with $\gamma_n = 0.8$ (underestimated) on model (1.3). $v_n \sim U(-1, 1)$.

the simulation result of OBE (SM-SA) algorithm (see Chap. 2 for details of the formulation) identifying this AR(3) model with under-estimated error bound $\gamma_n = 0.8$ for all n . As seen in the figure, the volume of the ellipsoid becomes negative quickly and the estimator diverges. This is the result of violating the underlying noise-bounding assumption of the algorithm. On the other hand, the simulation result of the algorithm using over-estimated error bound $\gamma_n = 1.2 \forall n$ is shown in Fig. 1.3. As seen in the figure, the estimator does not converge well to the true parameter ($a_1 = 2$ in the figures) even after 10,000 iterations.

The speed of convergence of OBE becomes slower if a looser overbounding sequence is employed. This is observed in Fig. 1.4 which shows the simulation results of 100 runs of the OBE algorithm on model (1.3). The numbers on the figure beside each dashed line are the estimated noise bounds.

In this dissertation, a new algorithm, OBE-ABE, and its variants are devised to improve the applicability of OBE algorithms while preserving or improving the superior performance of OBE algorithms in real-world applications. Proofs of a.s. conver-

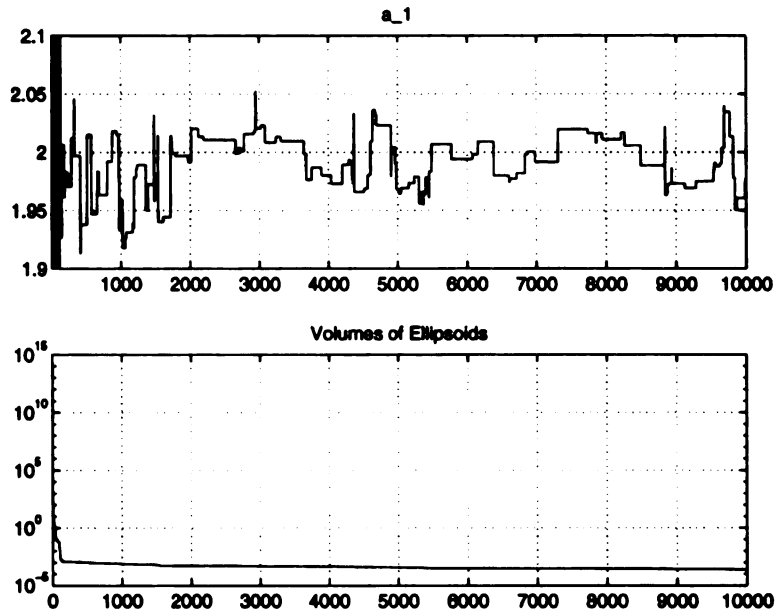


Figure 1.3: OBE algorithm with $\gamma_n = 1.2$ (overestimated) on model (1.3). $v_n \sim U(-1, 1)$.

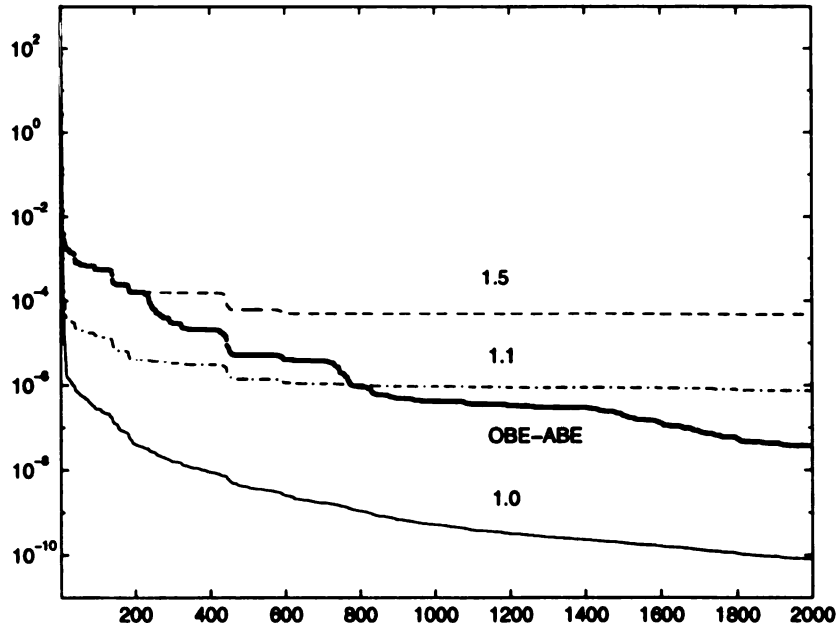


Figure 1.4: Volumes of ellipsoids (after 100 runs) of SM-SA with $\gamma_n = 1, 1.1, 1.5$ respectively, and OBE-ABE ($\gamma_0 = 1.5$) on model (1.3). $v_n \sim U(-1, 1)$.

gence and p. convergence of the new algorithms in the cases of iid, mixing, ergodic, and non-stationary noise with *asymmetric* bounds are provided. The new algorithms have efficient computation, excellent convergence performance and tracking capability. The details of the new algorithms are found in the subsequent chapters. A summary of the major contributions of this research is found in Chap. 7.

Chapter 2

OBE Algorithms with Automatic Bound Estimation

2.1 Introduction

In this chapter, the new algorithms, OBE-ABE, Sub-OBE-ABE, and the adaptive versions of these two algorithms are introduced. The algorithmic steps are listed in Tables 2.1 – 2.3. More details of these algorithms are found in subsequent chapters.

2.2 Formulation of OBE algorithms

A dynamic system or signal can often be modeled by a *linear-in-parameters* model

$$y_n = \theta_*^T \mathbf{x}_n + v_n \quad (2.1)$$

in which $\theta_* = [a_{1*}, \dots, a_{p*}, b_{0*}, b_{1*}, \dots, b_{q*}]^T$ is the unknown parameter vector to be identified; $\{\mathbf{x}_n\}$ is a sequence of observable vectors of dimension $m = p + q + 1$; and $\{v_n\}$ is an unobservable noise or model-error sequence. An important special case is the *auto-regressive with exogenous input* (ARX) [42] model in which $\mathbf{x}_n =$

$[y_{n-1}, \dots, y_{n-p}, u_n, \dots, u_{n-q}]^T$ is the observed data set composed of samples of the observable input sequence $\{u_n\}$ and output sequence $\{y_n\}$.

All versions of OBE algorithms [8, 11, 14, 20, 35] are based on the premise that v_n has a pointwise bound that is known *a priori*,

$$v_n^2 \leq \gamma_n, \quad \forall n \in \mathbb{N}. \quad (2.2)$$

Let A_i denote the parameter set at time i such that all elements in A_i are feasible parameter estimates consistent with (2.2). In conjunction with (2.1), it is clear that A_i is a (hyper-) strip region which can be expressed as

$$A_i = \left\{ \theta : \theta \in \mathbb{R}^m, (y_i - \theta^T \mathbf{x}_i)^2 \leq \gamma_i \right\}$$

As time i goes from 1 to n , the feasible set at time n , B_n , is the intersection of A_i , $\forall i \in [1, n]$, i.e.,

$$B_n = \bigcap_{i=1}^n A_i.$$

In general, B_n is a convex (hyper-) polytope in $\mathbb{R}^m$ which is mathematically difficult to track. Some SM algorithms are based on minimizing B_n at each step while OBE algorithms are based on minimizing a (hyper-) ellipsoid E_n [see (2.3)] which is a superset of B_n .

It can be shown [14] that the ellipsoid associated with any OBE algorithm is given by

$$E_n \stackrel{\text{def}}{=} \left\{ \theta : (\theta - \theta_n)^T \mathbf{P}_n^{-1} (\theta - \theta_n) < \kappa_n \right\} \quad (2.3)$$

where θ_n , κ_n and the matrix $\mathbf{P}_n$ are computed recursively using

$$G_n = \mathbf{x}_n^T \mathbf{P}_{n-1} \mathbf{x}_n \quad (2.4)$$

$$\varepsilon_n = y_n - \theta_{n-1}^T \mathbf{x}_n \quad (2.5)$$

$$\mathbf{P}_n = \frac{1}{\alpha_n} \left[\mathbf{P}_{n-1} - \frac{\beta_n \mathbf{P}_{n-1} \mathbf{x}_n \mathbf{x}_n^T \mathbf{P}_{n-1}}{\alpha_n + \beta_n G_n} \right] \quad (2.6)$$

$$\theta_n = \theta_{n-1} + \beta_n \mathbf{P}_n \mathbf{x}_n \varepsilon_n \quad (2.7)$$

$$\kappa_n = \alpha_n \kappa_{n-1} + \beta_n \gamma_n - \frac{\alpha_n \beta_n \varepsilon_n^2}{\alpha_n + \beta_n G_n}. \quad (2.8)$$

Recursions (2.4) – (2.8) comprise a general OBE algorithm. The ellipsoid center θ_n can be used as an estimator of the parameters θ_* at each n . The matrix $\mathbf{P}_n^{-1}$ satisfies the following recursion:

$$\mathbf{P}_n^{-1} = \alpha_n \mathbf{P}_{n-1}^{-1} + \beta_n \mathbf{x}_n \mathbf{x}_n^T. \quad (2.9)$$

The OBE algorithm is usually initialized with $\theta_0 = \mathbf{0}$, $\kappa_0 = \mu$ and $\mathbf{P}_0 = \frac{1}{\mu^2} \mathbf{I}$, where μ is a small number, typically 10^{-3} . In almost every OBE algorithm, the nonnegative data weights $\{\alpha_n\}$ and $\{\beta_n\}$ are calculated according to an optimization criterion which minimizes the “size” of the set E_n at each iteration¹. When such optimal weights do not exist, the updating need not take place. This prominent feature is called *selective updating*.

OBE algorithms are distinguished from one another by the choices of weighting sequences $\{\alpha_n\}$ and $\{\beta_n\}$ [14]. For example, the SM-SA algorithm has $\alpha_n = 1 - \lambda_n$ and $\beta_n = \lambda_n$ in which λ_n is given by

$$\lambda_n = \begin{cases} \frac{-b_n + \sqrt{b_n^2 - 4a_n c_n}}{2a_n}, & \text{if } c_n < 0 \\ 0, & \text{if } c_n \geq 0 \end{cases},$$

where

$$a_n = m\gamma_n - m\varepsilon_n^2 + mG_n^2\gamma_n - 2mG_n\gamma_n - \kappa_{n-1}G_n + \kappa_{n-1}G_n^2 + G_n\gamma_n - G_n^2\gamma_n - \varepsilon_n^2G_n$$

$$b_n = 2m\varepsilon_n^2 - 2m\gamma_n + 2mG_n\gamma_n + 2\kappa_{n-1}G_n - \kappa_{n-1}G_n^2 - G_n\gamma_n + \varepsilon_n^2G_n$$

¹The exception is the D/H OBE algorithm [8] in which κ_n is minimized at each iteration.

$$c_n = m\gamma_n - m\varepsilon_n^2 - \kappa_{n-1}G_n. \quad (2.10)$$

2.3 OBE-ABE Algorithm

While OBE algorithms require *a priori* knowledge of *exact* noise bounds to achieve convergence and fast speed of convergence (see Fig. 1.4 and Theorem 3.6 in Chap. 3), the new algorithm, OBE-ABE, does not require this knowledge. OBE-ABE theoretically requires a *lower* bound of tail probability but practically does not (see Remark 2 after Theorem 3.3). Initialized with any overestimated noise bound, OBE-ABE automatically (or blindly) updates its estimated noise bound γ_n to the unknown true bound γ_* , while concurrently shrinks the ellipsoids to get a consistent estimate. The OBE-ABE algorithm is found in Table 2.2. As seen in the table, the ABE procedure (below “otherwise”) features selective updating and is computationally negligible, hence preserves the computational efficiency of OBE. Further, the ABE procedure guarantees the convergence of the estimator, increases the speed of convergence, and increases the efficiency of updating the estimator (see Chap. 4). The convergence analysis of OBE-ABE algorithm is found in Chap 3.

2.4 Sub-OBE-ABE Algorithm

In optimal form, all versions of the OBE algorithm except D/H OBE have an $\mathcal{O}(m^2)$ check for innovation at each step. Hence, the computational complexity of OBE algorithms is similar to the popular RLS. To fully benefit from the selective updating feature and achieve superior computational superiority over RLS, a modification of the check for innovation to an $\mathcal{O}(m)$ operation is necessary. One approach is developed by Deller *et al.* [11, 13]. The modified OBE algorithm (Sub-OBE algorithm) in [11, 13] are $\mathcal{O}(\rho m^2)$ complexity with ρ indicating the fraction of data found to be innovative.

I. Initialization:

1. $\theta_0 = \mathbf{0}$, $\kappa_0 = \mu$ and $\mathbf{P}_0 = \frac{1}{\mu^2} \mathbf{I}$, where μ is a small number, typically 10^{-3} .
2. $\gamma_0 =$ any overestimated bound.
3. Choose ϵ (small positive number), M and N (see Remark 2 following Theorem 3.3).

II. Recursion:

For $n = 1 : N$

 If $c_n < 0$

 Execute recursions

$$\begin{aligned}
 G_n &= \mathbf{x}_n^T \mathbf{P}_{n-1} \mathbf{x}_n \\
 \varepsilon_n &= y_n - \theta_{n-1}^T \mathbf{x}_n \\
 \mathbf{P}_n &= \frac{1}{\alpha_n} \left[\mathbf{P}_{n-1} - \frac{\beta_n \mathbf{P}_{n-1} \mathbf{x}_n \mathbf{x}_n^T \mathbf{P}_{n-1}}{\alpha_n + \beta_n G_n} \right] \\
 \theta_n &= \theta_{n-1} + \beta_n \mathbf{P}_n \mathbf{x}_n \varepsilon_n \\
 \kappa_n &= \alpha_n \kappa_{n-1} + \beta_n \gamma_n - \frac{\alpha_n \beta_n \varepsilon_n^2}{\alpha_n + \beta_n G_n}.
 \end{aligned}$$

Otherwise,

 If a time interval I of length M over which $c_n \geq 0$ is found,

$$\gamma_n = \begin{cases} \gamma_{n-1} - d_J, & \text{if } d_J > 0 \\ \gamma_{n-1}, & \text{otherwise,} \end{cases}$$

where $d_J \stackrel{\text{def}}{=} \kappa_{J-1} G_J / m - \epsilon(2\sqrt{\gamma_{n-1}} - \epsilon)$ and $J = \arg \max_{n \in I} \varepsilon_n^2$.

Table 2.1: The OBE-ABE algorithm.

Hence, for small m (<10), the Sub-OBE algorithm is an $\mathcal{O}(m)$ algorithm since ρ is typically near 0.1 for a uniformly distributed $\{v_n\}$.

However, the same $\mathcal{O}(m)$ formula for checking of the Sub-OBE algorithm cannot be applied directly to the OBE-ABE algorithm, since the ABE procedure will fail [(3.61) no longer holds]. Hence, the Sub-OBE-ABE algorithm is endowed with another $\mathcal{O}(m)$ checking formula for innovation:

$$\bar{c}_n = m\gamma_n - m\varepsilon_n^2 - \kappa_{n-1}\mathbf{x}_n^T\mathbf{x}_n/d_{n-1},$$

where d_n has the recursion²:

$$d_n = \alpha_n d_{n-1} + \beta_n \mathbf{x}_n^T \mathbf{x}_n.$$

Justification of this check is found in Chap. 3. The steps of Sub-OBE-ABE algorithm are shown in Table 2.2. The Sub-OBE-ABE algorithm is compared to the OBE-ABE algorithm with regard to speed of convergence and computational complexity in Chap. 4.

2.5 Adaptive Sub-OBE-ABE

Tracking performance of adaptive OBE algorithms to time-varying parameters have been investigated in [11, 37, 38]. In those papers, adaptive OBE algorithms have been shown to have superior tracking capability to RLS, LMS, and their variants.

In [38], Rao and Huang modified D/H OBE algorithm by resetting κ_{n-1} , whenever $\kappa_n < 0$, to a value $\zeta + K_1$ or $\zeta + K_2$, where K_1 , and K_2 are two positive values obtained from the algorithm's parameters and data at time n , and ζ is a “safety fac-

²Preliminary version of Sub-OBE-ABE algorithm computes $1/d_n$ to be the minimum eigenvalue of $\mathbf{P}_n$ that features selective updating. Nayeri modifies it to this $\mathcal{O}(m)$ recursion.

I. Initialization:

1. $\theta_0 = \mathbf{0}$, $\kappa_0 = \mu$, $d_0 = \frac{1}{\mu^2}$, and $\mathbf{P}_0 = \frac{1}{\mu^2}\mathbf{I}$,
where μ is a small number, typically 10^{-3} .
2. $\gamma_0 =$ any overestimated bound.
3. Choose ϵ (small positive number), M and N (see Remark 2 following Theorem 3.3).

II. Recursion:

For $n = 1 : N$

 If $\bar{c}_n < 0$

 Execute recursions

$$\begin{aligned}
 G_n &= \mathbf{x}_n^T \mathbf{P}_{n-1} \mathbf{x}_n \\
 \varepsilon_n &= y_n - \theta_{n-1}^T \mathbf{x}_n \\
 \mathbf{P}_n &= \frac{1}{\alpha_n} \left[\mathbf{P}_{n-1} - \frac{\beta_n \mathbf{P}_{n-1} \mathbf{x}_n \mathbf{x}_n^T \mathbf{P}_{n-1}}{\alpha_n + \beta_n G_n} \right] \\
 \theta_n &= \theta_{n-1} + \beta_n \mathbf{P}_n \mathbf{x}_n \varepsilon_n \\
 \kappa_n &= \alpha_n \kappa_{n-1} + \beta_n \gamma_n - \frac{\alpha_n \beta_n \varepsilon_n^2}{\alpha_n + \beta_n G_n} \\
 d_n &= \alpha_n d_{n-1} + \beta_n \mathbf{x}_n^T \mathbf{x}_n.
 \end{aligned}$$

Otherwise,

 If a time interval I of length M over which $\bar{c}_n \geq 0$ is found,

$$\gamma_n = \begin{cases} \gamma_{n-1} - d_J, & \text{if } d_J > 0 \\ \gamma_{n-1}, & \text{otherwise,} \end{cases}$$

where $d_J \stackrel{\text{def}}{=} \kappa_{J-1} \mathbf{x}_J^T \mathbf{x}_J / (d_{J-1} m) - \epsilon(2\sqrt{\gamma_{n-1}} - \epsilon)$ and $J = \arg \max_{n \in I} \varepsilon_n^2$.

Table 2.2: The Sub-OBE-ABE algorithm.

tor” chosen to be 1. Equivalently, their adaptive OBE algorithm is obtained through modifying D/H OBE algorithm to include the resetting of κ_{n-1} to a value greater than 1 whenever $\kappa_n < 0$. Resetting κ_{n-1} is equivalent to expanding the ellipsoid.

In [11, 37], Deller *et al.* basically proposed three methods, *windowing*, *graceful forgetting* and *selective forgetting*, for modifying OBE algorithms to be adaptive to time-varying parameters. The basic spirit of those modifications is the various manipulations of forgetting factors (data weightings) similar to those of adaptive RLS (or WRLS). Among those methods, it is observed in [11, 37] that the best policy regarding tracking performance is the selective forgetting method which back-rotates previously accepted data sets when $\kappa_n < 0$ until $\kappa_n > 0$. This is equivalent to the expansion of the ellipsoid according to a systematic criterion.

In this dissertation, an *adaptive* Sub-OBE-ABE algorithm is presented by modifying Sub-OBE-ABE to include the resetting of $\kappa_n = 0.1$ whenever $\kappa_n < 0$. Despite its simplicity and computational efficiency, the adaptive Sub-OBE-ABE algorithm has shown excellent tracking performance in both slow and fast time-varying parameter cases (see Chap. 4). The steps of adaptive Sub-OBE-ABE algorithm are shown in Table 2.3.

I. Initialization:

1. $\theta_0 = \mathbf{0}$, $\kappa_0 = \mu$, $d_0 = \frac{1}{\mu^2}$, and $\mathbf{P}_0 = \frac{1}{\mu^2}\mathbf{I}$,
where μ is a small number, typically 10^{-3} .
2. $\gamma_0 =$ any overestimated bound.
3. Choose ϵ (small positive number), M and N (see Remark 2 following Theorem 3.3).

II. Recursion:

For $n = 1 : N$

 If $\bar{c}_n < 0$

 Execute recursions

$$\begin{aligned}
 G_n &= \mathbf{x}_n \mathbf{P}_{n-1} \mathbf{x}_n \\
 \varepsilon_n &= y_n - \theta_{n-1}^T \mathbf{x}_n \\
 \mathbf{P}_n &= \frac{1}{\alpha_n} \left[\mathbf{P}_{n-1} - \frac{\beta_n \mathbf{P}_{n-1} \mathbf{x}_n \mathbf{x}_n^T \mathbf{P}_{n-1}}{\alpha_n + \beta_n G_n} \right] \\
 \theta_n &= \theta_{n-1} + \beta_n \mathbf{P}_n \mathbf{x}_n \varepsilon_n \\
 \kappa_n &= \alpha_n \kappa_{n-1} + \beta_n \gamma_n - \frac{\alpha_n \beta_n \varepsilon_n^2}{\alpha_n + \beta_n G_n} \\
 d_n &= \alpha_n d_{n-1} + \beta_n \mathbf{x}_n^T \mathbf{x}_n.
 \end{aligned}$$

 If $\kappa_n < 0$, set $\kappa_n = 0.1$.

Otherwise,

 If a time interval I of length M over which $\bar{c}_n \geq 0$ is found,

$$\gamma_n = \begin{cases} \gamma_{n-1} - d_J, & \text{if } d_J > 0 \\ \gamma_{n-1}, & \text{otherwise,} \end{cases}$$

where $d_J \stackrel{\text{def}}{=} \kappa_{J-1} \mathbf{x}_J^T \mathbf{x}_J / (d_{J-1} m) - \epsilon(2\sqrt{\gamma_{n-1}} - \epsilon)$ and $J = \arg \max_{n \in I} \varepsilon_n^2$.

Table 2.3: The Adaptive Sub-OBE-ABE algorithm.

Chapter 3

Convergence Analysis

3.1 Introduction

This chapter provides proofs of a.s. convergence and p. convergence of OBE-ABE and Sub-OBE-ABE in the cases of iid, mixing, ergodic, and non-stationary noises. Throughout this chapter, the model considered is a stable ARX model

$$y_n = \theta_*^T \mathbf{x}_n + v_n \tag{3.1}$$

in which $\theta_* = [a_{1*}, \dots, a_{p*}, b_{0*}, b_{1*}, \dots, b_{q*}]$ is the unknown parameter vector to be identified; $\{\mathbf{x}_n\} = [y_{n-1}, \dots, y_{n-p}, u_n, \dots, u_{n-q}]$ is a sequence of observable data sets composed of the observable input sequence $\{u_n\}$ and output sequence $\{y_n\}$. The dimension of $\{\mathbf{x}_n\}$ is $m = p + q + 1$. The unobservable noise (or model-error) sequence $\{v_n\}$ is assumed to be *asymmetrically* bounded with *unknown* least upper bound $\sqrt{\gamma_*}$ or greatest lower bound $-\sqrt{\gamma_*}$ for all n . The results of this chapter are generalizable to cases in which y_n and v_n are vectors (e.g., [11, 37]).

For stochastic analysis, v_n , u_n , and y_n , etc., are modeled as random variables (r.v.) defined on a probability space $(\Omega, \mathcal{F}, P)$ where Ω is a sample space, $\mathcal{F}$ a σ -field, and

P a probability measure. Clearly, for the ARX model, if we let

$$\mathcal{F}_n \stackrel{\text{def}}{=} \sigma\{v_m, u_{m+1}, m \leq n\},$$

then $\mathbf{x}_n$ is $\mathcal{F}_{n-1}$ -measurable.

Note that the usual assumptions of WSS, SME, and whiteness of $\{v_n\}$ which are essential for the convergence of RLS, LMS, IV, etc., are not needed for OBE-ABE and its variants.

3.2 Definitions

For the concise statement and analysis of new theorems, the following definitions are introduced. Most definitions are cited from the literature. The first two below define the convergence of a random sequence to a random variable. These two types of convergence are treated in this dissertation.

Definition 3.1 [17, 42]. *An estimator θ_n is called p. consistent, or p. convergent (convergent in probability) if for all $\epsilon > 0$,*

$$\lim_{n \rightarrow \infty} P(\|\theta_n - \theta_*\| > \epsilon) = 0 \tag{3.2}$$

where $\|\cdot\|$ denotes any valid norm.

Note that p. convergence implies (is stronger than) *weak* convergence or, alternatively, convergence in *distribution*.

Definition 3.2 [17, 42]. *An estimator θ_n is called a.s. consistent, or a.s. convergent (convergent almost surely) if*

$$P(\lim_{n \rightarrow \infty} \theta_n = \theta_*) = 1. \tag{3.3}$$

Note that (3.3) is equivalent to

$$P(\|\theta_n - \theta_*\| > \epsilon \text{ i.o.}) = 0, \forall \epsilon > 0 \quad (3.4)$$

in which i.o. denotes *infinitely often* with respect to n .

Comparing (3.2) with (3.4), it follows that a.s. convergence implies p. convergence.

Equation (3.4) can be written as

$$P(\cup_{m=1}^{\infty} \cap_{n=m}^{\infty} (\|\theta_n - \theta_*\| > \epsilon)) = 0, \forall \epsilon > 0 \quad (3.5)$$

or equivalently,

$$P(\lim_{m \rightarrow \infty} \cap_{n=m}^{\infty} (\|\theta_n - \theta_*\| > \epsilon)) = 0 \forall \epsilon > 0. \quad (3.6)$$

Since the sets $\cap_{n=m}^{\infty} (\|\theta_n - \theta_*\| > \epsilon)$ is increasing with m , (3.6) can be written as

$$\lim_{m \rightarrow \infty} P(\cap_{n=m}^{\infty} (\|\theta_n - \theta_*\| > \epsilon)) = 0 \forall \epsilon > 0. \quad (3.7)$$

Although *surely* convergence [17] is theoretically stronger than a.s. convergence, the latter is regarded as the strongest type of convergence employed in practice (called *strong* convergence in some literature).

Definition 3.3 A random sequence $\{v_n\}$ is called asymptotically independent if

$$\lim_{n \rightarrow \infty} |P((v_i \in A) \cap (v_{i+n} \in B)) - P(v_i \in A) \cdot P(v_{i+n} \in B)| = 0, \quad \forall i \text{ and } \forall A, B \in \mathcal{F}.$$

For convenience, let us denote ϵ -neighborhoods of noise bounds as

$$D_{\epsilon}^+ = [\sqrt{\gamma_*} - \epsilon, \sqrt{\gamma_*}]$$

and

$$D_\epsilon^- = [-\sqrt{\gamma_*}, -\sqrt{\gamma_*} + \epsilon].$$

The following definition is seen in [44] which is a sufficient condition employed in their convergence analysis of the EPA algorithm.

Definition 3.4 [44]. *A random sequence $\{v_n\}$ is called uniformly conditionally heavily tailed (UCHT) if there exist $C_1 > C_2 > 0$ and an infinite subsequence $\{t_i\} \subset \mathbf{N}$, such that, with any sufficiently small $\epsilon > 0$,*

$$C_1 \epsilon \leq P(v_n \in (D_\epsilon^+ \cup D_\epsilon^-) | \mathcal{F}_{n-1}) \leq C_2 \epsilon \text{ a.s. } \forall n \in \{t_i\}. \quad (3.8)$$

Definition 3.5 *A random sequence $\{v_n\}$ is called uniformly conditionally tailed (UCT) if given $\epsilon > 0$, there exist a $\delta > 0$ and an infinite subsequence $\{t_i\} \subset \mathbf{N}$, such that*

$$P(v_n \in (D_\epsilon^+ \cup D_\epsilon^-) | \mathcal{F}_{n-1}) > \delta \text{ a.s. } \forall n \in \{t_i\}. \quad (3.9)$$

Note that, throughout the dissertation, the assumption that $\{v_n\}$ is asymmetrically bounded with the UCT condition means that either

$$P(v_n \in D_\epsilon^+ | \mathcal{F}_{n-1}) \geq \delta \text{ a.s.} \quad \text{and} \quad P(v_n \in D_\epsilon^- | \mathcal{F}_{n-1}) = 0 \text{ a.s.}$$

or

$$P(v_n \in D_\epsilon^+ | \mathcal{F}_{n-1}) = 0 \text{ a.s.} \quad \text{and} \quad P(v_n \in D_\epsilon^- | \mathcal{F}_{n-1}) \geq \delta \text{ a.s.}$$

holds for all n . Also, the assumption that $\{v_n\}$ is symmetrically bounded with UCT condition means that the following condition holds:

$$P(v_n \in D_\epsilon^+ | \mathcal{F}_{n-1}) \geq \delta_1 \text{ a.s.} \quad \text{and} \quad P(v_n \in D_\epsilon^- | \mathcal{F}_{n-1}) \geq \delta_2 \text{ a.s.}$$

Definition 3.6 A random sequence $\{v_n\}$ is called uniformly tailed (UT) if given $\epsilon > 0$, there exist a $\delta > 0$ and an infinite subsequence $\{t_i\} \subset \mathbb{N}$, such that

$$P(v_n \in (D_\epsilon^+ \cup D_\epsilon^-)) > \delta \text{ a.s. } \forall n \in \{t_i\}. \quad (3.10)$$

Note that UT and UCT conditions are less restrictive than UCHT. For example, “triangle” or “bounded-Gaussian” distributed iid sequence satisfies UCT and UT conditions but not UCHT.

The following definition of *almost periodic function* (a.p.) which is due to Bohr [6], and a lemma (Lemma A.3 in Appendix) which relates the type of random variable to an a.p. function are introduced for the proof of main theorems.

Definition 3.7 [1, 6]. A continuous function $\Phi(t)$ is almost periodic (a.p.) if given $\epsilon > 0$, there exists a length $l(\epsilon)$ such that every interval of length $l(\epsilon)$ contains at least a number τ , and $|\Phi(t + \tau) - \Phi(t)| < \epsilon$, $\forall t \in \mathbb{R}$.

The set of real numbers τ , $\{\tau\}_\epsilon$, defined above is called *relatively dense* [6]. Note that $l(\epsilon)$ exists if and only if there do not exist arbitrarily large gaps among the numbers τ . Clearly, a continuous *periodic* function is also a.p.

For the following definitions and some lemmas in the Appendix, let $K \stackrel{\text{def}}{=} \mathbb{N}$ or $\mathbb{Z}$,

$$\mathbb{R}^K \stackrel{\text{def}}{=} \{(\cdots, x_1, x_2, \cdots) : x_i \in \mathbb{R}, \forall i \in K\},$$

and

$$\mathcal{B}^K \stackrel{\text{def}}{=} \{(\cdots, A_1, A_2, \cdots) : A_i \in \mathcal{B}, \forall i \in K\}$$

where $\mathcal{B}$ is the Borel set.

Definition 3.8 [17]. $\{X_k : k \in \mathbb{K}\}$ is stationary if

$$(\cdots, X_1, X_2, \cdots) \stackrel{\mathcal{D}}{=} (\cdots, X_k, X_{k+1}, \cdots), \quad \forall k \in \mathbb{K}$$

where $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution.

Definition 3.9 [17]. The measurable transformation $T : \Omega \rightarrow \Omega$ is called a measure preserving transformation (m.p.t.) if $P \circ T^{-1} = P$, i.e., $P(\{\omega : T(\omega) \in A\}) = P(A)$.

Definition 3.10 [17]. A set $A \in \mathcal{F}$ is called invariant (or T -invariant) if $A = T^{-1}(A)$, i.e., $A = \{\omega : T(\omega) \in A\}$.

Definition 3.11 [17]. The invariant set $\mathcal{I} \stackrel{\text{def}}{=} \{A \in \mathcal{F} : T^{-1}(A) = A\}$ is called trivial if $P(A) \in \{0, 1\} \quad \forall A \in \mathcal{I}$.

Definition 3.12 [17]. The m.p.t. T is called ergodic if $\mathcal{I}$ is trivial.

Definition 3.13 [17]. The m.p.t. T is called mixing if

$$\lim_{n \rightarrow \infty} |P(A \cap T^{-n}(B)) - P(A) \cdot P(B)| = 0 \quad \forall A, B \in \mathcal{F}. \quad (3.11)$$

Note that the analogous definitions of mixing and ergodicity for a stationary sequence can be inferred from Definition 3.12, Definition 3.13, and Lemma A.5. By Definition 3.13, a stationary sequence is mixing if and only if it is asymptotically independent. Also, by Lemma A.6, mixing implies ergodicity.

3.3 Persistency of excitation condition

Persistency of excitation (PE) (defined below) is a necessary condition [5, 31, 42] for the convergence (consistency) of recursive algorithms such as LMS, RLS, and

their variants. For SM algorithms that assume bounded noise, PE is at least one of the sufficient conditions for convergence (see theorems in this chapter and [44]). Simulations even show [32] that PE is necessary for the consistency of the estimator of OBE algorithms.

The following definitions of PE that appear in literature are rewritten and related. The first definition is the most general one employed in proofs of the new theorems.

Definition 3.14 [44]. *A sequence of random vectors, $\{\mathbf{x}_n\}$, is called persistently exciting (PE)³ or omni-directional if for any nonsingular cone*

$$K = \{\mathbf{x} : \mathbf{x} = \alpha_1 \mathbf{e}_1 + \cdots + \alpha_m \mathbf{e}_m, \det(\mathbf{e}_1, \cdots, \mathbf{e}_m) \neq 0, \alpha_i > 0, \forall i\},$$

there exist ρ_1 and ρ_2 such that

$$\liminf_{n \rightarrow \infty} P(\mathbf{x}_n \in K \mid \mathcal{F}_{n-2}) \geq \rho_1 > 0 \text{ a.s.}, \quad (3.12)$$

and

$$E(\|\mathbf{x}_n\| \mid \mathcal{F}_{n-2}) \geq \rho_2 > 0 \text{ a.s. } \forall n. \quad (3.13)$$

The condition (3.12) means that the orientation of $\mathbf{x}_n$ is sufficiently varied in a conditional-probability sense while the condition (3.13) means that the magnitude of $\mathbf{x}_n$ cannot be too small on average. These conditions imply the following conditions which constitute another PE definition frequently used in the convergence analysis of conventional LSE recursive algorithms.

Definition 3.15 [31]. *A sequence of random vectors, $\{\mathbf{x}_n\}$, is called PE if there exist ρ_1 and ρ_2 such that*

$$\liminf_{n \rightarrow \infty} E(\mathbf{x}_n \mathbf{x}_n^T \mid \mathcal{F}_{n-2}) \geq \rho_1 \mathbf{I} > 0 \text{ a.s.} \quad (3.14)$$

³Please read “PE” as “persistency of excitation” or “persistently exciting” as appropriate in context.

and

$$E(\|\mathbf{x}_n\| \mid \mathcal{F}_{n-2}) \geq \rho_2 > 0 \text{ a.s. } \forall n. \quad (3.15)$$

Violation of (3.14) is effectively a.s. restriction of $\mathbf{x}_n$ to a proper subspace of $\mathbb{R}^m$. This justifies the assertion that (3.12) implies (3.14). Also note that the assumption of stability imposes an upper bound on (3.14).

The fact that

$$E(E(\mathbf{x}_n \mathbf{x}_n^T \mid \mathcal{F}_{n-2})) = E(\mathbf{x}_n \mathbf{x}_n^T),$$

leads to an equivalent definition of PE:

Definition 3.16 *A sequence of random vectors, $\{\mathbf{x}_n\}$, is called PE if there exist ρ_1 and ρ_2 such that*

$$\liminf_{n \rightarrow \infty} E(\mathbf{x}_n \mathbf{x}_n^T) \geq \rho_1 \mathbf{I} > 0 \text{ a.s.} \quad (3.16)$$

and

$$E(\|\mathbf{x}_n\| \mid \mathcal{F}_{n-2}) \geq \rho_2 > 0 \text{ a.s. } \forall n. \quad (3.17)$$

Condition (3.16) means that the autocorrelation matrix of $\mathbf{x}_n$ is positive definite asymptotically. For the stationary case, it becomes positive definite for all n as in condition (3.18) below.

Deterministic approaches to convergence analysis assume the stationarity and ergodicity (or at least WSS and SME) of the signals. In this case, Definition 3.16 is equivalent to the following.

Definition 3.17 [5]. *A stationary (or at least WSS) sequence of random vectors, $\{\mathbf{x}_n\}$, is called PE if there exist ρ_1 and ρ_2 such that*

$$E(\mathbf{x}_n \mathbf{x}_n^T) \geq \rho_1 \mathbf{I} > 0 \text{ a.s. } \forall n \quad (3.18)$$

and

$$E(\|\mathbf{x}_n\| \mid \mathcal{F}_{n-2}) \geq \rho_2 > 0 \text{ a.s. } \forall n. \quad (3.19)$$

By the Ergodic Theorem [17], Definition 3.17 is equivalent to the following.

Definition 3.18 [5, 42]. *A stationary and ergodic (or at least WSS and SME) sequence of random vectors, $\{\mathbf{x}_n\}$, is called PE if there exists an $N_1 \in \mathbb{N}$ and $\rho_1, \rho_2 > 0$ such that for all n*

$$\sum_{k=n+1}^{n+N_1} \mathbf{x}_k \mathbf{x}_k^T \geq \rho_1 \mathbf{I} > 0, \quad (3.20)$$

and

$$\|\mathbf{x}_n\| \geq \rho_2 > 0 \quad \forall n. \quad (3.21)$$

Note that the stability assumption imposes an upper bound on (3.20). Using matrix algebra, Definition 3.18 can be shown to be equivalent to the following.

Definition 3.19 [5]. *A stationary and ergodic (or at least WSS and SME) sequence of random vectors, $\{\mathbf{x}_n\}$, is called PE if for any unit vector $\mathbf{e} \in \mathbb{R}^m$, and for all n , there exist $N_1 \in \mathbb{N}$ and $\rho_1, \rho_2 > 0$, all independent of $\mathbf{e}$, such that*

$$\sum_{k=n+1}^{n+N_1} \|\mathbf{x}_k^T \mathbf{e}\| \geq \rho_1 > 0, \quad (3.22)$$

and

$$\|\mathbf{x}_n\| \geq \rho_2 > 0 \quad \forall n. \quad (3.23)$$

3.4 Lemmas

The following lemmas and those in the Appendix (cited from the literature) are essential for the proofs of the main theorems in this work. For convenience, define

$$c_n^- = \begin{cases} c_n, & \text{if } c_n < 0 \\ 0, & \text{otherwise.} \end{cases}$$

and

$$c_n^+ = \begin{cases} c_n, & \text{if } c_n \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

where c_n is the check for innovation (2.10). Then, by Lemma A.2, $\lim_{n \rightarrow \infty} c_n^- = 0$.

However, since c_n is a random variable,

$\lim_{n \rightarrow \infty} c_n^- = 0$ does not imply $\lim_{n \rightarrow \infty} P(c_n < 0) = 0$. For example, let

$$c_n = \begin{cases} 1, & \text{with probability 0.5} \\ -1/n, & \text{with probability 0.5.} \end{cases}$$

Then, $\lim_{n \rightarrow \infty} c_n^- = 0$ while $\lim_{n \rightarrow \infty} P(c_n < 0) = 0.5$. Hence, the following lemmas are needed for the proof of Lemma 3.3 which, in turn, is essential for the proofs of the main theorems.

Lemma 3.1 *Let X, Y be random variables. If X is continuously distributed, then $X + Y$ is continuously distributed. If, in addition, $Y \neq 0$, then XY is continuously distributed.*

Proof: Let $Z = X + Y$, then,

$$f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(x, z - x) dx = \int_{-\infty}^{\infty} f_X(x) f_{Y|X}(z - x|x) dx,$$

where f denotes the various probability density functions. By assumption, $f_X(x)$ does not contain a Dirac delta function, so $f_Z(z)$ does not. Hence, $X + Y$ is continuous.

Let $W = XY$, then,

$$f_W(w) = \int_{-\infty}^{\infty} \frac{1}{|y|} f_{XY} \left(\frac{w}{y}, y \right) dy = \int_{-\infty}^{\infty} \frac{1}{|y|} f_X \left(\frac{w}{y} \right) f_{Y|X} \left(y, \frac{w}{y} \right) dy.$$

Since $Y \neq 0$ and $f_X(\frac{w}{y})$ does not contain a Dirac delta function, it follows that $f_W(w)$ does not contain a Dirac delta function. Hence, XY is continuous. $\square$

Lemma 3.2 *Assume that model (3.1) is a stable ARX model, and $\{u_n\}$ and $\{v_n\}$ are bounded. If both $\{u_n\}$ and $\{v_n\}$ are asymptotically independent sequences, and $\{u_n\}$ is independent of $\{v_n\}$, then y_n converges to a continuous random variable as $n \rightarrow \infty$.*

Proof: First, rewrite (3.1) as

$$y_n - \sum_{j=1}^p a_j y_{n-j} = \sum_{j=0}^q b_j u_{n-j} + v_n. \quad (3.24)$$

Let

$$V_n \stackrel{\text{def}}{=} \sum_{j=0}^q b_j u_{n-j} + v_n.$$

Note that $\{V_n\}$ is an asymptotically independent random sequence. Let h_n be the discrete-time impulse response of the system (3.24), i.e., $y_n = h_n$, if V_n is a Kronecker delta function. Then,

$$y_n = \sum_{j=0}^n h_j V_{n-j}. \quad (3.25)$$

By the Lebesgue Decomposition Theorem [17], a random variable can be decomposed into a discrete part and a continuous part. However, by (3.25) and Lemma 3.1, y_n is continuous if V_n is continuous. Hence, V_n can be assumed discrete for the remainder

of the proof. That is, assume

$$P(V_n = q_k) = p_k > 0, \quad k = 1, 2, \dots, K \quad (3.26)$$

where q_k is bounded since V_n is bounded by assumption.

Let $\Phi_{V_n}(t)$ denote the characteristic function of V_n . Then, from (3.26),

$$\Phi_{V_n}(t) = \sum_{k=1}^K p_k e^{itq_k}, \quad \text{where } i = \sqrt{-1}.$$

Hence

$$\Phi_{h_j V_{n-j}}(t) = \sum_{k=1}^K p_k e^{it h_j q_k}. \quad (3.27)$$

Note that $e^{it h_j q_k}$ is a periodic function of t with primary period $\tau_j = 2\pi/h_j q_k$ for all j, k , in which $h_j \neq 0$ and $q_k \neq 0$. Since h_j is a stable infinite impulse response, it follows that $\lim_{j \rightarrow \infty} h_j = 0$. Hence, given $\epsilon_1 > 0$, there exists an N such that for all $j > N$, $|h_j| < \epsilon_1$ and there exists $j > N$ such that $|h_j| > 0$. Hence, $\tau_n = 2\pi/h_n q_k \rightarrow \infty$ as $n \rightarrow \infty$. This means that as $n \rightarrow \infty$, the set $\{\tau\}_\epsilon$ of $\Phi_{h_j V_{n-j}}(t)$ is not relatively dense. That is, there does not exist an $l(\epsilon)$ for $\Phi_{h_j V_{n-j}}(t)$ to satisfy Definition 3.7 as $n \rightarrow \infty$. Hence, $\Phi_{h_j V_{n-j}}(t)$ is a non-a.p. function as $n \rightarrow \infty$.

For finite n , Φ_{y_n} is a.p. by Lemma A.3 since y_n is discrete by (3.25). Hence, Φ_{y_n} is non-a.p. as $n \rightarrow \infty$ by the asymptotic independence of $\{V_n\}$. Thus, by Lemma A.3, y_n is continuous as $n \rightarrow \infty$. $\square$

Note that, by Lemma 3.1, if $\{u_n\}$ or $\{v_n\}$ or both are continuously distributed, the assumptions that $\{u_n\}$ and $\{v_n\}$ are asymptotically independent sequences and the independence between $\{u_n\}$ and $\{v_n\}$ are not required in the above lemma.

The following lemma asserts that if the noise bounds are overestimated, then the time between updates becomes infinite as $n \rightarrow \infty$.

Lemma 3.3 *Assume that model (3.1) is a stable ARX model, and $\{u_n\}$ and $\{v_n\}$ are bounded. If PE holds, both $\{u_n\}$ and $\{v_n\}$ are asymptotically independent sequences, $\{u_n\}$ is independent of $\{v_n\}$, and the noise bounds are overestimated for OBE algorithms, i.e., if there exists an $\epsilon_1 > 0$ and an $N \in \mathbb{N}$ such that for all $n > N$, $\gamma_n - v_n^2 > \epsilon_1$, then the expected updating interval of the estimator θ_n approaches infinity as n approaches infinity.*

Proof: Let $\tilde{\theta}_n = \theta_n - \theta_*$. From (2.2) and (2.5),

$$\gamma_n - \varepsilon_n^2 = \gamma_n - (y_n - \theta_{n-1}^T \mathbf{x}_n)^2 = \gamma_n - (v_n - \tilde{\theta}_{n-1}^T \mathbf{x}_n)^2$$

Hence, from (2.10),

$$c_n = m\gamma_n - m\varepsilon_n^2 - \kappa_{n-1}G_n = m\gamma_n - m(v_n - \tilde{\theta}_{n-1}^T \mathbf{x}_n)^2 - \kappa_{n-1}G_n. \quad (3.28)$$

Let

$$\phi_n \stackrel{\text{def}}{=} m(v_n - \tilde{\theta}_{n-1}^T \mathbf{x}_n)^2 + \kappa_{n-1}G_n \geq 0. \quad (3.29)$$

From (3.28),

$$c_n = m\gamma_n - \phi_n. \quad (3.30)$$

From Lemma A.2, it follows that

$$\liminf_{n \rightarrow \infty} c_n = \liminf_{n \rightarrow \infty} (c_n^+ + c_n^-) \geq 0.$$

This means that, for any $\epsilon > 0$, there exists an N such that for all $n > N$,

$$c_n > -m\epsilon. \quad (3.31)$$

It follows from (3.30) and (3.31) that, for all $n > N$,

$$\phi_n < m(\gamma_n + \epsilon). \quad (3.32)$$

From (3.30) and (3.32), it follows that

$$P(c_n < 0) = P(m\gamma_n < \phi_n < m(\gamma_n + \epsilon)) \quad \forall n > N. \quad (3.33)$$

This probability approaches zero as $\epsilon \rightarrow 0$, hence as $n \rightarrow \infty$. To prove this, first note that, as $n \rightarrow \infty$, y_n in (3.1) is continuous by Lemma 3.2. From (2.4),

$$G_n = \mathbf{x}_n^T \mathbf{P}_{n-1} \mathbf{x}_n = \sum_{i=1}^p \nu_{n,i} y_{n-i}^2 + \sum_{i=0}^q \nu_{n,i+p+1} u_{n-i}^2$$

where $\nu_{n,i} > 0$, $\forall i$, are the eigenvalues of the positive-definite matrix $\mathbf{P}_{n-1}$. Hence, by Lemma 3.1, G_n is continuous as $n \rightarrow \infty$. Also, from Lemma 3.7 (to follow), κ_n does not converge to zero since γ_n are overestimated $\forall n > N$ by assumption. Thus, it follows from Lemma 3.1 that $\kappa_{n-1} G_n$ is continuous as $n \rightarrow \infty$. Finally, from (3.29) and Lemma 3.1, ϕ_n is continuous as $n \rightarrow \infty$. This proves that the probability in (3.33) approaches zero as $n \rightarrow \infty$.

Since an OBE algorithm updates its estimator only when $c_n < 0$, the probability of an update diminishes as $n \rightarrow \infty$. This implies that the expected updating interval approaches infinity as time increases indefinitely. $\square$

Note that, if $\{u_n\}$ or $\{v_n\}$ or both are continuously distributed, the assumptions that $\{u_n\}$ and $\{v_n\}$ are asymptotically independent sequences and the independence between $\{u_n\}$ and $\{v_n\}$ are not required in the above lemma.

The following lemma asserts that the sum of two independent, ergodic, stationary sequences is stationary and ergodic.

Lemma 3.4 *If both $\{X_k : k \in \mathbb{K}\}$ and $\{Y_k : k \in \mathbb{K}\}$ are ergodic, stationary sequences, and $\{X_k : k \in \mathbb{K}\}$ is independent of $\{Y_k : k \in \mathbb{K}\}$, then $\{Z_k = X_k + Y_k : k \in \mathbb{K}\}$ is also stationary and ergodic.*

Proof: Since $\{X_n\}$ and $\{Y_n\}$ are stationary, by Definition 3.8, there exist sets A , B , C , and D such that

$$A = \{\omega : (\cdots, X_{-1}, X_0, X_1, \cdots) \in C\} \quad (3.34)$$

$$= \{\omega : (\cdots, X_{n-1}, X_n, X_{n+1}, \cdots) \in C\}, \quad (3.35)$$

and

$$B = \{\omega : (\cdots, Y_{-1}, Y_0, Y_1, \cdots) \in D\} \quad (3.36)$$

$$= \{\omega : (\cdots, Y_{n-1}, Y_n, Y_{n+1}, \cdots) \in D\}. \quad (3.37)$$

Hence, A and B are invariant sets by Definition 3.10. Also, by the ergodicity of $\{X_n\}$ and $\{Y_n\}$, Definition 3.11, and Definition 3.12,

$$P(A) \in \{0, 1\} \quad \text{and} \quad P(B) \in \{0, 1\}.$$

By (3.34) and (3.36), let F be such that

$$F = A \cap B = \{\omega : (\cdots, X_{-1}, X_0, X_1, \cdots) \in C \cap (\cdots, Y_{-1}, Y_0, Y_1, \cdots) \in D\}$$

Then, there exists an E such that

$$\begin{aligned} F &= \{\omega : (\cdots, (X_{-1} + Y_{-1}), (X_0 + Y_0), (X_1 + Y_1), \cdots) \in E\} \\ &= \{\omega : (\cdots, Z_{-1}, Z_0, Z_1, \cdots) \in E\}. \end{aligned} \quad (3.38)$$

Similarly, by (3.35) and (3.37),

$$F = \{\omega : (\cdots, Z_{n-1}, Z_n, Z_{n+1}, \cdots) \in E\}. \quad (3.39)$$

Thus, by (3.38), (3.39), and Definition 3.10, F is shift-invariant with respect to $\{Z_n\}$. Also, by the independence of $\{X_n\}$ and $\{Y_n\}$,

$$P(F) = P(A \cap B) = P(A) \cdot P(B) \in \{0, 1\}.$$

Hence, by Definition 3.11, F is trivial. Therefore, by Definition 3.12, $\{Z_n\}$ is stationary and ergodic. $\square$

3.5 Almost Sure Convergence of OBE-ABE

In this section, the theorems of a.s. convergence of the OBE-ABE algorithm are introduced and proven in iid, mixing, and ergodic noise cases. For a.s. convergence, both the input sequence $\{u_n\}$ and the noise sequence $\{v_n\}$ are assumed stationary, and $\{u_n\}$ is assumed independent of $\{v_n\}$. However, for p. convergence, the stationarity of $\{u_n\}$ and $\{v_n\}$ are not required. Further, if $\{u_n\}$ or $\{v_n\}$ or both are continuously distributed, the conditions for a.s. and p. convergence can be further relaxed. The following is the main theorem for a.s. convergence in the mixing case.

Theorem 3.1 *Assume that the stationary sequences $\{u_n\}$ and $\{v_n\}$ are independent. If PE holds, UCT holds, and $\{u_n\}$ and $\{v_n\}$ are mixing, then the estimator of the OBE-ABE algorithm is a.s. convergent.*

Proof: Since PE (3.12) and UCT (3.9) hold, there exists an infinite subsequence $\{t_i\} \in \mathbb{Z}$ such that for all $n \in \{t_i\}$, (3.9) and

$$P(\mathbf{x}_n \in K \mid \mathcal{F}_{n-2}) \geq \rho_1 > 0 \text{ a.s.}, \quad (3.40)$$

hold where K is defined in (3.12). Throughout the proof, all n considered are in $\{t_i\}$. For convenience, let $\{n\}$ replace $\{t_i\}$ as the time coordinate in the proof. The meaning of time intervals, for example, should change accordingly.

Let $\{v_k : k \in \mathbb{Z}\}$ denote the noise sequence and $\{u_k : k \in \mathbb{Z}\}$ the input sequence of model (3.1). Rewrite (3.1) as

$$y_n - \sum_{j=1}^p a_j y_{n-j} = \sum_{j=0}^q b_j u_{n-j} + v_n. \quad (3.41)$$

Let

$$w_n \stackrel{\text{def}}{=} \sum_{j=0}^q b_j u_{n-j}, \quad (3.42)$$

and $V_n \stackrel{\text{def}}{=} w_n + v_n$. Then, (3.41) becomes

$$y_n = \sum_{j=-\infty}^n h_{n-j} V_j \quad (3.43)$$

where h_n is the infinite impulse response, y_n , if V_n is set to a Kronecker delta function. Note that, by Lemma A.7, $\{w_n\}$ is stationary and ergodic. Also, by Lemma 3.4, $\{V_n\}$ is stationary and ergodic.

Let $\{I_k\}$ be a sequence of time intervals of length M over which the OBE-ABE algorithm does not update its estimator. Then, for each $k = 1, 2, \dots, K$,

$$\theta \stackrel{\text{def}}{=} [\theta(1) \ \theta(2) \ \dots \ \theta(m)]$$

is a constant vector for all n in each I_k . Hence, by (2.5),

$$\varepsilon_n = y_n - \theta_{n-1}^T \mathbf{x}_n = y_n - \sum_{i=1}^m \theta(i) y_{n-i} = \sum_{i=0}^m \theta(i) y_{n-i}, \quad (3.44)$$

where $\theta(0) \stackrel{\text{def}}{=} 1$. From (3.43) and (3.44), it follows that

$$\varepsilon_n = \sum_{i=0}^m \sum_{j=-\infty}^{n-i} \theta(i) h_{n-i-j} V_j.$$

By changing the order of summations, this equation becomes, for all $n > m$,

$$\varepsilon_n = \sum_{j=n-m+1}^n V_j \sum_{i=0}^{n-j} \theta(i) h_{n-i-j} + \sum_{j=-\infty}^{n-m} V_j \sum_{i=0}^m \theta(i) h_{n-i-j}.$$

Letting $p = n - j$,

$$\varepsilon_n = \sum_{p=0}^{m-1} V_{n-p} \sum_{i=0}^p \theta(i) h_{p-i} + \sum_{p=m}^{\infty} V_{n-p} \sum_{i=0}^m \theta(i) h_{p-i} \stackrel{\text{def}}{=} \sum_{p=0}^{m-1} V_{n-p} \cdot g_1 + \sum_{p=m}^{\infty} V_{n-p} \cdot g_2. \quad (3.45)$$

Note that both g_1 and g_2 are independent of n . Hence, (3.45) can be written as

$$\varepsilon_n = g(V_n, V_{n-1}, V_{n-2}, \dots) \quad (3.46)$$

where $g : \mathbb{R}^K \rightarrow \mathbb{R}$ is measurable and independent of n . Hence, by Lemma A.7, $\{\varepsilon_n\}$ is stationary and ergodic as $M \rightarrow \infty$.

Now, for each $k = 1, 2, \dots, K$, it follows from (2.10) that

$$c_n = m(\gamma_n - \varepsilon_n^2) - \kappa_{n-1} G_n \geq 0, \quad \forall n \in I_k. \quad (3.47)$$

Also from (2.5) and letting $\tilde{\theta}_n \stackrel{\text{def}}{=} \theta_n - \theta_*$,

$$\varepsilon_n^2 = (y_n - \theta_{n-1}^T \mathbf{x}_n)^2 = (v_n - \tilde{\theta}_{n-1}^T \mathbf{x}_n)^2. \quad (3.48)$$

Let A_n^- be the event that $v_n \in D_\epsilon^-$ and $\tilde{\theta}_{n-1}^T \mathbf{x}_n \geq 0$. Let A_n^+ be the event that $v_n \in D_\epsilon^+$ and $\tilde{\theta}_{n-1}^T \mathbf{x}_n < 0$. Also, let $A_n \stackrel{\text{def}}{=} A_n^+ \cup A_n^-$ and denote by A_n^c the complement of A_n . Also note that $\tilde{\theta}_{n-1}^T \mathbf{x}_n$ is $\mathcal{F}_{n-1}$ -measurable.

Let B_k denote the event that A_n occurs at least once in I_k . Then, for all $n \in I_{k+1}$, $B_k \in \mathcal{F}_{n-2}$. Hence, it follows from (3.40) that

$$P(\tilde{\theta}_{n-1}^T \mathbf{x}_n \geq 0 \mid B_k) \geq \rho_1 > \rho > 0 \quad (3.49)$$

and

$$P(\tilde{\theta}_{n-1}^T \mathbf{x}_n < 0 \mid B_k) \geq \rho_2 > \rho > 0. \quad (3.50)$$

Since $B_k \in \mathcal{F}_{n-2} \subset \mathcal{F}_{n-1}$ and $\tilde{\theta}_{n-1}^T \mathbf{x}_n$ is $\mathcal{F}_{n-1}$ -measurable, it follows from UCT (3.9) that

$$P(v_n \in (D_\epsilon^+ \cup D_\epsilon^-) \mid \tilde{\theta}_{n-1}^T \mathbf{x}_n \geq 0, B_k) \geq \delta_1 > \frac{\delta}{\rho} > 0 \quad (3.51)$$

and

$$P(v_n \in (D_\epsilon^+ \cup D_\epsilon^-) \mid \tilde{\theta}_{n-1}^T \mathbf{x}_n < 0, B_k) \geq \delta_2 > \frac{\delta}{\rho} > 0. \quad (3.52)$$

Since A_n^+ and A_n^- are mutually exclusive for sufficiently small ϵ , it follows from (3.49), (3.50), (3.51), (3.52), and the notes below (3.9) that

$$\begin{aligned} P(A_n \mid B_k) &= P(A_n^- \mid B_k) + P(A_n^+ \mid B_k) \\ &= P(v_n \in D_\epsilon^- \mid \tilde{\theta}_{n-1}^T \mathbf{x}_n \geq 0, B_k) \cdot P(\tilde{\theta}_{n-1}^T \mathbf{x}_n \geq 0 \mid B_k) \\ &\quad + P(v_n \in D_\epsilon^+ \mid \tilde{\theta}_{n-1}^T \mathbf{x}_n < 0, B_k) \cdot P(\tilde{\theta}_{n-1}^T \mathbf{x}_n < 0 \mid B_k) \\ &> P(v_n \in D_\epsilon^- \mid \tilde{\theta}_{n-1}^T \mathbf{x}_n \geq 0, B_k) \cdot \rho + P(v_n \in D_\epsilon^+ \mid \tilde{\theta}_{n-1}^T \mathbf{x}_n < 0, B_k) \cdot \rho \\ &> \delta > 0. \end{aligned} \quad (3.53)$$

Then, by (3.53) and the definition of A_n , it follows that

$$P(|\varepsilon_n| > \sqrt{\gamma_*} - \epsilon \mid B_k) \geq P(A_n \mid B_k) > \delta > 0. \quad (3.54)$$

Thus, by the Poincaré Recurrence Theorem (Theorem A.1) and (3.54),

$$P(|\varepsilon_n| > \sqrt{\gamma_*} - \epsilon \text{ i.o. } \mid B_k) = 1 \text{ as } M \rightarrow \infty.$$

It follows that, for sufficiently large M and conditionally on B_{k-1} , the following inequality holds with probability 1:

$$\max_{n \in I_k} \varepsilon_n^2 \geq (\sqrt{\gamma_*} - \epsilon)^2 = \gamma_* - \epsilon(2\sqrt{\gamma_*} - \epsilon) \geq \gamma_* - \epsilon(2\sqrt{\gamma_{n-1}} - \epsilon). \quad (3.55)$$

Now, let $J_k = \arg \max_{n \in I_k} \varepsilon_n^2$. Then, by letting $n = J_k$, (3.47) can be written as (replacing J_k by J for simplicity)

$$\gamma_J - \varepsilon_J^2 - \kappa_{J-1} G_J / m \geq 0. \quad (3.56)$$

By (3.55) and (3.56),

$$\gamma_J - \gamma_* \geq \kappa_{J-1} G_J / m - \epsilon(2\sqrt{\gamma_{n-1}} - \epsilon). \quad (3.57)$$

Let

$$d_J \stackrel{\text{def}}{=} \kappa_{J-1} G_J / m - \epsilon(2\sqrt{\gamma_{n-1}} - \epsilon). \quad (3.58)$$

Hence, by (3.57), γ_n is updated (whenever an interval I_k is found at time n) by the recursion

$$\gamma_n = \begin{cases} \gamma_{n-1} - d_J, & \text{if } d_J > 0 \\ \gamma_{n-1}, & \text{otherwise} \end{cases}, \quad (3.59)$$

is an upper bound of γ_* a.s. for each k .

Note that PE and the mixing of $\{u_n\}$ and $\{v_n\}$ implies that, for overestimated γ_n , Lemma 3.2 and then Lemma 3.3 hold. Hence, intervals I_k exist i.o. It follows that, with $\epsilon \rightarrow 0$, (3.59) and Lemma 3.7 imply that γ_n would continue to be updated by the OBE-ABE algorithm until $d_J \rightarrow 0$ and $\gamma_n \rightarrow \gamma_*$ as $n \rightarrow \infty$. Thus, $\kappa_J \rightarrow 0$ when $\epsilon \rightarrow 0$ since $G_n > 0$ due to PE and the stability assumption. Hence, $\kappa_n \rightarrow 0$ due to its non-increasing property. Therefore,

$$P(\omega : \lim_{n \rightarrow \infty} \theta_n = \theta_*) = 1.$$

This completes the proof of a.s. convergence. $\square$

For p. convergence, the mixing assumptions of both $\{u_n\}$ and $\{v_n\}$ in the above theorem can be relaxed to the requirements that both $\{u_n\}$ and $\{v_n\}$ are asymptotically independent sequences (see Theorem 3.3). Further, if $\{u_n\}$ or $\{v_n\}$ or both are continuously-distributed random sequences, then the mixing condition in the theorem above can be relaxed to an ergodic condition for a.s. convergence. The following theorem validates this assertion.

Theorem 3.2 *Assume that the stationary sequences $\{u_n\}$ and $\{v_n\}$ are independent. If $\{u_n\}$ or $\{v_n\}$ or both are continuously-distributed random sequences, PE holds, UCT holds, and $\{u_n\}$ and $\{v_n\}$ are ergodic, then the estimator of the OBE-ABE algorithm is a.s. convergent.*

Proof: Continuously-distributed $\{v_n\}$ or $\{u_n\}$ or both implies that Lemmas 3.2 and 3.3 hold without the assumption that v_n and $\{u_n\}$ are asymptotically independent sequences. The theorem is then proved by following the same steps as in the proof of Theorem 3.1. $\square$

For p. convergence, the assumptions that $\{u_n\}$ and $\{v_n\}$ are independent and stationary, and that $\{u_n\}$ and $\{v_n\}$ are ergodic are not required (see Theorem 3.4). The following corollaries are iid noise cases for a.s. convergence. They are special cases of Theorem 3.1 and 3.2, respectively.

Corollary 3.1 *Assume that the stationary sequences $\{u_n\}$ and $\{v_n\}$ are independent. If PE holds, UT holds, $\{u_n\}$ is mixing, and $\{v_n\}$ is iid, then the estimator of the OBE-ABE algorithm is a.s. convergence.*

Proof: Since UT and UCT imply each other with the iid assumption, and iid implies mixing, a.s. convergence follows immediately from Theorem 3.1. $\square$

Corollary 3.2 *Assume that the stationary sequences $\{u_n\}$ and $\{v_n\}$ are independent. If $\{v_n\}$ or $\{u_n\}$ or both are continuously-distributed random sequences, PE holds, UT holds, $\{u_n\}$ is ergodic, and $\{v_n\}$ is iid, then the estimator of the OBE-ABE algorithm is a.s. convergent.*

Proof: Since UT and UCT imply each other with the iid assumption, and iid implies ergodic, a.s. convergence follows immediately from Theorem 3.2. $\square$

The following corollaries assert the a.s. convergence of the OBE-ABE algorithm when the probability distribution of the noise sequence $\{v_n\}$ is known.

Corollary 3.3 *Assume that the stationary sequences $\{u_n\}$ and $\{v_n\}$ are independent. If PE holds, $\{u_n\}$ is ergodic, and $\{v_n\}$ is iid with uniform distribution, then the estimator of the OBE-ABE algorithm is a.s. convergent.*

Proof: For a uniform distribution, $P(v_n \in (D_\epsilon^+ \cup D_\epsilon^-)) \geq \frac{\epsilon}{2\sqrt{\gamma_n}} > 0$. So, UT holds. By Corollary 3.2, the estimator is a.s. convergent. $\square$

Corollary 3.4 *Assume that the stationary sequences $\{u_n\}$ and $\{v_n\}$ are independent. If PE holds, $\{u_n\}$ is mixing, and $\{v_n\}$ is iid with binary Bernoulli distribution: $P(v_n = \sqrt{\gamma_*}) > \delta$ or $P(v_n = -\sqrt{\gamma_*}) > \delta, \forall n$, then the estimator of the OBE-ABE algorithm is a.s. convergence.*

Proof: Since UT holds in this case, a.s. convergence follows from Corollary 3.1. $\square$

Corollary 3.5 *Assume that the stationary sequences $\{u_n\}$ and $\{v_n\}$ are independent. If PE holds, $\{u_n\}$ is ergodic and continuously distributed, and $\{v_n\}$ is iid with binary Bernoulli distribution: $P(v_n = \sqrt{\gamma_*}) > \delta$ or $P(v_n = -\sqrt{\gamma_*}) > \delta, \forall n$, then the estimator of the OBE-ABE algorithm is a.s. convergent.*

Proof: Since UT holds in this case, the a.s. convergence follows from Corollary 3.2. $\square$

Note that all the theorems and corollaries in this section are also valid for conventional OBE algorithms (and the EPA algorithm) with a known exact noise bound $\sqrt{\gamma_*}$ or $-\sqrt{\gamma_*}$, because, for the OBE-ABE algorithm, $\gamma_n \rightarrow \gamma_*$ as $n \rightarrow \infty$ as seen in the proof of Theorem 3.1.

3.6 Convergence in Probability of OBE-ABE

Although a.s. convergence implies p. convergence, the sufficient conditions for a.s. convergence can be relaxed for p. convergence. Specifically, in addition to other relaxations in some cases, the stationarity of $\{u_n\}$ and $\{v_n\}$ are not required for p. convergence. The main theorem for p. convergence of OBE-ABE algorithm is in the following.

Theorem 3.3 *Assume that the sequences $\{v_n\}$ and $\{u_n\}$ are independent. If PE holds, UCT holds, and $\{u_n\}$ and $\{v_n\}$ are asymptotically independent sequences, then*

the estimator of the OBE-ABE algorithm is $p.$ convergent.

Proof: Since PE (3.12) and UCT (3.9) hold, there exists an infinite subsequence $\{t_i\} \in \mathbb{Z}$ such that for all $n \in \{t_i\}$, (3.9) and

$$P(\mathbf{x}_n \in K \mid \mathcal{F}_{n-2}) \geq \rho_1 > 0 \quad \text{a.s.}, \quad (3.60)$$

hold where K is defined in (3.12). Throughout the proof, all n considered are in $\{t_i\}$. For convenience, let $\{n\}$ replace $\{t_i\}$ as the time coordinate in the proof. The meaning of time intervals, for example, should change accordingly.

Let $\{I_k\}$ be a sequence of time intervals of length $M \in \mathbb{N}$ over which the OBE-ABE algorithm does not update its estimator. Then, for each $k = 1, 2, \dots, K$, it follows from (2.10) that

$$c_n = m(\gamma_n - \varepsilon_n^2) - \kappa_{n-1}G_n \geq 0, \quad \forall n \in I_k. \quad (3.61)$$

Also from (2.5),

$$\varepsilon_n^2 = (y_n - \theta_{n-1}^T \mathbf{x}_n)^2 = (v_n - \tilde{\theta}_{n-1}^T \mathbf{x}_n)^2. \quad (3.62)$$

Let A_n^- be the event that $v_n \in D_\epsilon^-$ and $\tilde{\theta}_{n-1}^T \mathbf{x}_n \geq 0$, and A_n^+ be the event that $v_n \in D_\epsilon^+$ and $\tilde{\theta}_{n-1}^T \mathbf{x}_n < 0$. Also, let $A_n \stackrel{\text{def}}{=} A_n^+ \cup A_n^-$, and denote by A_n^c the complement of A_n . Also note that $\tilde{\theta}_{n-1}^T \mathbf{x}_n$ is $\mathcal{F}_{n-1}$ -measurable.

Let B_k denote the event that A_n occurs at least once in I_k . Then, for all $n \in I_{k+1}$, $B_k \in \mathcal{F}_{n-2}$. Hence, it follows from (3.60) that

$$P(\tilde{\theta}_{n-1}^T \mathbf{x}_n \geq 0 \mid B_k) \geq \rho_1 > \rho > 0 \quad (3.63)$$

and

$$P(\tilde{\theta}_{n-1}^T \mathbf{x}_n < 0 \mid B_k) \geq \rho_2 > \rho > 0. \quad (3.64)$$

Since $B_k \in \mathcal{F}_{n-2} \subset \mathcal{F}_{n-1}$ and $\tilde{\theta}_{n-1}^T \mathbf{x}_n$ is $\mathcal{F}_{n-1}$ -measurable, it follows from UCT (3.9) that

$$P(v_n \in (D_\epsilon^+ \cup D_\epsilon^-) \mid \tilde{\theta}_{n-1}^T \mathbf{x}_n \geq 0, B_k) \geq \delta_1 > \frac{\delta}{\rho} > 0 \quad (3.65)$$

and

$$P(v_n \in (D_\epsilon^+ \cup D_\epsilon^-) \mid \tilde{\theta}_{n-1}^T \mathbf{x}_n < 0, B_k) \geq \delta_2 > \frac{\delta}{\rho} > 0. \quad (3.66)$$

Since A_n^+ and A_n^- are mutually exclusive for sufficiently small ϵ , it follows from (3.63), (3.64), (3.65), and (3.66) that

$$\begin{aligned} P(A_n \mid B_k) &= P(A_n^- \mid B_k) + P(A_n^+ \mid B_k) \\ &= P(v_n \in D_\epsilon^- \mid \tilde{\theta}_{n-1}^T \mathbf{x}_n \geq 0, B_k) \cdot P(\tilde{\theta}_{n-1}^T \mathbf{x}_n \geq 0 \mid B_k) \\ &\quad + P(v_n \in D_\epsilon^+ \mid \tilde{\theta}_{n-1}^T \mathbf{x}_n < 0, B_k) \cdot P(\tilde{\theta}_{n-1}^T \mathbf{x}_n < 0 \mid B_k) \\ &> P(v_n \in D_\epsilon^- \mid \tilde{\theta}_{n-1}^T \mathbf{x}_n \geq 0, B_k) \cdot \rho + P(v_n \in D_\epsilon^+ \mid \tilde{\theta}_{n-1}^T \mathbf{x}_n < 0, B_k) \cdot \rho \\ &> (\delta_1 + \delta_2) \cdot \rho \stackrel{\text{def}}{=} \delta > 0. \end{aligned} \quad (3.67)$$

Hence,

$$P(A_n^c \mid B_k) = 1 - P(A_n \mid B_k) < 1 - \delta.$$

Similarly, since $A_{n-2}^c \in \mathcal{F}_{n-2}$, it follows that

$$\begin{aligned} P(A_n^c \mid A_{n-2}^c, B_k) &= 1 - P(A_n \mid A_{n-2}^c, B_k) = 1 - [P(A_n^- \mid A_{n-2}^c, B_k) + P(A_n^+ \mid A_{n-2}^c, B_k)] \\ &= 1 - \{P(v_n \in D_\epsilon^- \mid \tilde{\theta}_{n-1}^T \mathbf{x}_n \geq 0, A_{n-2}^c, B_k) \cdot P(\tilde{\theta}_{n-1}^T \mathbf{x}_n \geq 0 \mid A_{n-2}^c, B_k) \\ &\quad + P(v_n \in D_\epsilon^+ \mid \tilde{\theta}_{n-1}^T \mathbf{x}_n < 0, A_{n-2}^c, B_k) \cdot P(\tilde{\theta}_{n-1}^T \mathbf{x}_n < 0 \mid A_{n-2}^c, B_k)\} \\ &< 1 - \delta. \end{aligned}$$

Thus,

$$P(A_n^c \cap A_{n-2}^c \mid B_k) = P(A_{n-2}^c \mid B_k) \cdot P(A_n^c \mid A_{n-2}^c, B_k) < (1 - \delta)^2.$$

Similarly,

$$P(A_n^c | A_{n-2}^c A_{n-4}^c B_k) < 1 - \delta.$$

Hence,

$$\begin{aligned} P(A_n^c \cap A_{n-2}^c \cap A_{n-4}^c | B_k) &= P(A_{n-2}^c A_{n-4}^c | B_k) \cdot P(A_n^c | A_{n-2}^c A_{n-4}^c B_k) \\ &< (1 - \delta)^2 \cdot (1 - \delta) = (1 - \delta)^3. \end{aligned}$$

By induction, (for notational simplicity, let M be even)

$$P(\cap_{i=1}^{M/2} A_{2i}^c | B_k) < (1 - \delta)^{M/2}.$$

Thus,

$$\begin{aligned} P(B_k | B_{k-1}) &= P(A_n \text{ occurs at least once in } I_k | B_{k-1}) = P(\cup_{i=1}^M A_i | B_{k-1}) \\ &= 1 - P(\cap_{i=1}^M A_i^c | B_{k-1}) \geq 1 - P(\cap_{i=1}^{M/2} A_{2i}^c | B_{k-1}) \\ &> 1 - (1 - \delta)^{M/2}. \end{aligned} \tag{3.68}$$

Note that by (3.62), the fact that A_n occurs at least once in I_k implies that $\max_{n \in I_k} \varepsilon_n^2 \geq (\sqrt{\gamma_*} - \epsilon)^2$. Hence, from (3.62) and (3.68), it follows that

$$\max_{n \in I_k} \varepsilon_n^2 \geq (\sqrt{\gamma_*} - \epsilon)^2 = \gamma_* - \epsilon(2\sqrt{\gamma_*} - \epsilon) \geq \gamma_* - \epsilon(2\sqrt{\gamma_{n-1}} - \epsilon) \tag{3.69}$$

with probability $1 - (1 - \delta)^{M/2}$ (> 0.99995 if $\delta = 0.01$ and $M > 2000$).

In the case of iid noise and the assumption that $\{v_n\}$ is independent of $\{u_n\}$, the conditional probabilities in the above derivation are not required and can be replaced with marginal probabilities. Hence, (3.69) holds with probability $1 - (1 - \delta)^M$ in this case.

Let $J_k = \arg \max_{n \in I_k} \varepsilon_n^2$. Then, by letting $n = J_k$, (3.61) can be written as (replacing J_k by J for simplicity.)

$$\gamma_J - \varepsilon_J^2 - \kappa_{J-1} G_J / m \geq 0. \quad (3.70)$$

By (3.69) and (3.70),

$$\gamma_J - \gamma_* \geq \kappa_{J-1} G_J / m - \epsilon(2\sqrt{\gamma_{n-1}} - \epsilon). \quad (3.71)$$

Let

$$d_J \stackrel{\text{def}}{=} \kappa_{J-1} G_J / m - \epsilon(2\sqrt{\gamma_{n-1}} - \epsilon). \quad (3.72)$$

Hence, by (3.71), γ_n is updated (whenever an interval I_k is found at time n) by the recursion

$$\gamma_n = \begin{cases} \gamma_{n-1} - d_J, & \text{if } d_J > 0 \\ \gamma_{n-1}, & \text{otherwise} \end{cases}, \quad (3.73)$$

is an upper bound of γ_* with probability $1 - (1 - \delta)^{M/2}$ for each k .

Since the σ -fields $\mathcal{F}_n$ increase with n , i.e., $\mathcal{F}_{n-1} \subset \mathcal{F}_n$, it follows from (3.68) that, for each k ,

$$P(B_k | B_{k-1} B_{k-2} \cdots B_1) > 1 - (1 - \delta)^{M/2}.$$

Hence,

$$\begin{aligned} P(\cap_{k=1}^K B_k) &= P(B_K | B_{K-1} B_{K-2} \cdots B_1) \cdot P(B_{K-1} | B_{K-2} B_{K-3} \cdots B_1) \cdots P(B_2 | B_1) \cdot P(B_1) \\ &> [1 - (1 - \delta)^{\frac{M}{2}}]^K. \end{aligned} \quad (3.74)$$

The assumptions that PE holds, the independence between $\{u_n\}$ and $\{v_n\}$, and that both $\{u_n\}$ and $\{v_n\}$ are asymptotically independent sequences, imply that, for

overestimated γ_n , Lemma 3.2 and then Lemma 3.3 hold. Hence, intervals I_k exist i.o. It follows that with $\epsilon \rightarrow 0$, (3.73) and Lemma 3.7 imply that γ_n would continue to be updated by the OBE-ABE algorithm until $d_J \rightarrow 0$ and $\gamma_n \rightarrow \gamma_*$ as $n \rightarrow \infty$. Thus, $\kappa_J \rightarrow 0$ when $\epsilon \rightarrow 0$ since $G_n > 0$ due to PE and the stability assumption. Hence, $\kappa_n \rightarrow 0$ due to its non-increasing property.

Note that $K < N/M$ where N is the total “time” (number of data sets). Let $M = \lfloor \sqrt{2N} \rfloor + 1$ where $\lfloor \cdot \rfloor$ denotes the integer part. Then, (3.74) becomes

$$P(\cap_{k=1}^K B_k) > [1 - (1 - \delta)^{\frac{M}{2}}]^{\frac{N}{M}} \geq [1 - (1 - \delta)^{\sqrt{N/2}}]^{\sqrt{N/2}} \rightarrow 1 \quad \text{as } N \rightarrow \infty.$$

This is verified by applying L’Hopital’s rule as shown below to find that $(1 - q^n)^n \rightarrow 1$, as $n \rightarrow \infty$, where $0 \leq q < 1$:

$$\begin{aligned} \lim_{n \rightarrow \infty} (1 - q^n)^n &= \lim_{n \rightarrow \infty} \exp[n \ln(1 - q^n)] = \exp[\lim_{n \rightarrow \infty} n \ln(1 - q^n)] \\ &= \exp\left(\lim_{n \rightarrow \infty} \frac{\ln(1 - q^n)}{\frac{1}{n}}\right) = \exp\left(\lim_{n \rightarrow \infty} \frac{-\frac{1}{1-q^n} q^n \ln q}{-\frac{1}{n^2}}\right) \\ &= \exp\left(\lim_{n \rightarrow \infty} \frac{n^2 \ln q}{q^{-n} - 1}\right) = \exp\left(\lim_{n \rightarrow \infty} \frac{2n \ln q}{-q^{-n} \ln q}\right) \\ &= \exp\left(\lim_{n \rightarrow \infty} \frac{2n}{-q^{-n}}\right) = \exp\left(\lim_{n \rightarrow \infty} \frac{2}{q^{-n} \ln q}\right) \\ &= \exp(0) = 1. \end{aligned}$$

Thus,

$$\lim_{n \rightarrow \infty} P(\|\theta_n - \theta_*\| > \epsilon) = 0.$$

This completes the proof of p. convergence. $\square$

Remark 1: Theoretically, instead of the *a priori* knowledge of γ_* to achieve a.s. or p. convergence, the OBE-ABE algorithm requires only a *lower bound*, $\delta(\epsilon)$, of the

ϵ -tail probability of v_n [see (3.10)].

Remark 2: For the application of the OBE-ABE algorithm, since N is finite, an ϵ is first arbitrarily chosen to be a small positive number (κ_N decreases with ϵ). Given this ϵ , a δ is estimated. Any lower bound satisfying UT (3.10) (in the iid case, for example) will work. Then, a large enough N and $M(= \lfloor \sqrt{N} \rfloor + 1$, for example) are chosen so that the probability $[1 - (1 - \delta)^M]^{\frac{N}{M}}$ is close to one. For example, if $\delta = 0.1$, choose $M = 101$ and $N = 10,000$, for example, to get the probability $[1 - (1 - \delta)^M]^{\frac{N}{M}} > 0.997$. If $\delta = 0.02$, choose $M = 601$ and $N = 360,000$, for example, to get the probability $[1 - (1 - \delta)^M]^{\frac{N}{M}} > 0.996$. Since the probability $[1 - (1 - \delta)^M]^{\frac{N}{M}}$ is the most conservative value, if N is fixed in practice and is not too large ($N < 10,000$, for example), M can be simply chosen large enough to make $1 - (1 - \delta)^M$ close to one. If, in a real-world application, δ is still difficult to estimated after choosing ϵ , then simply start the OBE-ABE algorithm with a small M ($M = 20$ for example) and then increase M after each trial if necessary. The negativity of κ_n can serve as an indication of a too small M . The algorithm is guaranteed to be p. convergent (with probability close to 1) if M is large enough when sufficient conditions (PE, UCT) are satisfied.

If $\{u_n\}$ or $\{v_n\}$ or both are continuously-distributed random sequences, then the requirements that both $\{u_n\}$ and $\{v_n\}$ be asymptotically independent sequences, and that $\{u_n\}$ be independent of $\{v_n\}$ are not required for p. consistency. The following theorem validates this assertion.

Theorem 3.4 *If $\{u_n\}$ or $\{v_n\}$ or both are continuously-distributed random sequences, PE holds, and UCT holds, then the estimator of the OBE-ABE algorithm is p. convergent.*

Proof: Continuously-distributed v_n implies that Lemmas 3.2 and 3.3 hold without the assumption that both $\{u_n\}$ and $\{v_n\}$ are asymptotically independent sequences, and that $\{u_n\}$ is independent of $\{v_n\}$. The theorem is then proved by following the steps of the proof of Theorem 3.3. $\square$

The following corollaries involve iid noise cases for p. convergence. They are special cases of Theorems 3.3 and 3.4, respectively.

Corollary 3.6 *Assume that the sequences $\{u_n\}$ and $\{v_n\}$ are independent. If PE holds, UT holds, $\{u_n\}$ is asymptotically independent, and $\{v_n\}$ is iid, then the estimator of the OBE-ABE algorithm is a.s. convergent.*

Proof: Since UT and UCT imply each other with the iid assumption, and iid implies asymptotically independent, p. convergence follows immediately from Theorem 3.3. $\square$

Corollary 3.7 *If $\{u_n\}$ or $\{v_n\}$ or both are continuously-distributed random sequences, PE holds, UT holds, and $\{v_n\}$ is iid, then the estimator of the OBE-ABE algorithm is p. convergent.*

Proof: Since UT and UCT imply each other with the iid assumption, p. convergence follows immediately from Theorem 3.4. $\square$

The following corollaries assert the p. convergence of the OBE-ABE algorithm when the probability distribution of the noise sequence $\{v_n\}$ is known.

Corollary 3.8 *If PE holds, and $\{v_n\}$ is iid with uniform distribution, then the estimator of the OBE-ABE algorithm is p. convergent.*

Proof: For a uniform distribution, $P(v_n \in (D_\epsilon^+ \cup D_\epsilon^-)) \geq \frac{\epsilon}{2\sqrt{\gamma_n}} > 0$. So, UT holds. By Corollary (3.7), the estimator is p. convergent. $\square$

Corollary 3.9 *Assume that the sequences $\{u_n\}$ and $\{v_n\}$ are independent. If PE holds, $\{u_n\}$ is asymptotically independent, and $\{v_n\}$ is iid with binary Bernoulli distribution: $P(v_n = \sqrt{\gamma_*}) > \delta$ or $P(v_n = -\sqrt{\gamma_*}) > \delta, \forall n$, then the estimator of the OBE-ABE algorithm is a.s. convergent.*

Proof: Since UT holds in this case, the p. convergence follows from Corollary 3.6. $\square$

Corollary 3.10 *If PE holds, $\{u_n\}$ is continuously distributed, and $\{v_n\}$ is iid with binary Bernoulli distribution: $P(v_n = \sqrt{\gamma_*}) > \delta$ or $P(v_n = -\sqrt{\gamma_*}) > \delta, \forall n$, then the estimator of the OBE-ABE algorithm is p. convergent.*

Proof: Since UT holds in this case, the p. convergence follows from Corollary 3.7. $\square$

Note that all the theorems and corollaries in this section are also valid for conventional OBE algorithms (and the EPA algorithm) with a known exact noise bound $\sqrt{\gamma_*}$ or $-\sqrt{\gamma_*}$, because, for the OBE-ABE algorithm, $\gamma_n \rightarrow \gamma_*$ as $n \rightarrow \infty$ as seen in the proof of Theorem 3.3.

3.7 Almost Sure and Probability Convergence of Sub-OBE-ABE

All the convergence theorems developed in previous sections are valid for Sub-OBE-ABE algorithm in various noise cases with the same sufficient conditions respectively. The sketch of proofs of all the theorems for Sub-OBE-ABE is in the following. The algorithmic steps of the algorithm are found in Table 2.3.

For convenience, let us rewrite the optimal check coefficient c_n , suboptimal check coefficient $\bar{c}_n$, and recursion formula for d_n here:

$$c_n = m\gamma_n - m\varepsilon_n^2 - \kappa_{n-1}G_n \quad (3.75)$$

$$\bar{c}_n = m\gamma_n - m\varepsilon_n^2 - \kappa_{n-1}\mathbf{x}_n^T\mathbf{x}_n/d_{n-1} \quad (3.76)$$

$$d_n = \alpha_n d_{n-1} + \beta_n \mathbf{x}_n^T \mathbf{x}_n. \quad (3.77)$$

Also, define

$$\bar{c}_n^- = \begin{cases} c_n, & \text{if } \bar{c}_n < 0 \\ 0, & \text{otherwise.} \end{cases} \quad (3.78)$$

and

$$\bar{c}_n^+ = \begin{cases} c_n, & \text{if } \bar{c}_n \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

Hence, $c_n = \bar{c}_n^+ + \bar{c}_n^-$. The following lemmas are needed for the proof of the main theorem.

Lemma 3.5 $\bar{c}_n < 0$ implies $c_n < 0$

Proof: Since the matrix $\mathbf{P}_n$ is real symmetric,

$$G_n = \mathbf{x}_n^T \mathbf{P}_{n-1} \mathbf{x}_n \geq \nu_{\min}(\mathbf{P}_{n-1}) \mathbf{x}_n^T \mathbf{x}_n = \frac{1}{\nu_{\max}(\mathbf{P}_{n-1}^{-1})} \mathbf{x}_n^T \mathbf{x}_n \quad (3.79)$$

where $\nu_{\min}(\cdot)$ and $\nu_{\max}(\cdot)$ denote the minimum and maximum eigenvalues, respectively. Also, as in (2.9)

$$\mathbf{P}_n^{-1} = \alpha_n \mathbf{P}_{n-1}^{-1} + \beta_n \mathbf{x}_n \mathbf{x}_n^T.$$

This implies

$$\text{tr}(\mathbf{P}_n^{-1}) = \alpha_n \text{tr}(\mathbf{P}_{n-1}^{-1}) + \beta_n \mathbf{x}_n^T \mathbf{x}_n. \quad (3.80)$$

Since $\mathbf{P}_n^{-1}$ is positive definite,

$$\text{tr}(\mathbf{P}_n^{-1}) \geq \nu_{\max}(\mathbf{P}_n^{-1})$$

That is,

$$\frac{1}{\text{tr}(\mathbf{P}_n^{-1})} \leq \frac{1}{\nu_{\max}(\mathbf{P}_n^{-1})}. \quad (3.81)$$

Comparing (3.80) and (3.77), we see that $d_n = \text{tr}(\mathbf{P}_n^{-1})$. Also, by (3.79) and (3.81),

$$G_n \geq \mathbf{x}_n \mathbf{x}_n^T / d_{n-1}. \quad (3.82)$$

Hence, by (3.75), (3.76) and (3.82), the lemma is proved. $\square$

Note that a version of Lemma A.2 is valid for Sub-OBE-ABE as is stated in Lemma 3.6 below. The proof is the same except that c_n^- in the original proof [27] is replaced by $\bar{c}_n^-$.

Lemma 3.6 *For Sub-OBE-ABE algorithm, $\lim_{n \rightarrow \infty} \bar{c}_n^- = 0$, where $\bar{c}_n^-$ is defined in (3.78).*

Theorem 3.5 *All the convergence theorems of OBE-ABE are valid for Sub-OBE-ABE with the same respective conditions.*

Proof: That Lemma 3.3 is valid for Sub-OBE-ABE is shown by substituting c_n^+ and c_n^- in the proof with $\bar{c}_n^+$ and $\bar{c}_n^-$, respectively. Also note that $P(\bar{c}_n^- < 0) \leq P(c_n^- < 0)$ which is implied by Lemma 3.5. Since other lemmas (including Lemmas 3.5 and 3.6) which are necessary for the proof of the main theorems of Sub-OBE-ABE algorithm are not affected by the suboptimal check $\bar{c}_n$, it follows that the proofs of convergence for OBE-ABE are also valid for Sub-OBE-ABE provided that c_n and G_n in the proofs are changed to $\bar{c}_n$ and $\mathbf{x}_n^T \mathbf{x}_n / d_{n-1}$, respectively. $\square$

3.8 A Necessary Condition

All convergence theorems in the previous sections have two common sufficient conditions: PE and UCT (or UT in the iid case). PE has been shown to be a necessary condition [5, 31, 42] for the convergence of any LSE algorithm. Another necessary condition for the convergence of those algorithms is the whiteness of the noise sequence. This is a logical result since white noise carries the least power of all WSS noises.

SM algorithms are based on minimizing the feasible set (polytope or ellipsoid) according to assumed known noise bounds. Although some SM algorithms (especially OBEs) have similar recursion formulae to RLS, a LSE algorithm, the convergence of SM algorithms does not rely on the whiteness of the noise sequence. Even in the white noise case, the estimator of an OBE algorithm (i.e. the center of ellipsoid at each step) may be biased away from the true parameter vector due to the optimal data weighting process. This is reasonable since whiteness is an “average” property, not a pointwise property (e.g. UCT and UT are pointwise properties).

If UCT or UT does not hold, then the ellipsoid of SM algorithms will not shrink to a point, hence convergence is not guaranteed. The following lemma which is extended from Lemma A.1 by considering an almost sure (a.s.) ω -set in its proof, validates this assertion.

Lemma 3.7 *Assume that the PE (3.12) holds. If there exists an $\epsilon > 0$ and $N \in \mathbb{N}$, such that $\gamma_n - v_n^2 > \epsilon$, $\forall n > N$, a.s., then the sequence of the ellipsoids of OBE algorithms does not asymptotically shrink to a point a.s.*

Theorem 3.6 *Assume that PE holds. Then, UCT is a necessary condition for the sequence of the ellipsoids of any OBE algorithm to shrink to a point a.s.*

Proof: Suppose that UCT does not hold. Then, considering the notes below (3.9),

there exist an $\epsilon > 0$ and an $N \in \mathbf{N}$ such that, for all $n > N$, $P(v_n \in D_\epsilon^+ | \mathcal{F}_{n-1}) = 0$ and $P(v_n \in D_\epsilon^- | \mathcal{F}_{n-1}) = 0$ a.s. Hence, $\gamma_n - v_n^2 > \epsilon$, $\forall n > N$ a.s. Thus, by Lemma 3.7, the sequence of the ellipsoids does not shrink to a point a.s. $\square$

Chapter 4

Simulation Studies

4.1 Introduction

In this chapter, the performance of each of the new algorithms is investigated through simulations. Although the algorithms have been proven to be convergent in iid, mixing, ergodic, and non-stationary noise cases in the previous chapter, simulations can provide other characteristics which are not explicitly inferred from theoretical analysis. For example, simulations can illustrate how the speed of convergence of OBE-ABE algorithm is affected by the ABE procedure. Simulations also reveal that the suboptimal check in Sub-OBE-ABE does not slow down the speed of convergence. This is not surprising since the noise bounds estimated by the ABE procedure are not affected by the suboptimal check. Simulations also reveal the effect of asymmetric noise bounds on the performance of convergence.

The version of OBE algorithms used in this dissertation is SM-SA algorithm due to its interpretable optimization criterion and numerical stability. Colored noise cases, in addition to iid cases, are also demonstrated.

4.2 IID Noise Cases

In this section, the performance of OBE-ABE, Sub-OBE-ABE, and conventional OBE are compared in iid noise cases with symmetric and asymmetric bounds. Comparisons of performance of OBE-type algorithms with RLS in the iid case are found in [27]. Hence, they are omitted here. When compared to RLS, OBE algorithms have better speed of convergence in many noise cases.

Throughout this section, an AR(3) model is simulated as follows:

$$y_n = a_{1*}y_{n-1} + a_{2*}y_{n-2} + a_{3*}y_{n-3} + v_n \quad (4.1)$$

where $a_{1*} = 2, a_{2*} = -1.48, a_{3*} = 0.34$ are unknown parameters to be identified. Three types of noise sequences $\{v_n\}$ are employed to generate three observable output sequences $\{y_n\}$. Specifically,

CASE 1: AR(3) model as (4.1). $v_n \sim B(-1, 1)$ is iid and non-zero-mean with binary Bernoulli distribution.

$$v_n = \begin{cases} 1, & \text{with probability 0.7} \\ -1, & \text{with probability 0.3.} \end{cases} \quad (4.2)$$

CASE 2: AR(3) model as (4.1). $v_n \sim U(-1, 1)$ is uniformly distributed on $[-1, 1]$.

The following is a case of asymmetric noise bounds:

CASE 3: AR(3) model as (4.1). $v_n \sim B(-0.5, 1)$ is iid, non-zero-mean, asymmetri-

cally bounded with binary Bernoulli distribution. Specifically,

$$v_n = \begin{cases} 1, & \text{with probability 0.7} \\ -0.5, & \text{with probability 0.3.} \end{cases} \quad (4.3)$$

The following case is constructed for the illustration of the final ellipsoid of the algorithm and the trajectory of the estimator.

CASE 4: AR(2) model is constructed as follows:

$$y_n = -0.1y_{n-1} + 0.56y_{n-2} + v_n \quad (4.4)$$

in which v_n is as in CASE 1, (4.2).

Note that the noise bound ($\gamma_* = 1$) and the non-zero mean of v_n are assumed *unknown* in these cases.

4.2.1 OBE-ABE vs. OBE

In this subsection, OBE-ABE and OBE are compared with regard to the speed of convergence and computational efficiency using CASEs 1 – 3. These results support the validity of the theorems in the previous chapter.

For CASE 1, the simulation results of OBE-ABE ($\gamma_0 = 1.5$, $\epsilon = 0$, and $M = 20$) and OBE ($\gamma_0 = 1.5$) are shown in Figs. 4.1 – 4.3. As seen in the figures, neither estimator is affected by the non-zero mean of v_n and both algorithms converge to the true parameter ($a_{1*} = 2$). This supports the assertions of related theorems in Chap. 3.

However, with the help of automatic bound estimation, OBE-ABE converges faster

(as seen in Fig. 4.1) than OBE. This result is not unexpected in light of Lemma 3.7 and related theorems in Chap. 3.

For CASE 2, the simulation results of OBE-ABE ($\gamma_0 = 1.5$, $\epsilon = 0.05$, and $M = 50$) and OBE ($\gamma_0 = 1.5$) are shown in Figs. 4.4 – 4.6. The difference of convergence performance between OBE-ABE and OBE is similar to those of CASE 1.

Further, a comparison between CASE 1 and CASE 2 shows that the speed of convergence of both algorithms is proportional to the ϵ -tail probability, δ , defined in UT (3.10) or UCT (3.9).

For CASE 3, the simulation results of OBE-ABE ($\gamma_0 = 1.5$, $\epsilon = 0$, and $M = 70$) and OBE ($\gamma_0 = 1.5$) are shown in Figs. 4.7 – 4.9. As seen in the figures, the convergence of (overestimated) OBE is affected by the asymmetry of noise bounds while OBE-ABE converges consistently as estimated bounds γ_n automatically converge to the true bound ($= 1$).

For CASE 4, the final ellipsoid of OBE-ABE ($\gamma_0 = 1.5$, $\epsilon = 0$, and $M = 20$) and the trajectory of the estimator after 1000 steps are shown Fig. 4.10. The trajectory of the estimator starts from the 17th step. The ellipsoid does not degenerate to a line segment. This conforms to the positiveness of $\mathbf{P}_n$ and κ_n under PE and stable assumptions.

These simulations show that the number of updates for the estimator of OBE-ABE algorithm is smaller than that of the (overestimated bound) OBE algorithm if both algorithms terminate at same ellipsoidal volume. This means that OBE-ABE shrinks the volumes of ellipsoids *more* drastically than OBE each time that the updating occurs. The improved computational efficiency of OBE-ABE is not surprising since the ABE procedure, having trivial computational cost as seen in CASE 1 through CASE 3, blindly gets *more correct* information (noise bound) than those of OBE.

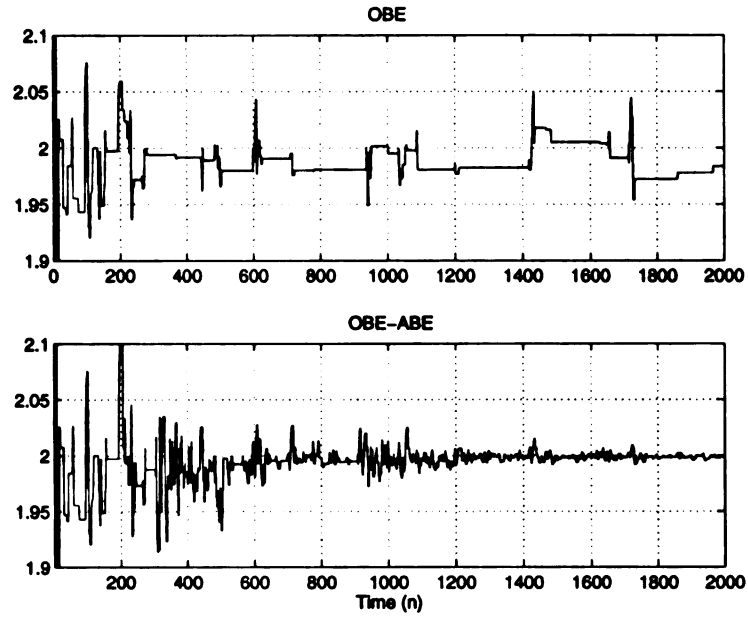


Figure 4.1: Estimators of $a_{1*} = 2$. CASE 1 : $v_n \sim B(-1,1)$ is non-zero mean for model (4.1).

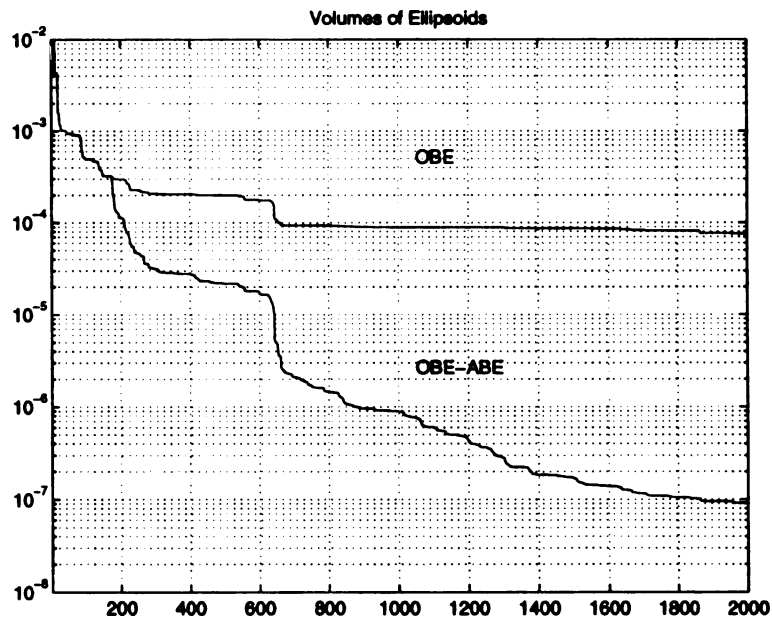


Figure 4.2: CASE 1 : $v_n \sim B(-1,1)$ is non-zero mean for model (4.1).

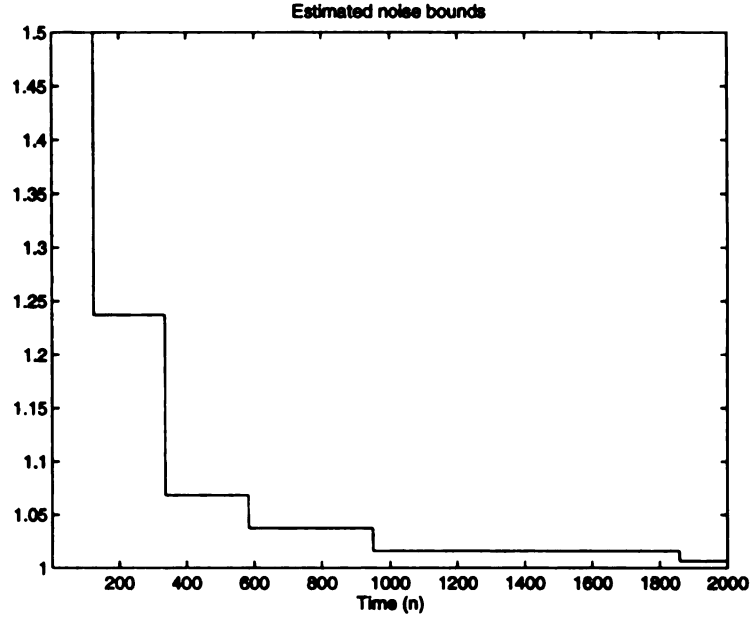


Figure 4.3: CASE 1 : $v_n \sim B(-1, 1)$ is non-zero mean for model (4.1).

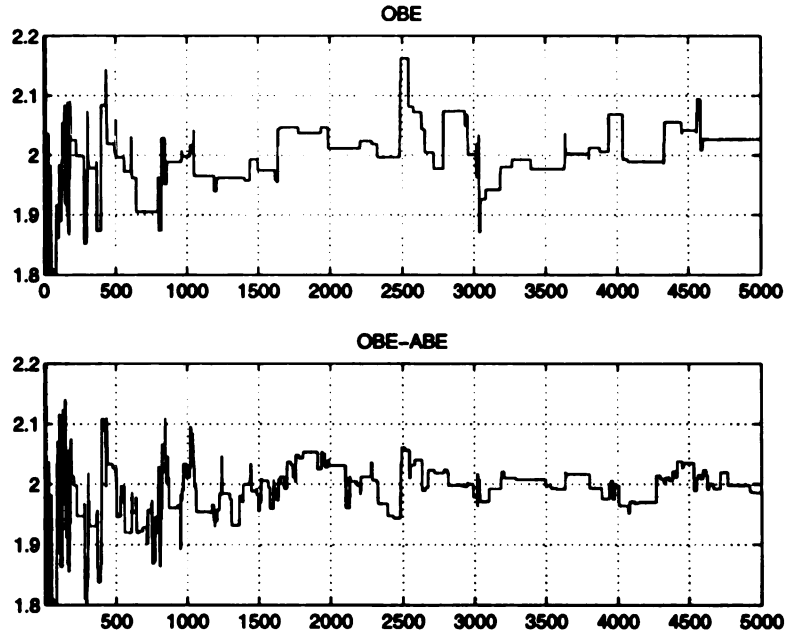


Figure 4.4: Estimators of $a_{1*} = 2$. CASE 2 : $v_n \sim U(-1, 1)$ for model (4.1).

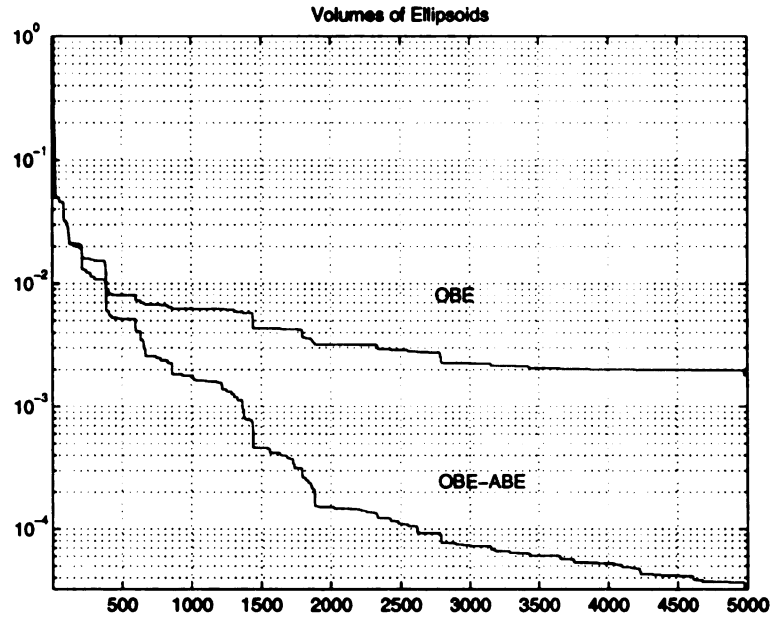


Figure 4.5: CASE 2 : $v_n \sim U(-1, 1)$ for model (4.1).

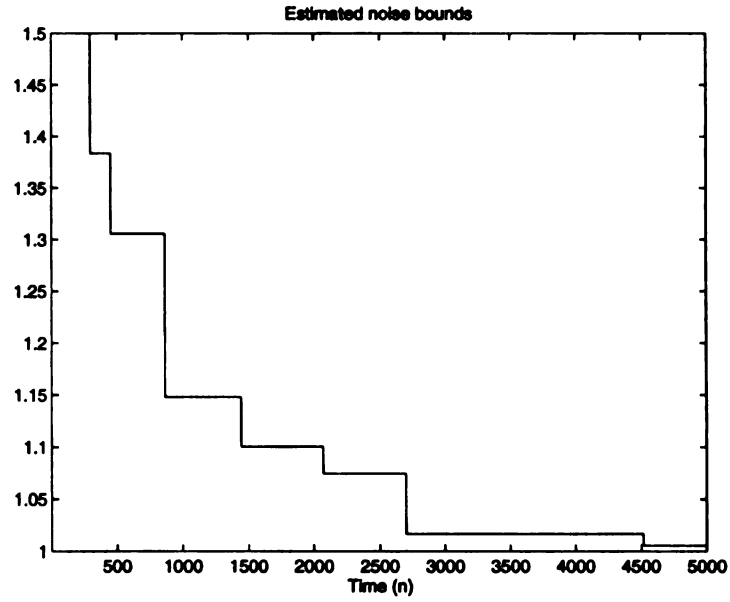


Figure 4.6: CASE 2 : $v_n \sim U(-1, 1)$ for model (4.1).

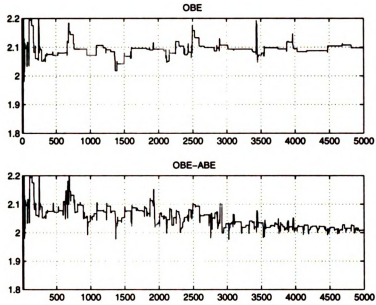


Figure 4.7: Estimators of $a_{1*} = 2$. CASE 3 : $v_n \sim B(-0.5, 1)$, (4.3), for model (4.1).

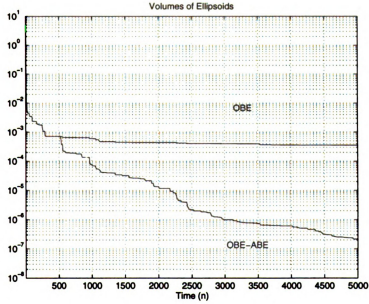


Figure 4.8: CASE 3 : $v_n \sim B(-0.5, 1)$, (4.3), for model (4.1).

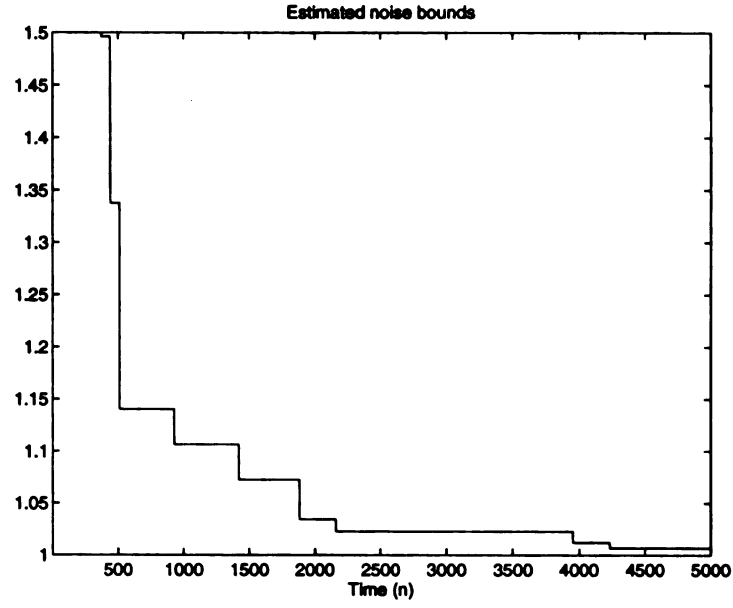


Figure 4.9: CASE 3 : $v_n \sim B(-0.5, 1)$, (4.3), for model (4.1).

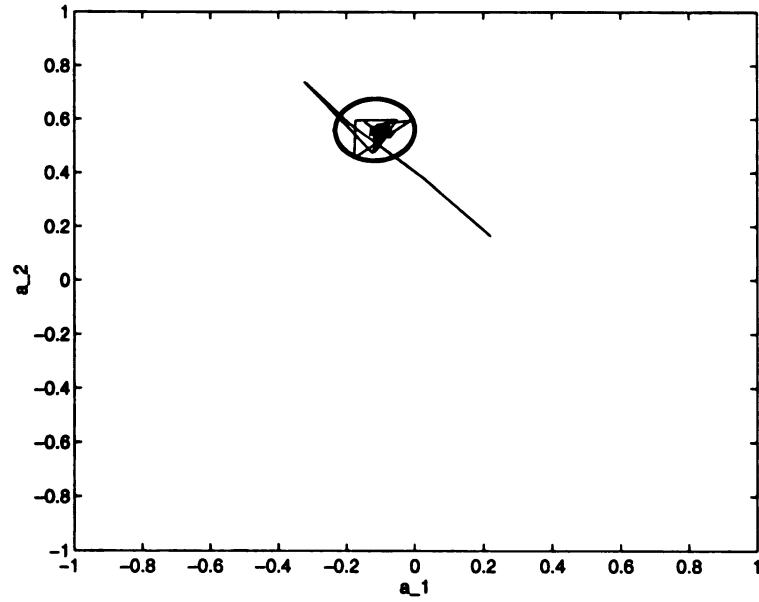


Figure 4.10: CASE 4 : The final ellipsoid and the trajectory of the estimator. $v_n \sim B(-1, 1)$, (4.2), for model (4.4).

4.2.2 Sub-OBE-ABE vs. OBE-ABE

In this subsection, Sub-OBE-ABE and OBE-ABE are compared with regard to the speed of convergence and computational efficiency using CASEs 1 and 2. These convergence results support the validity of the theorems in the previous chapter.

For CASE 1, the simulation results of OBE-ABE ($\gamma_0 = 1.5$, $\epsilon = 0$, and $M = 20$) and Sub-OBE-ABE ($\gamma_0 = 1.5$, $\epsilon = 0$, and $M = 20$) are shown in Figs. 4.11 and 4.12. As seen in the figures, neither estimator is affected by the non-zero mean of v_n , and both algorithms converge to the true parameter ($a_{1*} = 2$).

In this case, Sub-OBE-ABE, which selects 18% of the data, has similar convergence performance to OBE-ABE which selects 50% of the data. This reveals that the excellent performance of ABE procedure is not affected by the suboptimal check.

For CASE 2, the simulation results of OBE-ABE ($\gamma_0 = 1.5$, $\epsilon = 0.05$, and $M = 50$) and Sub-OBE-ABE ($\gamma_0 = 1.5$, $\epsilon = 0.05$, and $M = 30$) are shown in Figs. 4.13 and 4.14. As seen in the figures, neither estimator is affected by the non-zero mean of v_n and both algorithms converge to the true parameter ($a_{1*} = 2$). Sub-OBE-ABE which selects very few (1.5%) data in this case has similar convergence performance to OBE-ABE which selects 5% of the data. However, both in CASEs 1 and 2, the speed of convergence of Sub-OBE-ABE in the beginning (e.g. first 150 steps) is slower than that of OBE-ABE. This is because, in that stage, Sub-OBE-ABE has big ellipsoids and is busy updating the estimator using the suboptimal check without updating the estimated noise bound using the ABE procedure.

4.3 Colored Noise Cases

In this section, the performances of OBE-ABE, Sub-OBE-ABE, and conventional OBE are compared in colored noise cases with symmetric and asymmetric bounds.

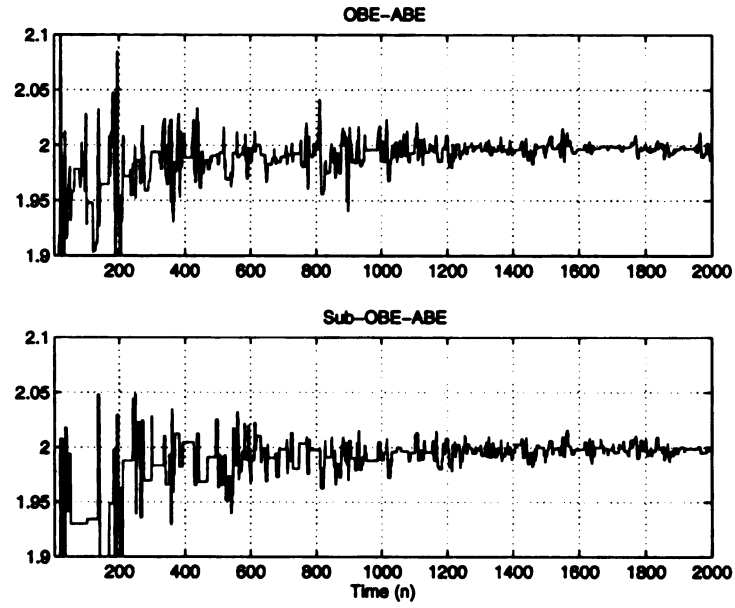


Figure 4.11: Estimators of $a_{1*} = 2$. CASE 1 : $v_n \sim B(-1,1)$ is non-zero mean for model (4.1).

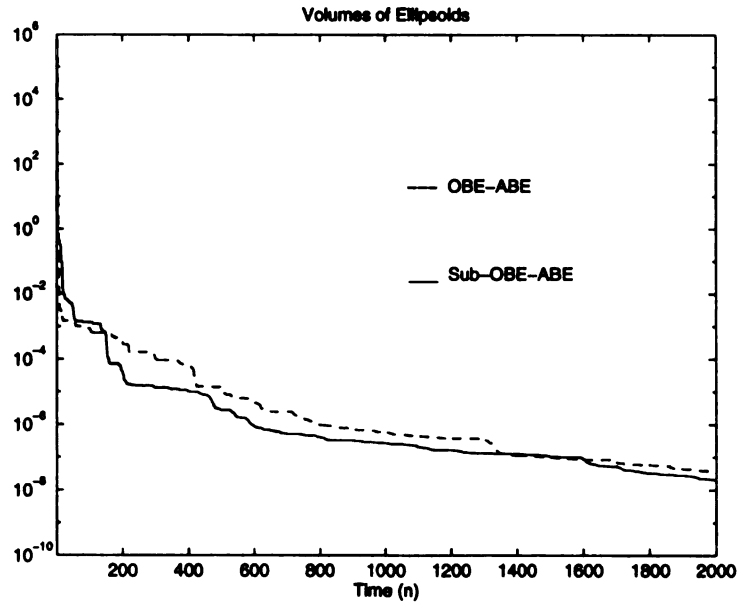


Figure 4.12: CASE 1 : $v_n \sim B(-1,1)$ is non-zero mean for model (4.1).

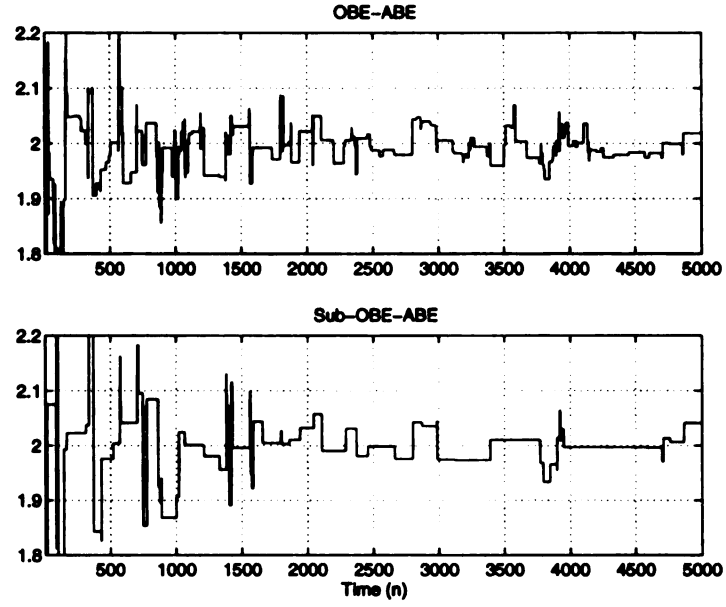


Figure 4.13: Estimators of $a_{1*} = 2$. CASE 2 : $v_n \sim U(-1, 1)$ for model (4.1).

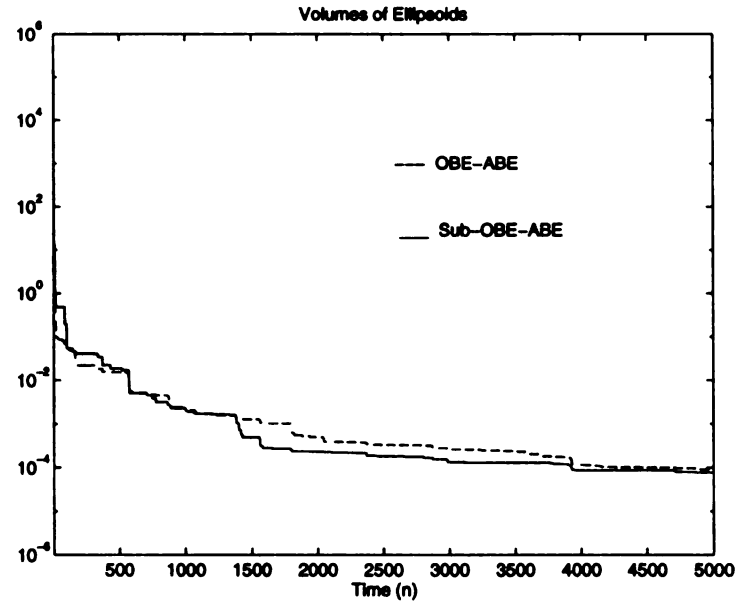


Figure 4.14: CASE 2 : $v_n \sim U(-1, 1)$ for model (4.1).

Four colored noise cases are constructed as follows:

CASE 5: AR(3) model as (4.1). v_n is a colored non-zero-mean noise sequence related to a colored sequence $\{w_n\}$ as follows:

$$v_n = \begin{cases} 1, & \text{if } w_n > -1 \\ -1, & \text{otherwise} \end{cases} \quad (4.5)$$

in which the colored sequence $\{w_n\}$ is generated by a uniformly-distributed iid white noise sequence $z_n \sim U(-1, 1)$ as follow:

$$w_n = -0.8w_{n-1} + z_n.$$

CASE 6: v_n is the same as in CASE 5, (4.5), while the order of the AR model increases to 12. Specifically, a stable AR(12) model is employed in this case as follows:

$$y_n = a_{1*}y_{n-1} + a_{2*}y_{n-2} + \cdots + a_{12*}y_{n-12} + v_n \quad (4.6)$$

where $a_{1*} = -0.1, a_{2*} = 0.9175, a_{3*} = -0.191, a_{4*} = -0.2253, a_{5*} = 0.2601, a_{6*} = 0.0046, a_{7*} = -0.0367, a_{8*} = -0.0209, a_{9*} = -0.0082, a_{10*} = 0.0095, a_{11*} = -0.0052,$ and $a_{12*} = -0.0041$ are unknown parameters to be identified.

CASE 7: AR(3) model as (4.1). v_n is a colored non-zero-mean noise sequence with asymmetric bounds generated by a colored sequence $\{w_n\}$, i.e.,

$$v_n = \begin{cases} 1, & \text{if } w_n > -1 \\ -0.5, & \text{otherwise} \end{cases} \quad (4.7)$$

The following case is constructed for the illustration of the final ellipsoid of the algo-

rithm and the trajectory of the estimator in colored noise case.

CASE 8: AR(2) model as in CASE 4, (4.4). v_n is as in CASE 5, (4.5).

4.3.1 OBE-ABE vs. OBE

In this subsection, OBE-ABE and OBE are compared with regard to the speed of convergence and computational efficiency using CASEs 5 – 7. These results support the validity of the theorems in the previous chapter.

For CASE 5, the simulation results of OBE-ABE ($\gamma_0 = 1.5$, $\epsilon = 0$, and $M = 40$) and OBE ($\gamma_0 = 1.5$) are shown in Figs. 4.15 – 4.17. As seen in the figures, neither estimator is affected by the color of v_n and both algorithms converge to the true parameter ($a_{1*} = 2$). This supports the assertions of related theorems in Chap. 3.

For CASE 6, the simulation results of OBE-ABE ($\gamma_0 = 1.5$, $\epsilon = 0$, and $M = 40$) and OBE ($\gamma_0 = 1.5$) in this case are shown in Figs. 4.18 – 4.20. As seen in the figures, the slow speed of convergence of OBE due to overestimated noise bound results in poor convergence, while OBE-ABE converges very well to the true parameter ($a_{1*} = -0.1$). This shows that the excellent convergence performance is not affected by model order while (overestimated bound) OBE is.

For CASE 7, the simulation results of OBE-ABE ($\gamma_0 = 1.5$, $\epsilon = 0$, and $M = 250$) and OBE ($\gamma_0 = 1.5$) are shown in Figs. 4.21 – 4.23. As seen in the figures, the convergence of (overestimated) OBE is affected by the asymmetry of noise bounds while OBE-ABE converges consistently as estimated bounds γ_n automatically converge to the true bound.

For CASE 8, the final ellipsoid of OBE-ABE ($\gamma_0 = 1.5$, $\epsilon = 0$, and $M = 40$) and the trajectory of the estimator after 1000 steps are shown in Fig. 4.10. The trajectory

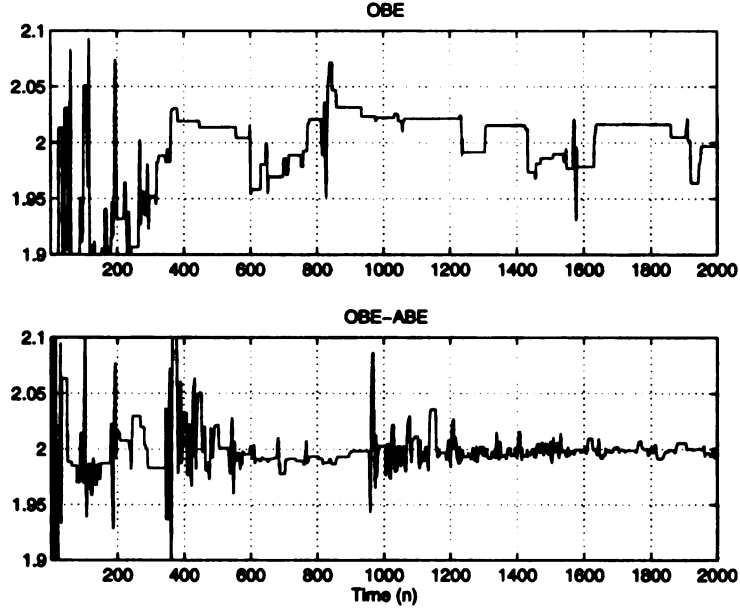


Figure 4.15: Estimators of $a_{1*} = 2$. CASE 5: v_n is colored in (4.5) for model (4.1).

of the estimator starts from the 17th step. The ellipsoid does not degenerate to a line segment. This conforms to the positiveness of P_n and κ_n under PE and stable assumptions.

4.3.2 Sub-OBE-ABE vs. OBE-ABE

In this subsection, Sub-OBE-ABE and OBE-ABE are compared with regard to the speed of convergence and computational efficiency using CASEs 5 and 6. These results support the validity of the theorems in the previous chapter.

For CASE 5, the simulation results of OBE-ABE ($\gamma_0 = 1.5$, $\epsilon = 0$, and $M = 20$) and Sub-OBE-ABE ($\gamma_0 = 1.5$, $\epsilon = 0$, and $M = 20$) are shown in Figs. 4.25 and 4.26. As seen in the figures, neither estimator is affected by the color and non-zero mean of v_n and both algorithms converge to the true parameter ($a_{1*} = 2$). This supports the assertions of related theorems in Chap. 3.

In this case, Sub-OBE-ABE, which selects 15% of the data, has similar convergence performance to OBE-ABE which selects 45% of the data. This reveals that the

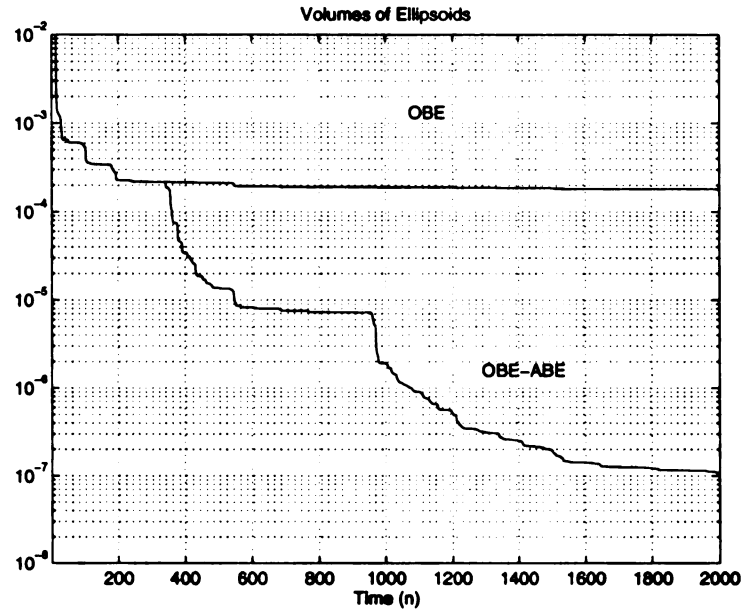


Figure 4.16: CASE 5 : v_n is colored in (4.5) for model (4.1).

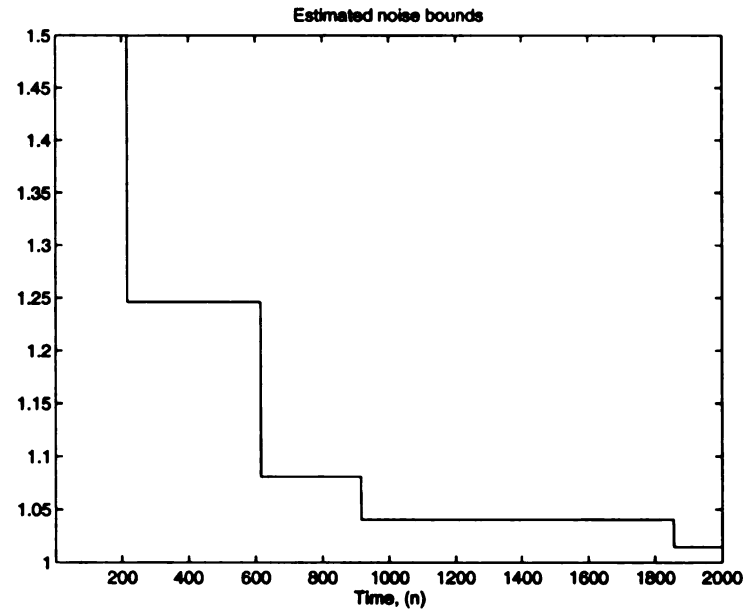


Figure 4.17: CASE 5 : v_n is colored in (4.5) for model (4.1).

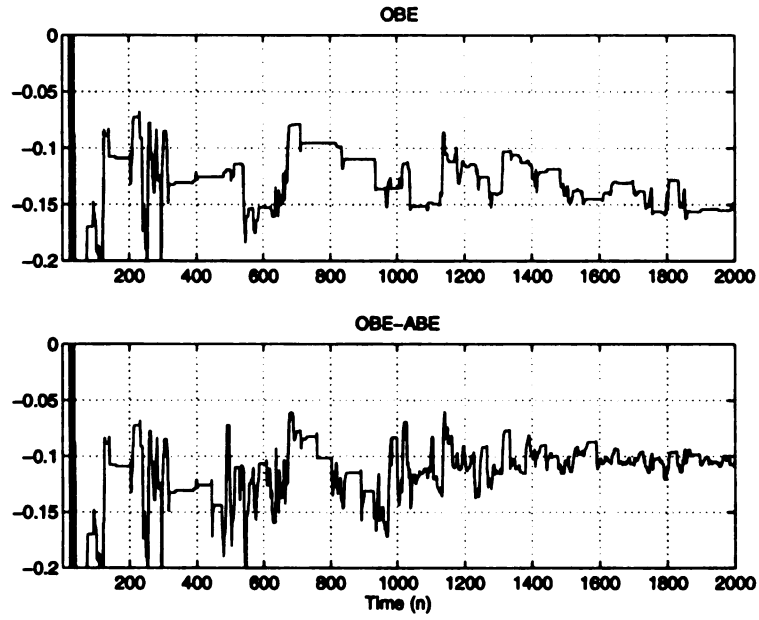


Figure 4.18: Estimators of $a_{1*} = -0.1$. CASE 6: v_n is colored in (4.5) for model (4.6).

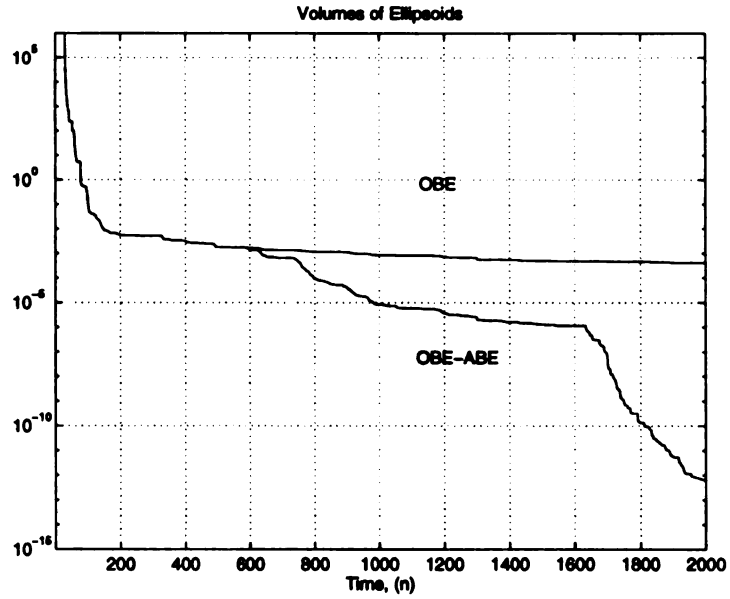


Figure 4.19: CASE 6 : v_n is colored in (4.5) for model (4.6).

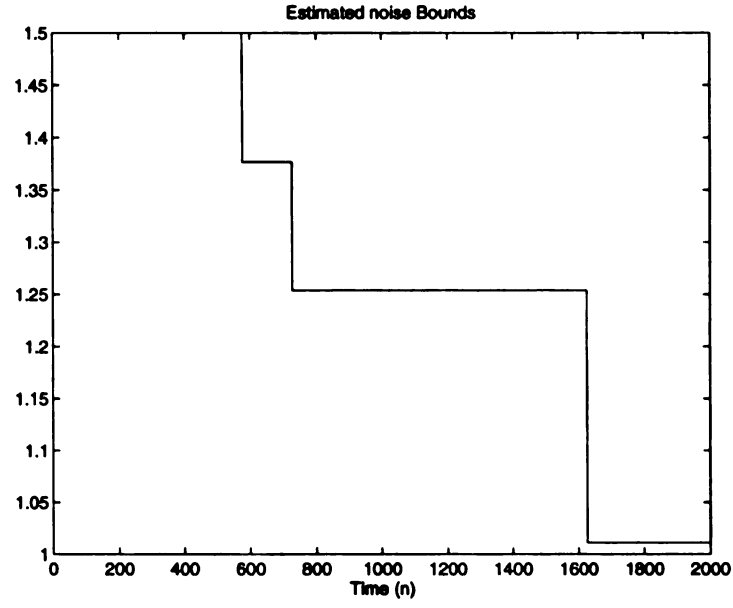


Figure 4.20: CASE 6 : v_n is colored in (4.5) for model (4.6).

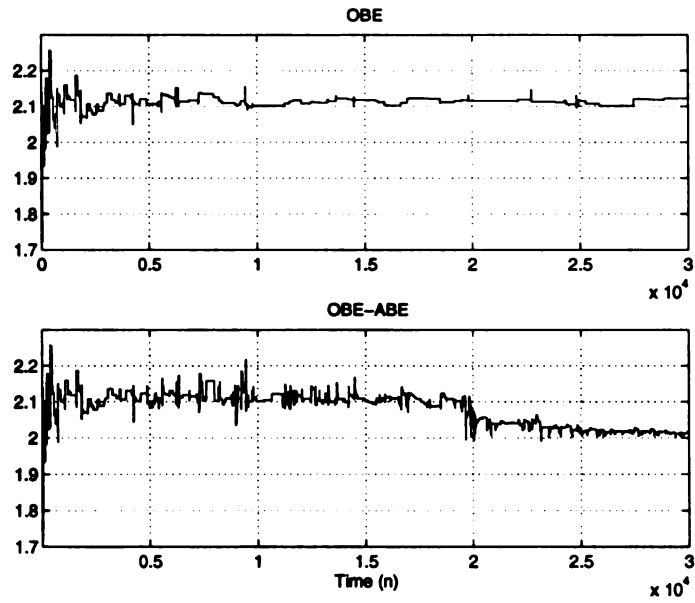


Figure 4.21: Estimators of $a_{1*} = 2$. CASE 7 : Colored noise $v_n \sim B(-0.5, 1)$, (4.7), for model (4.1).

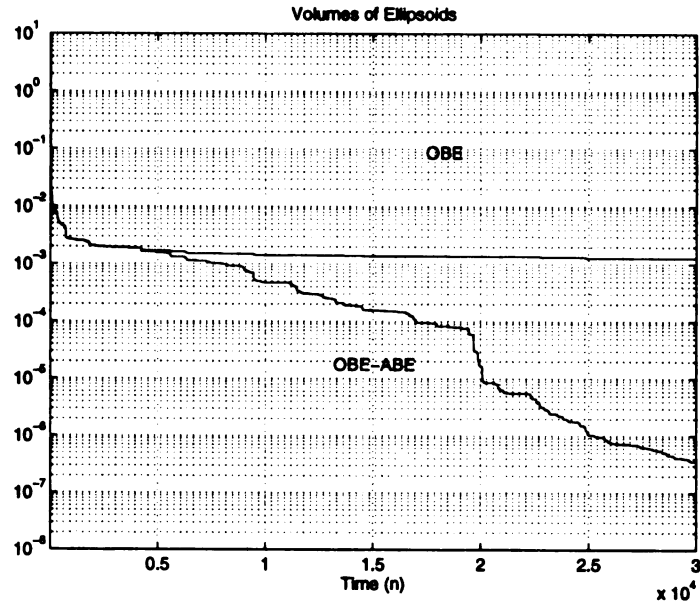


Figure 4.22: CASE 7 : Colored noise $v_n \sim B(-0.5, 1)$, (4.7), for model (4.1).

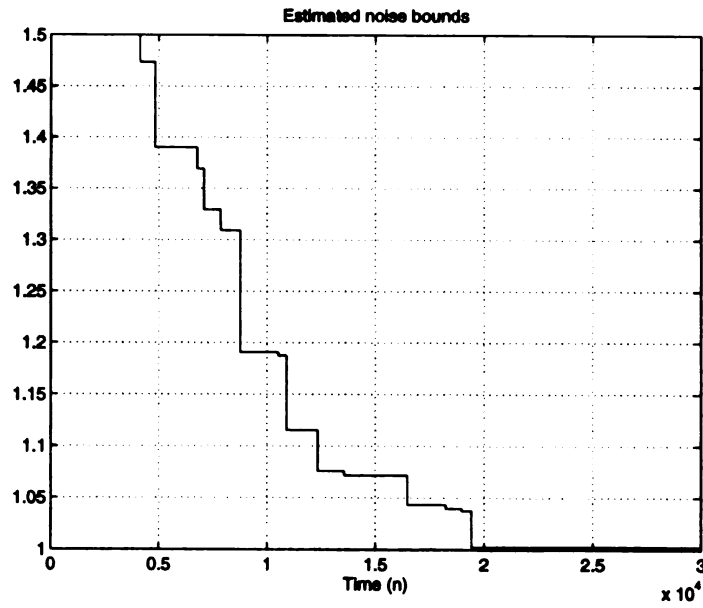


Figure 4.23: CASE 7 : Colored noise $v_n \sim B(-0.5, 1)$, (4.7), for model (4.1).

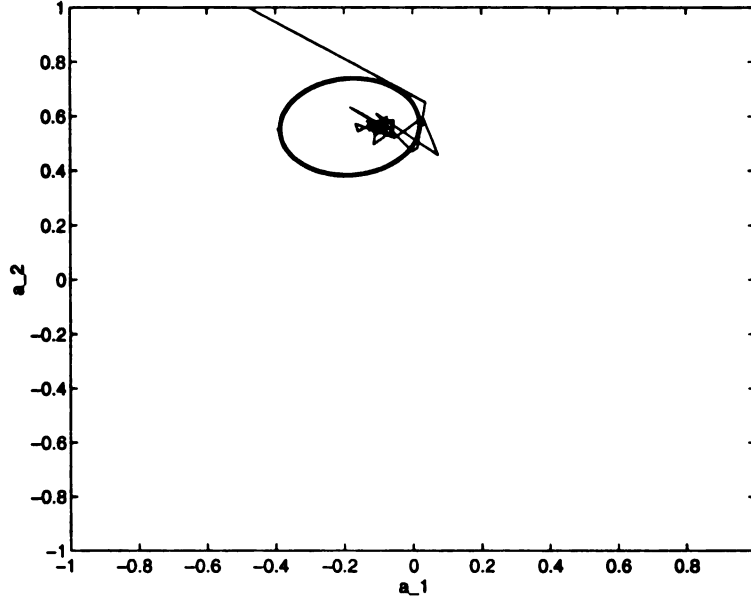


Figure 4.24: CASE 8 : The final ellipsoid and the trajectory of the estimator. $v_n \sim B(-1, 1)$, (4.5), for model (4.4).

excellent performance of the ABE procedure is not affected by the suboptimal check.

For CASE 6, the simulation results of OBE-ABE ($\gamma_0 = 1.5$, $\epsilon = 0$, and $M = 30$) and Sub-OBE-ABE ($\gamma_0 = 1.5$, $\epsilon = 0$, and $M = 30$) are shown in Figs. 4.27 and 4.28. As seen in the figures, neither estimator is affected by the non-zero mean of v_n and both algorithms converge to the true parameter ($a_{1*} = 2$). However, Sub-OBE-ABE, which selects very few (1.5%) data in this higher order case, has slower speed of convergence in 2000 steps than OBE-ABE which selects 5% of the data.

4.4 Tracking Performance of Adaptive Sub-OBE-ABE

In this section, tracking performance of Sub-OBE-ABE algorithm is illustrated through study of two examples of time-varying systems: one gradually changing, the other abruptly changing. Specifically, an AR(2) model with time-varying parameters is

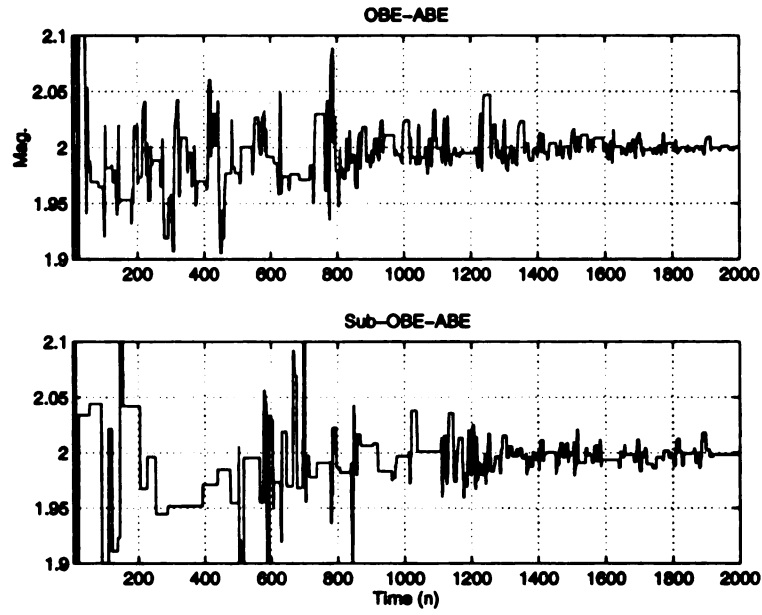


Figure 4.25: Estimators of $a_{1*} = 2$. CASE 5: v_n is colored in (4.5) for model (4.1).

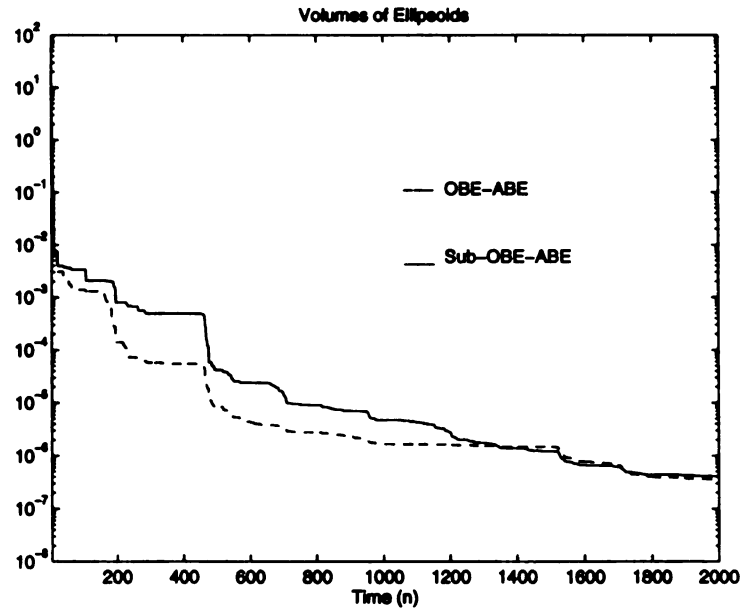


Figure 4.26: CASE 5 : v_n is colored in (4.5) for model (4.1).

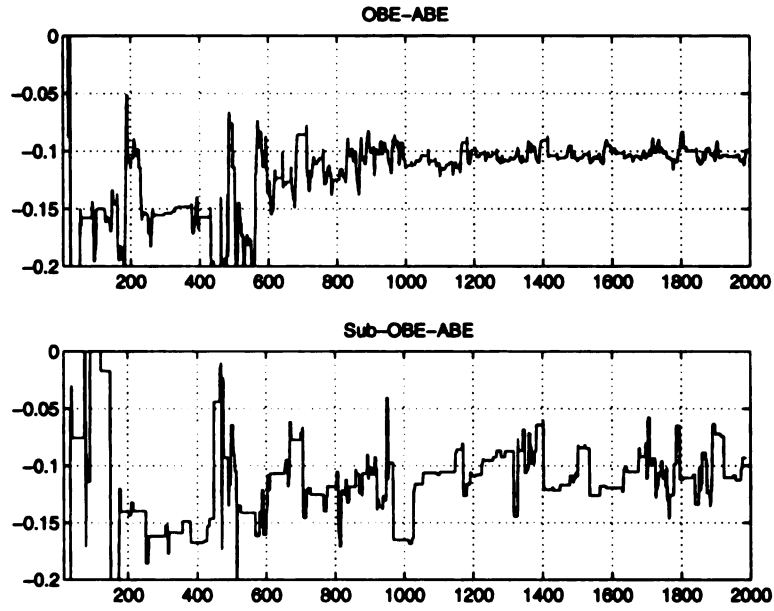


Figure 4.27: Estimators of $a_{1*} = -0.1$. CASE 6: v_n is colored in (4.5) for model (4.6).

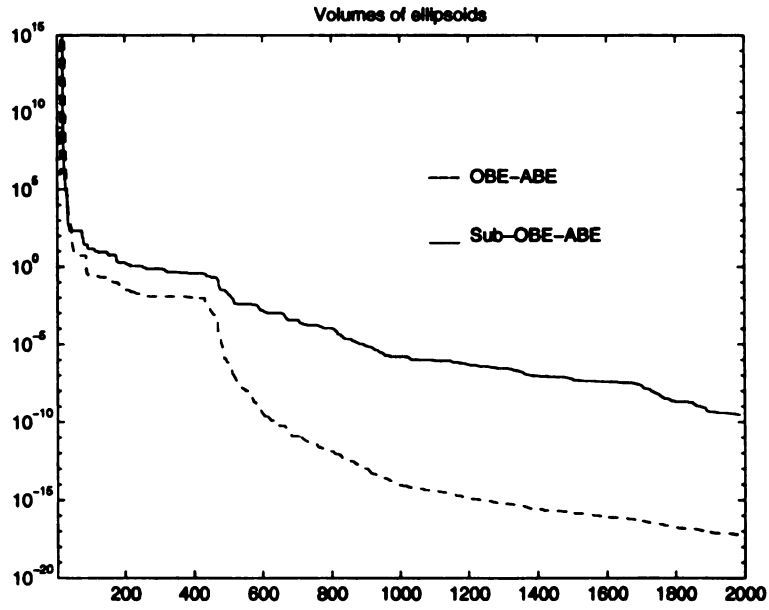


Figure 4.28: CASE 6: v_n is colored in (4.5) for model (4.6).

constructed as follows:

$$y_n = a_{1*}y_{n-1} + a_{2*}y_{n-2} + v_n \quad (4.8)$$

in which $a_{2*} = -0.68$ and a_{1*} varies between -1.6 and 1.6 (the same as in [11]). This is equivalent to varying the system's conjugate poles $0.8 \pm j0.2$ to and from $-0.8 \pm j0.2$. The variations of a_{1*} (abruptly or gradually) are shown as dashed lines in Figs. 4.31 and 4.39. The noise sequence $\{v_n\}$ is iid and uniformly distributed on $[-1,1]$.

Although “non-adaptive” OBE algorithms are well-known to have good tracking capability in slowly time-varying systems, they eventually lose tracking capability. This result is shown in Fig. 4.29 and 4.30 ($\gamma_n = 1.5$).

On the other hand, the adaptive Sub-OBE-ABE algorithm ($\gamma_0 = 1.5$, $\epsilon = 0.02$, $M = 70$), which selects 4.5% and 3% (ρ in each figure) of the data, respectively, keeps track of the varying parameters very well, as shown in Figs. 4.31 and 4.32.

To further investigate its tracking capability, faster-varying systems are simulated as in Figs. 4.33 – 4.36. As shown in the figures, the adaptive Sub-OBE-ABE algorithm ($\gamma_0 = 1.5$, $\epsilon = 0.02$, $M = 70$), which selects no more than 7.5% of the data, keeps track of these faster time-varying parameters equally well.

Comparing these diagrams with those in [11, 37, 38] shows that the adaptive Sub-OBE-ABE algorithm has better tracking capability.

The tracking capability of the Sub-OBE-ABE (or any OBE) algorithm is proportional to the tail probability of noise. This is a natural result since the larger the tail probability, the more frequent Sub-OBE-ABE updates its estimator. To demonstrate this, a much faster time-varying systems with the same uniformly distributed noise v_n is simulated as shown in Fig. 4.37. As shown in the figure, the tracking performance is not as good as in previous cases. However, if v_n is now binary Bernoulli distributed with 0.5 tail probability, the tracking performance is improved as shown in Fig. 4.38.

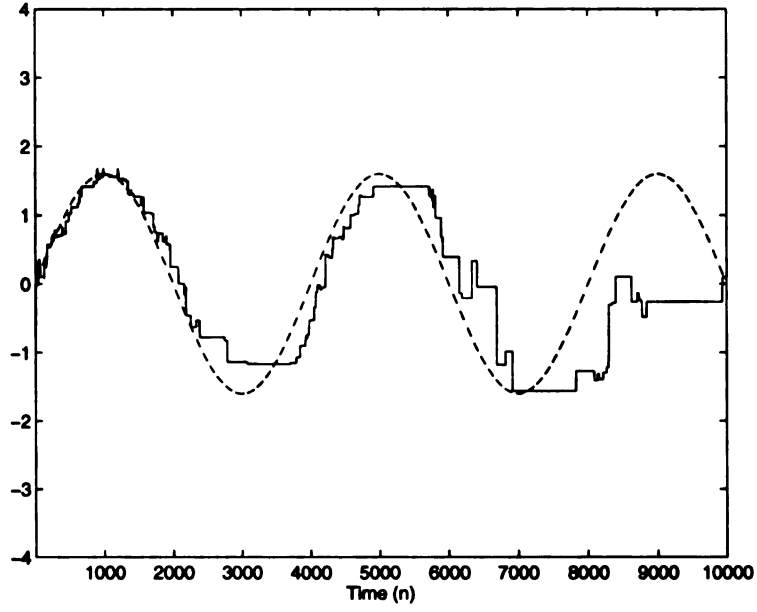


Figure 4.29: Non-adaptive OBE algorithm on the time-varying system (4.8). $\gamma_n = 1.5$.

An even faster time-varying system with the same binary Bernoulli distributed noise is shown in Fig. 4.39. As shown in the figure, the tracking performance of Sub-OBE-ABE algorithm is still excellent in this case. In these cases, Sub-OBE-ABE selects, on average, 32% of the data.

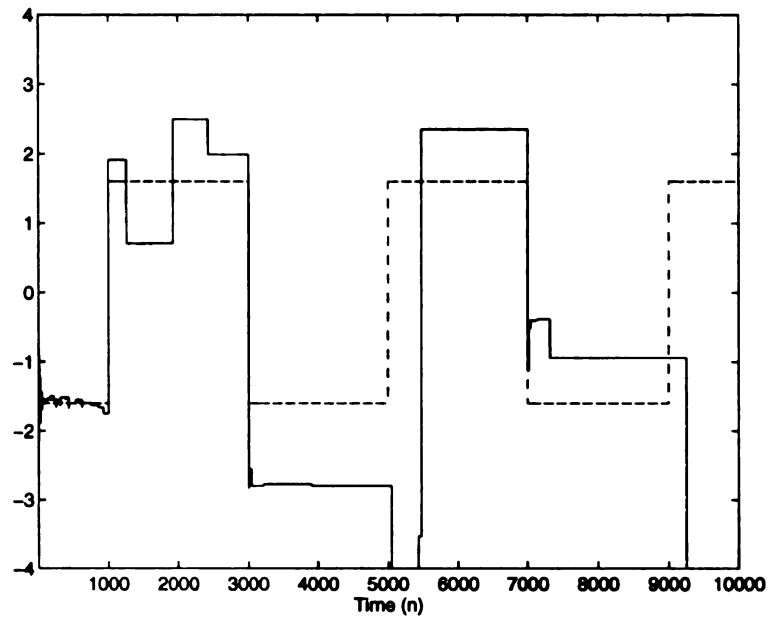


Figure 4.30: Non-adaptive OBE (SM-SA) algorithm on the time-varying system (4.8). $\gamma_n = 1.5$.

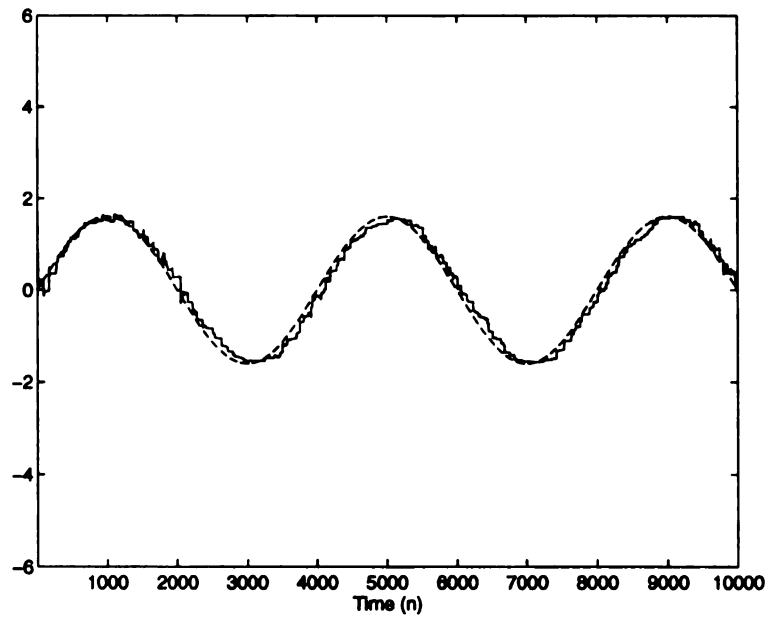


Figure 4.31: Adaptive Sub-OBE-ABE algorithm on the time-varying system (4.8). $\rho = 4.5\%$.

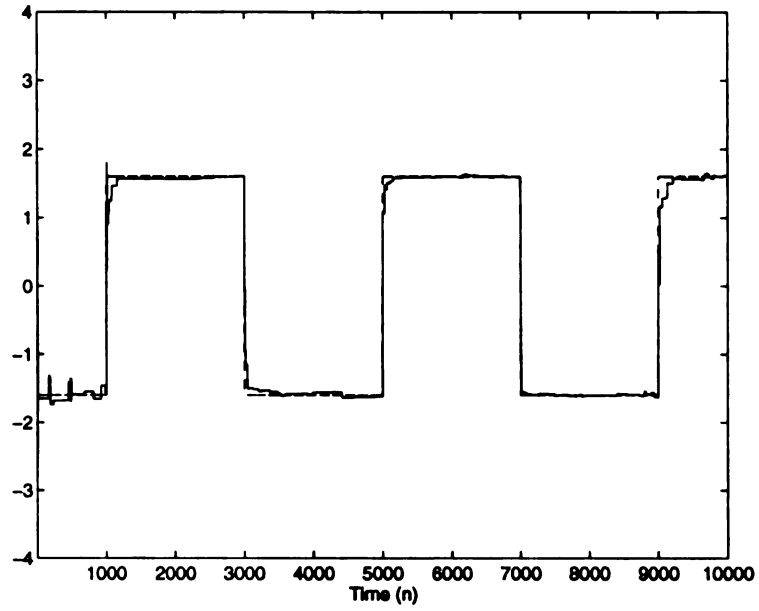


Figure 4.32: Adaptive Sub-OBE-ABE on the time-varying system (4.8). $\rho = 3\%$.

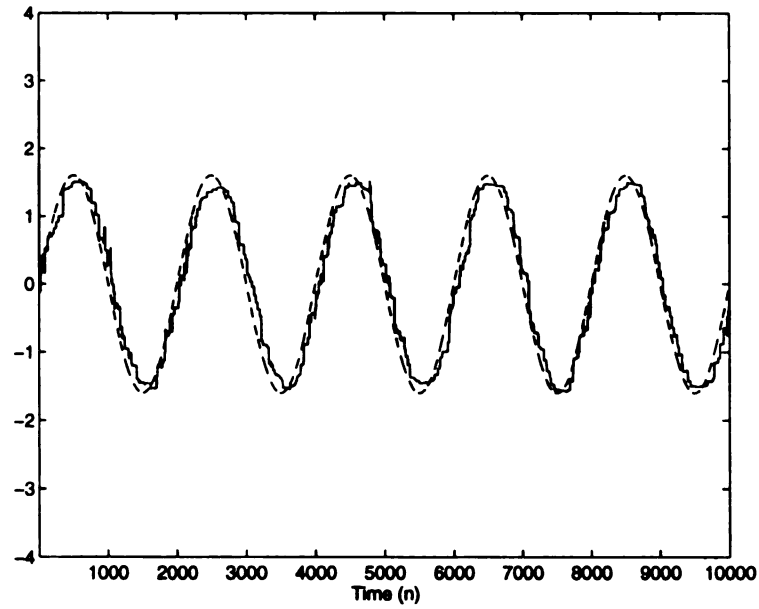


Figure 4.33: Adaptive Sub-OBE-ABE algorithm on the time-varying system (4.8). $\rho = 5.5\%$.

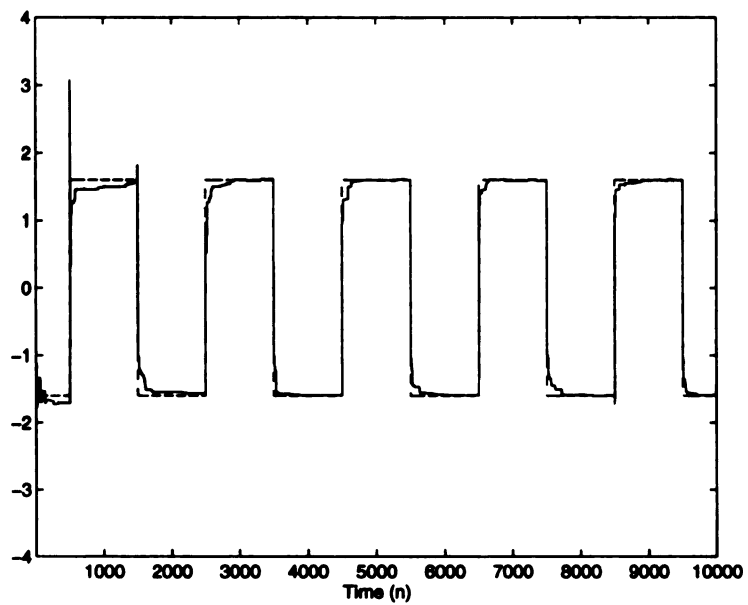


Figure 4.34: Adaptive Sub-OBE-ABE on the time-varying system (4.8). $\rho = 3.5\%$.

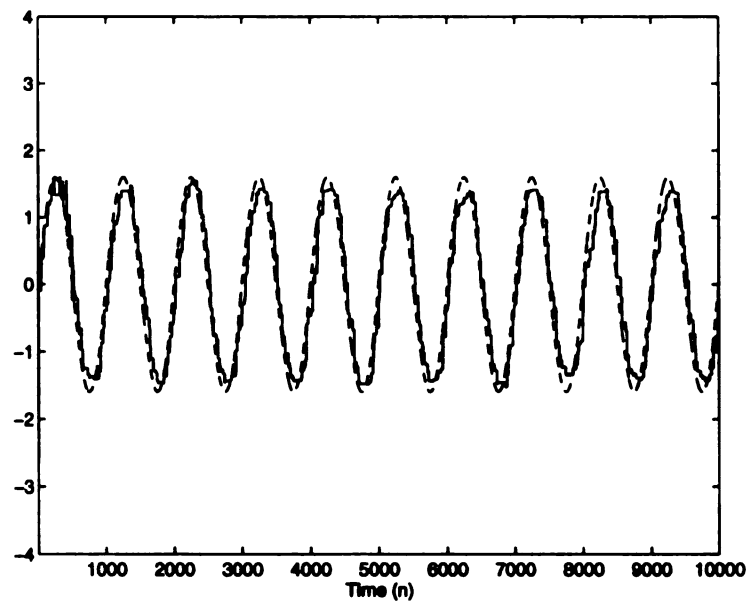


Figure 4.35: Adaptive Sub-OBE-ABE on the time-varying system (4.8). $\rho = 7.5\%$.

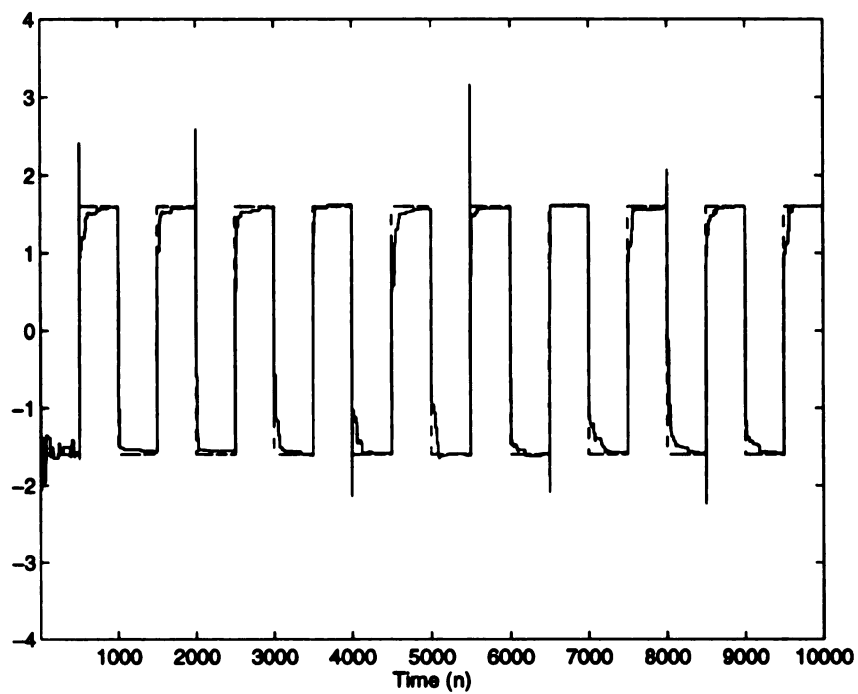


Figure 4.36: Adaptive Sub-OBE-ABE on the time-varying system (4.8). $\rho = 5.5\%$.

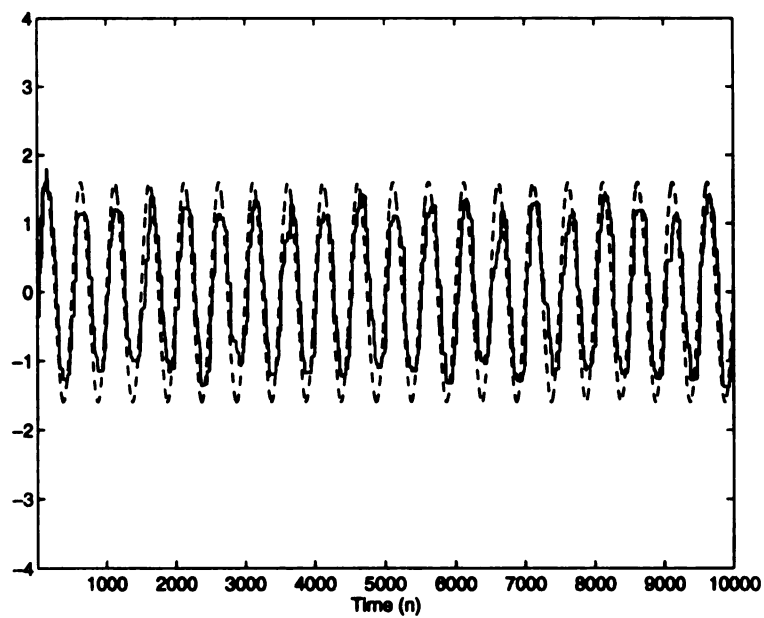


Figure 4.37: Adaptive Sub-OBE-ABE on the time-varying system (4.8). $\rho = 10\%$.

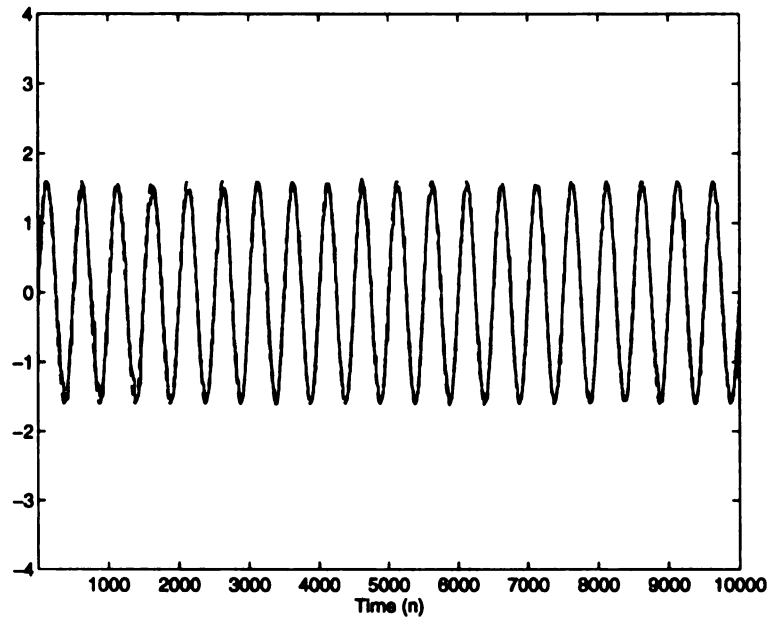


Figure 4.38: Adaptive Sub-OBE-ABE on the time-varying system (4.8) ($v_n \sim B(-1, 1)$). $\rho = 31\%$.

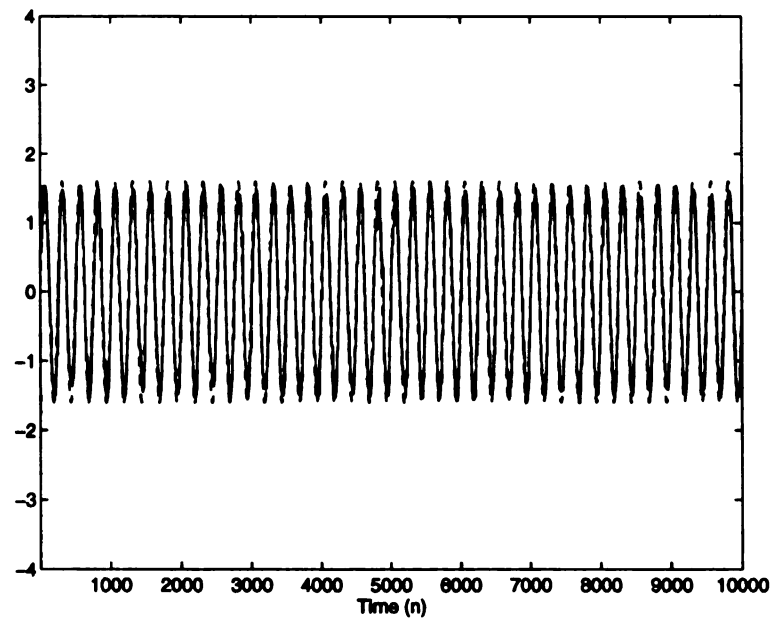


Figure 4.39: Adaptive Sub-OBE-ABE on the time-varying system (4.8) ($v_n \sim B(-1, 1)$). $\rho = 33\%$.

Chapter 5

Application to Linear Prediction Analysis of Speech

5.1 Introduction

OBE-ABE algorithm and its variants presented in the previous chapters are proven in Chap. 3 to be a.s. and p. convergent under various conditions including colored and non-stationary noise cases. These algorithms have excellent performance in simulations which are designed to satisfy the sufficient conditions of convergence. However, in real-world applications where model structures or noise characteristics are not ideal as in simulations, the performance of most algorithms suffer. In order to further investigate the performance of the new OBE-ABE algorithms in real-world applications, they are applied to the *linear prediction* (LP) analysis of speech [12, 40] in this chapter and to the blind-deconvolution problem in the next chapter.

5.2 LP model of speech

Speech is a very dynamic signal. Even a short frame (e.g., 25.6 msec) of speech can be regarded as quasi-stationary at best. However, for analytical tractability, a short

frame of 25.6 msec (256 points for 10 kHz sampling rate) speech signal is usually modeled as an AR(m) process as follows:

$$y_n = \sum_{i=1}^m a_{i*} y_{n-i} + v_n = \theta_*^T \mathbf{x}_n + v_n. \quad (5.1)$$

$\theta_*^T = [a_{1*} \cdots a_{m*}]$ is the vector of unknown LP coefficients to be identified and the *unobservable* excitation input sequence $\{v_n\}$ is a white noise for unvoiced sounds and a quasi-periodic pulse train of appropriate pitch for voiced sounds.

The LP model has been the industry standard for decades in speech analysis because of its mathematical tractability and acceptable results for processing or recognition purposes. Although *cepstral coefficients* [12] or *temporal cepstral derivative* [40] are employed in some speech recognition systems for robustness or improved performance, they are often derived from LP coefficients.

The autocorrelation method using the Levinson-Durbin algorithm [12, 40] to achieve an $\mathcal{O}(m)$ [$\mathcal{O}(3m)$ if overlapped frames are taken into account] computation has been the industry standard for identifying the LP coefficients $\{a_{n*}\}$. It is a batch method based on LSE optimization. Its recursive counterpart is the widely-used RLS. Both algorithms work very well on identifying LP coefficients as shown in the figures in the following sections. However, RLS has $\mathcal{O}(m^2)$ computational complexity since it uses 100% of the data to update the estimator.

Any recursive algorithm must converge fast enough in, for example, 256 steps in order to be successfully applied to LP analysis of speech. Known to have fast speed of convergence and efficient computation [$\mathcal{O}(\rho m^2)$], OBE algorithms (with suboptimal checking), besides RLS, is another candidate for this purpose.

5.3 LP analysis using OBE-ABE

The first attempt to applying OBE algorithm to LP analysis of speech is performed by Deller and Luk [15]. This attempt reveals the fact that the convergence performance of OBE is seriously affected by the precision of the estimated noise bounds which are not available in most real-world applications.

With the ABE procedure, the OBE-ABE algorithm makes possible the application of OBE to real-world applications including LP analysis of speech. Due to OBE-ABE's fast speed of convergence, robustness to measurement noise (see Chap. 6) and non-stationarity, and blind-deconvolution capability in colored noise, the results shown below are comparable to those of RLS or the autocorrelation method.

Theoretically, OBE-ABE can be initialized with any overestimated bound γ_0 , small ϵ , and large enough M . Since each speech frame in the LP model is typically only 256 points (for 10 KHz sampling rate) and quasi-stationary, the initialization must be deliberately chosen to assure convergence. To this purpose, several speech signals which are drawn from TIMIT (Texas Instruments and Massachusetts Institute of Technology) speech database [18] are employed in determining the optimal initialization. Also, each frame of speech is normalized to have maximum unity magnitude. However, no windowing (e.g., *Hamming window*) is needed. An acceptable initialization is $\gamma_0 = 1.5$, $\epsilon = 0.25$, and $M = 12$, for both voiced and unvoiced sounds. This makes the application of OBE-ABE straightforward, without requiring *a priori* knowledge of the voice/unvoiced status of the frame. This initialization produces acceptable results using only, on average, 12% of the data for voiced sounds and 8% for the unvoiced. By experimentation, an excellent initialization is found ($\gamma_0 = 1$, $\epsilon = 0.01$, and $M = 20$) for all the voiced and unvoiced. This initialization results in excellent spectral envelopes of LP coefficients which are comparable to those of the autocorrelation method and RLS. The data selected in this case are, on average, 25%

for voiced sounds and 60% for the unvoiced.

As typical examples of LP analysis, the first set of spectra of LP coefficients for vowels /I/ (voiced phoneme, from utterance “six”), /E/ (from “seven”), and /u/ (from “two”) are shown in Figs. 5.40 – 5.42, and for unvoiced plosive /t/ (from “two”), /f/ (from “four”), and /s/ (from “six”) are shown in Figs. 5.43 – 5.45. The upper portions of the figures show the spectra of the speech frames themselves based on a 512-point FFT. The lower portions show the spectra of LP coefficients ($m = 14$) obtained by autocorrelation method (solid line), and OBE-ABE (dashed line). As seen in the figures, OBE-ABE produces similar spectra to those of conventional batch method.

The second set of spectra of LPC coefficients for vowels /I/ (voiced phoneme, from utterance “six”), /E/ (from “seven”), and /u/ (from “two”) are shown in Figs. 5.46 – 5.48, and for unvoiced plosive /t/ (from “two”), /f/ (from “four”), and /s/ (from “six”) are shown in Figs. 5.49 – 5.51. Similar to the first set, the upper portions of the figures show the spectra of the speech frames themselves based on a 512-point FFT. The lower portions show the spectra of LP coefficients ($m = 14$) obtained by RLS (solid line), and OBE-ABE (dashed line). As seen in the figures, OBE-ABE produces similar spectra to those of RLS while selects, on average, no more than half as many data as RLS.

To further investigate OBE-ABE’s capability, a simulated speech frame (256 points) of phoneme /u/ is generated by the LP model (5.1) using a pulse train (pulses separated by 80 points, amplitude 0.5, and first pulse located at the 10th point) as the excitation $\{v_n\}$. Both OBE-ABE and autocorrelation method are applied to this case for the identification of the LP coefficients. The results are shown in Fig. 5.52. As seen in the figure, OBE-ABE produces better result since its LP spectrum is more resemblant to that of the simulated speech itself. This conforms to the fact that OBE-ABE converges consistently in colored noise case while other algorithms based

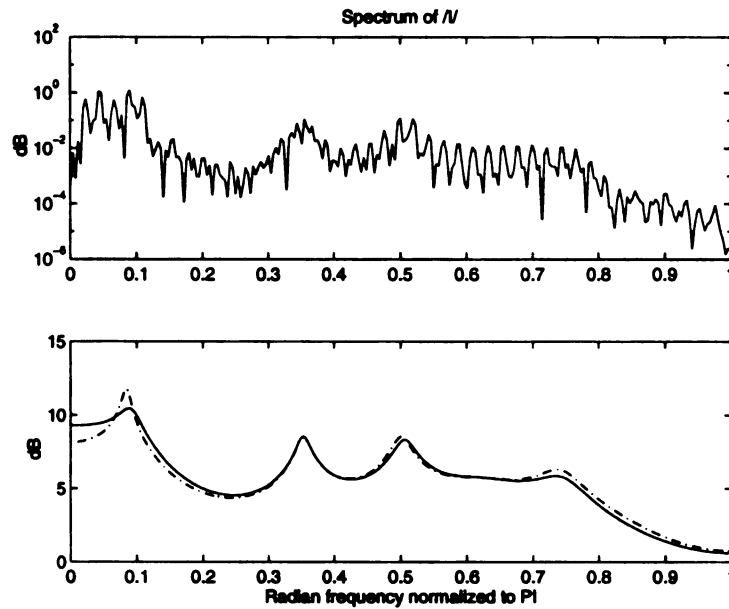


Figure 5.40: Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and autocorrelation (solid line).

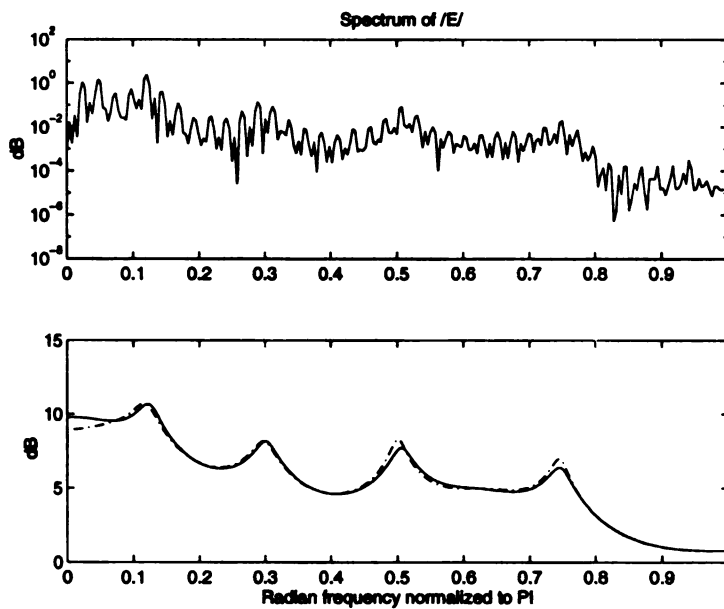


Figure 5.41: Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and autocorrelation (solid line).

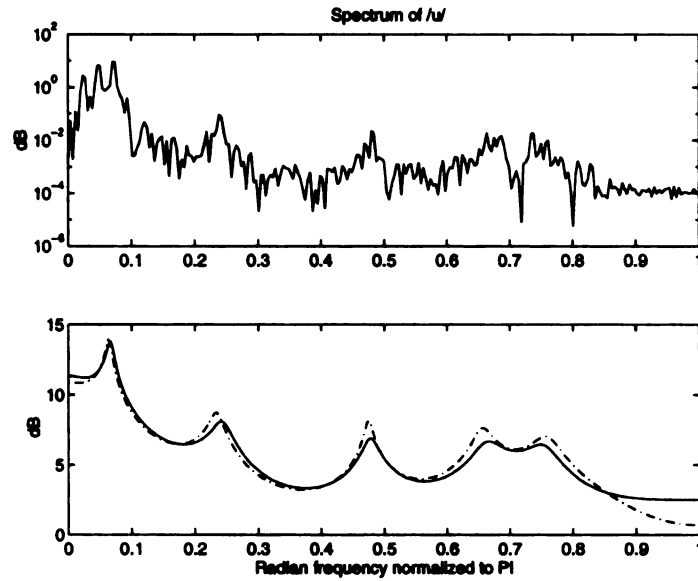


Figure 5.42: Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and autocorrelation (solid line).

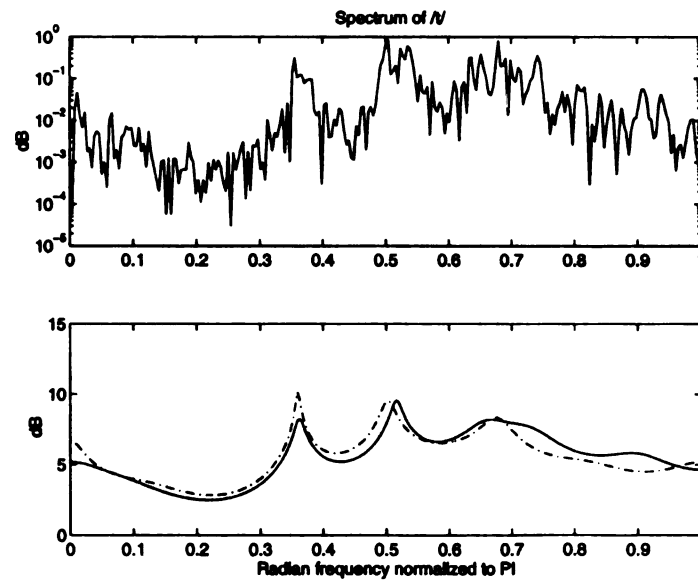


Figure 5.43: Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and autocorrelation (solid line).

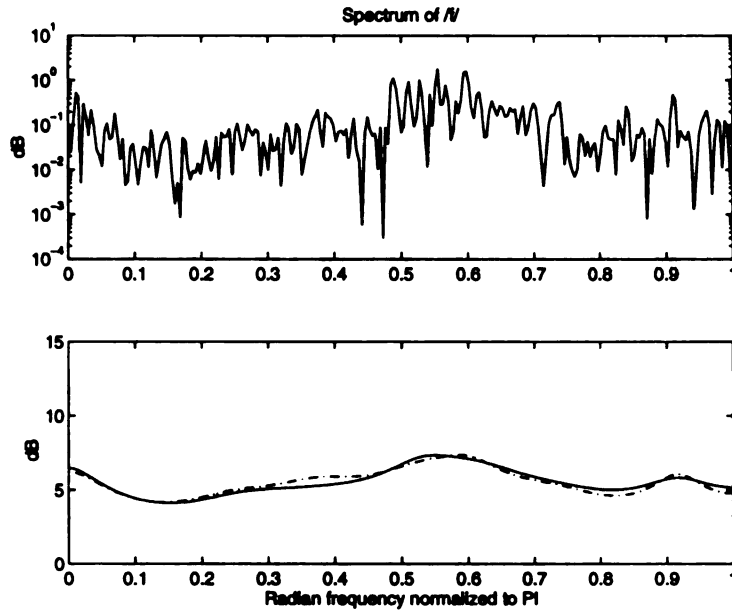


Figure 5.44: Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and autocorrelation (solid line).

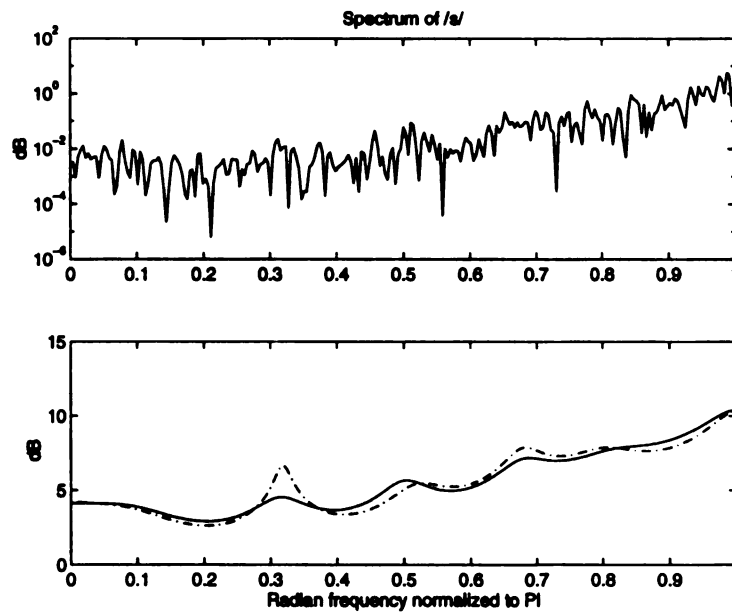


Figure 5.45: Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and autocorrelation (solid line).

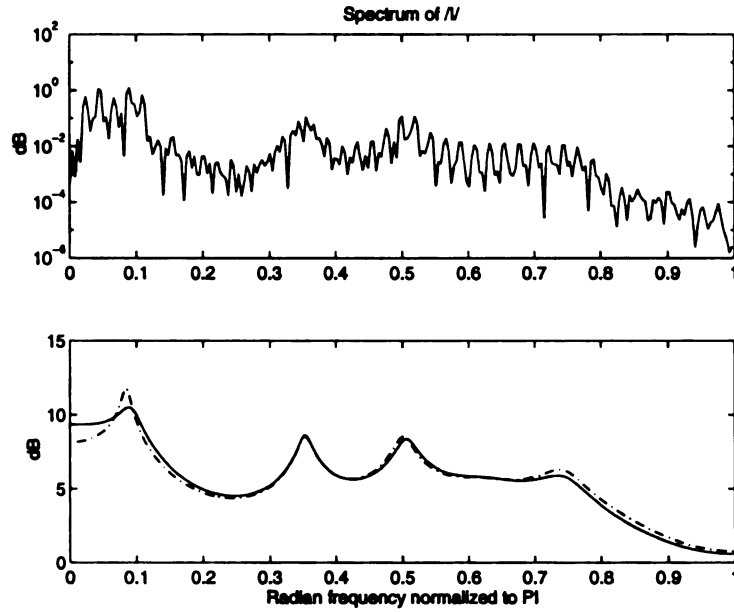


Figure 5.46: Spectra of voiced /I/ phoneme. Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and RLS (solid line).

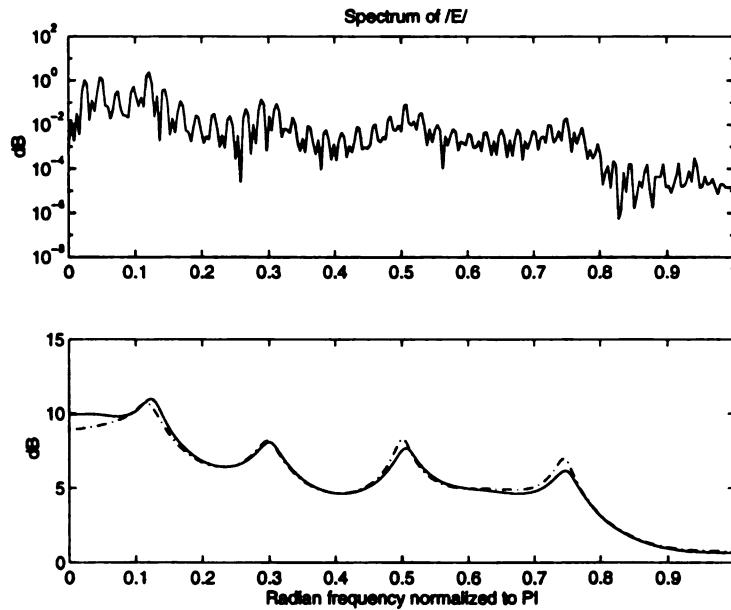


Figure 5.47: Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and RLS (solid line).

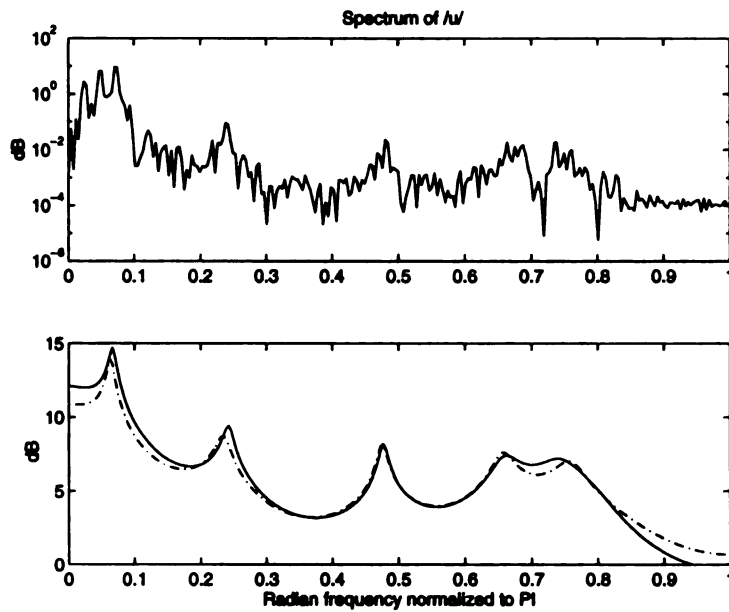


Figure 5.48: Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and RLS (solid line).

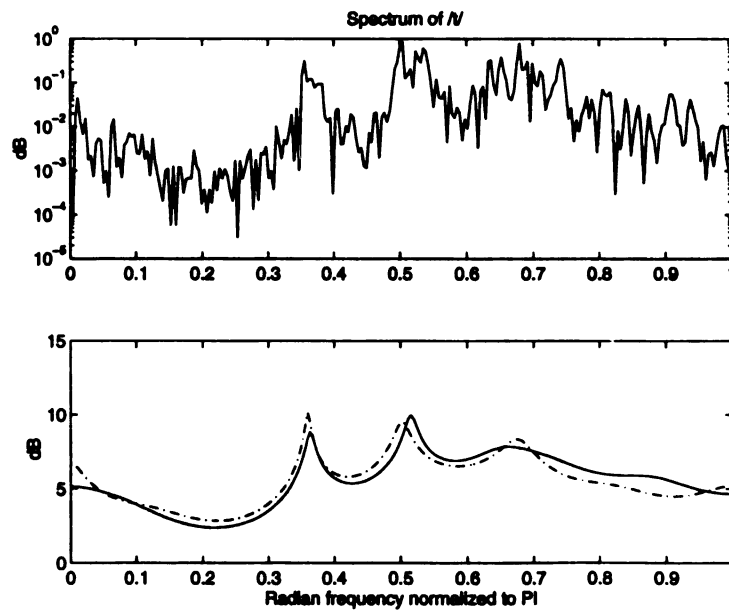


Figure 5.49: Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and RLS (solid line).

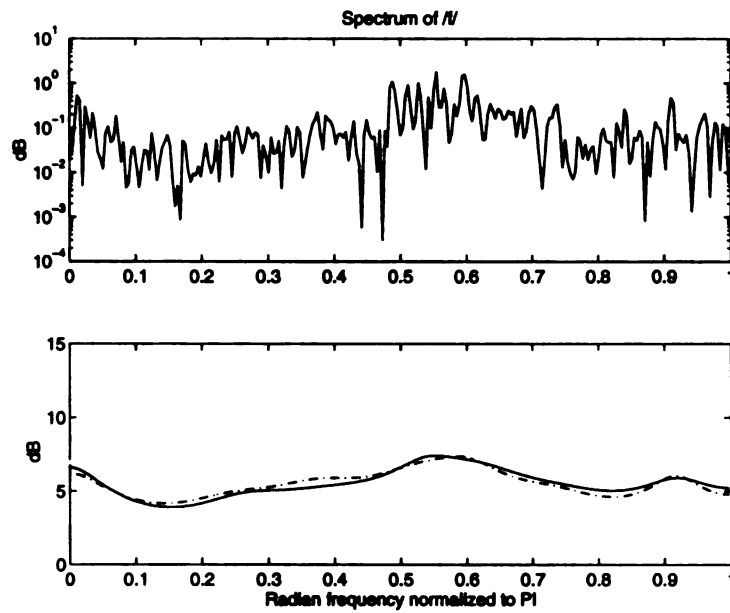


Figure 5.50: Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and RLS (solid line).

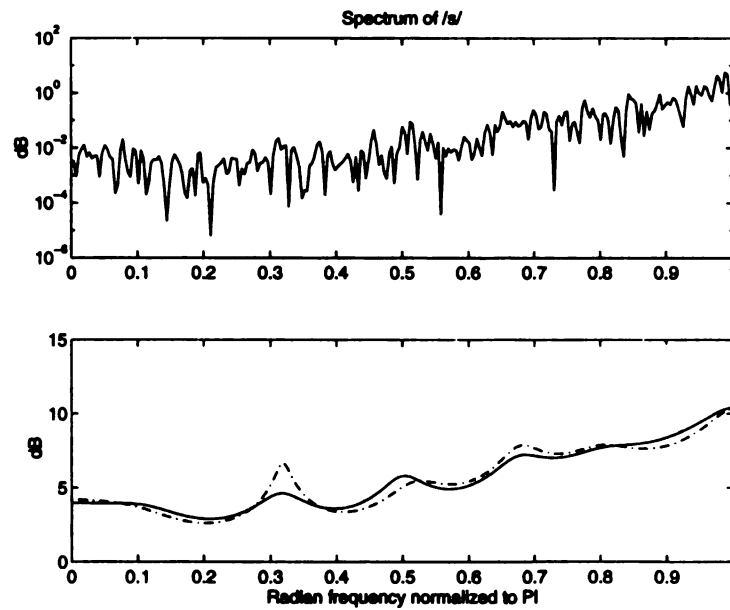


Figure 5.51: Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and RLS (solid line).

on LS concept (e.g., LS, RLS, LMS) are inconsistent.

5.4 Sub-OBE-ABE and computational complexity

The Sub-OBE-ABE algorithm has been shown in Chap. 4 to have similar speed of convergence to OBE-ABE while having more efficient $\mathcal{O}(\rho m^2)$ computation. However, within short periods of 256 steps, Sub-OBE-ABE does not converge as well as OBE-ABE with the same initialization, due to the suboptimal check for innovation. This phenomenon has been explained in Chap. 4 for Fig. 4.12 and 4.14.

Through experiments, two set of initializations of Sub-OBE-ABE algorithm are found respectively for the voiced and unvoiced sounds. For voiced sounds, OBE-ABE initialized with $\gamma_0 = 0.25$, $\epsilon = 0.0001$, and $M = 12$ has excellent results as shown in Figs. 5.53 – 5.55. Seven percent of the data is selected in the voiced case on average. For unvoiced sounds, the initialization is $\gamma_0 = 0.5$, $\epsilon = 0.0001$, and $M = 8$. The results for unvoiced sounds are shown in Figs. 5.56 – 5.58. Thirteen percent of the data is selected in the unvoiced case on average.

As seen in the figures, Sub-OBE-ABE achieves similar spectra to those of RLS or the autocorrelation method while selecting, on average, only 10% of the data. Since the number of LPC coefficients, m , is on the order of 10 [12, 40] for most speech applications, Sub-OBE-ABE is virtually an $\mathcal{O}(m)$ recursive algorithm in this application.

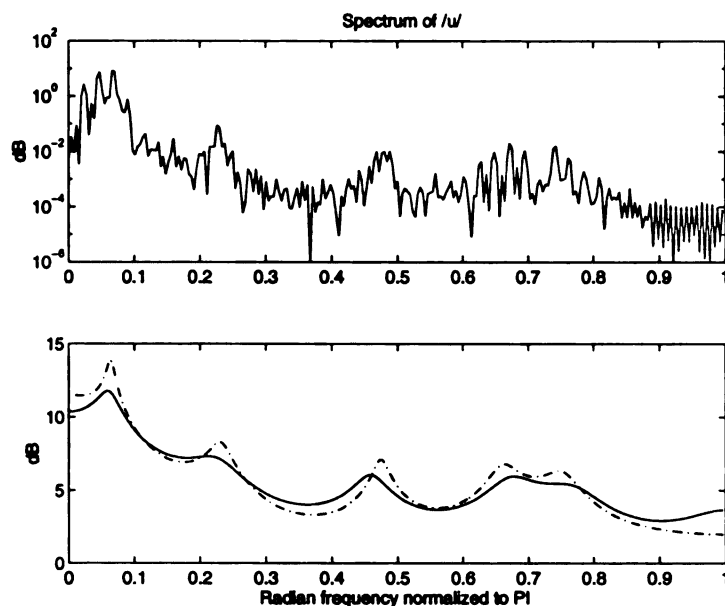


Figure 5.52: Spectra of simulated voiced /u/ phoneme. Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on OBE-ABE (dashed line), and autocorrelation method (solid line).

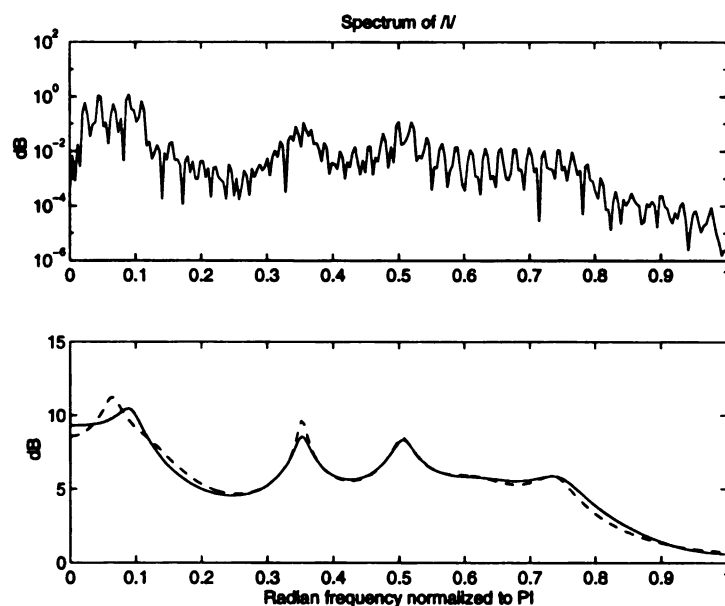


Figure 5.53: Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on Sub-OBE-ABE (dashed line), and autocorrelation (solid line).

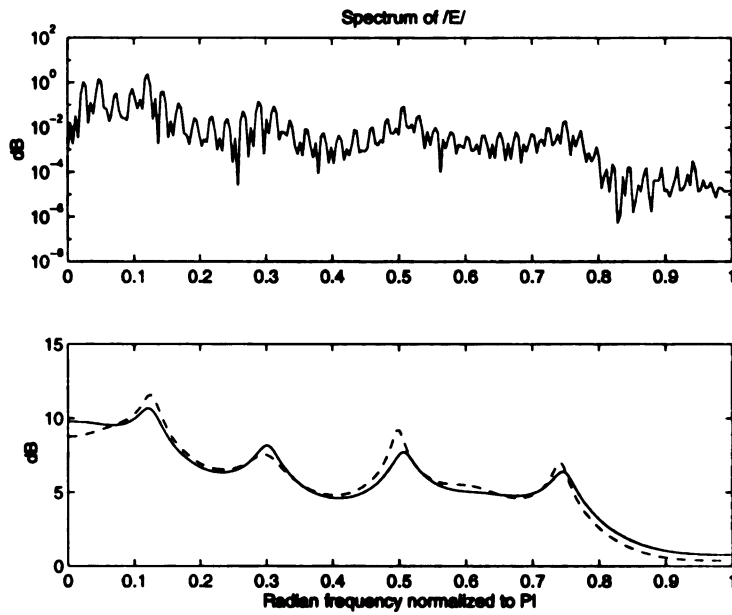


Figure 5.54: Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on Sub-OBE-ABE (dashed line), and autocorrelation (solid line).

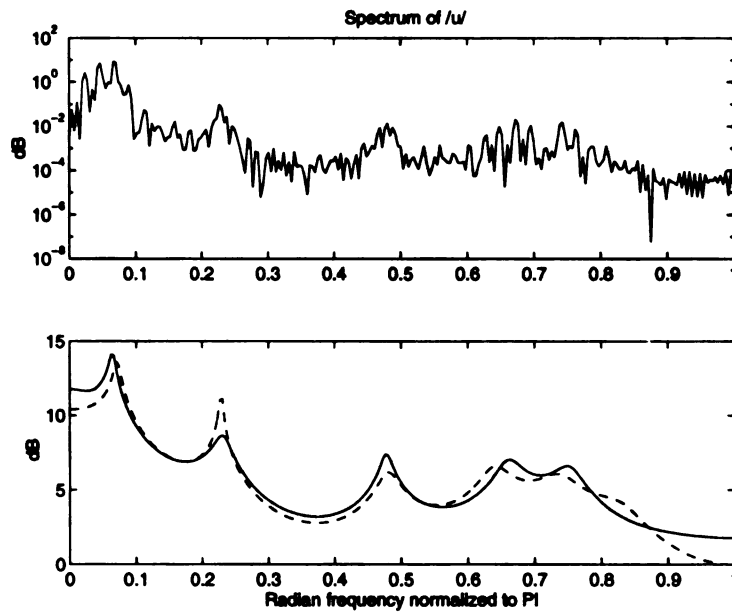


Figure 5.55: Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on Sub-OBE-ABE (dashed line), and autocorrelation (solid line).

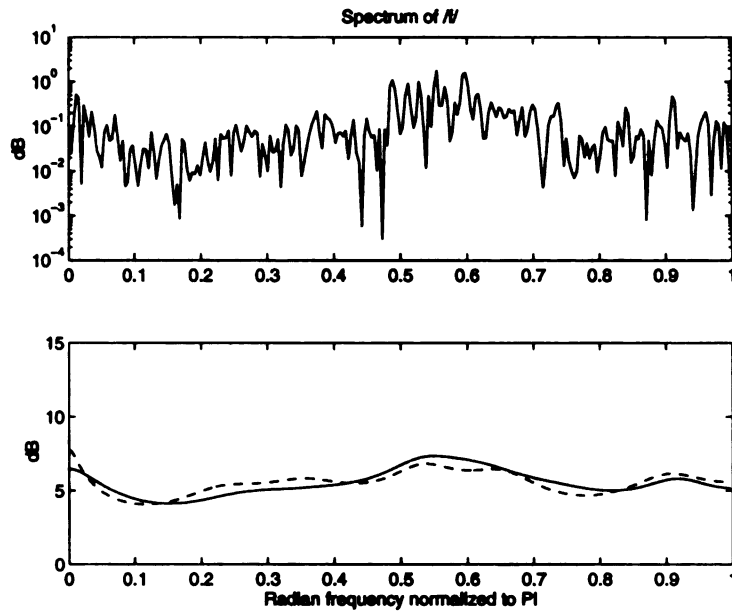


Figure 5.56: Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on Sub-OBE-ABE (dashed line), and autocorrelation (solid line).

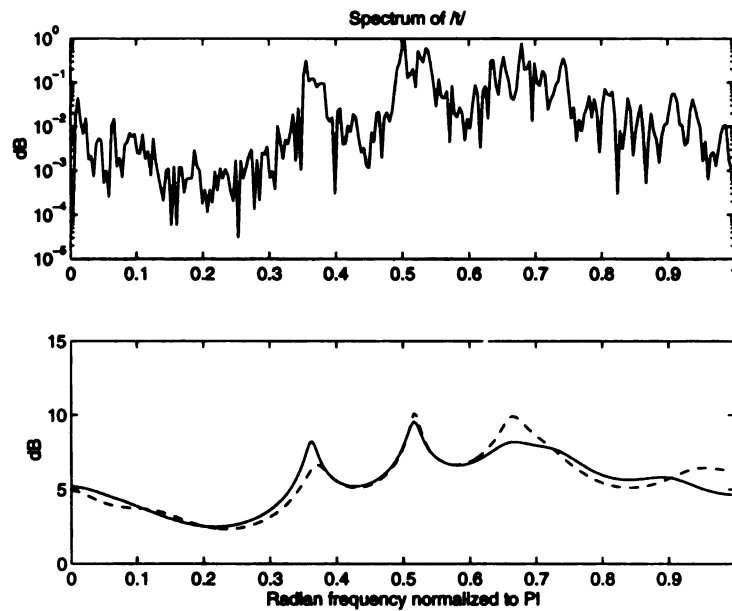


Figure 5.57: Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on Sub-OBE-ABE (dashed line), and autocorrelation (solid line).

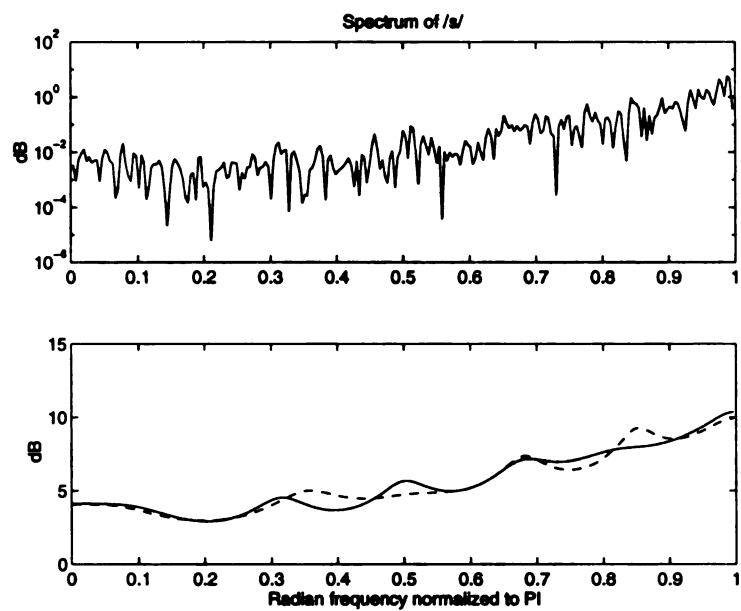


Figure 5.58: Upper graph: Direct 512-point FFT spectrum. Lower graph: Spectra based on Sub-OBE-ABE (dashed line), and autocorrelation (solid line).

Chapter 6

Application to Blind Deconvolution

6.1 Introduction

The blind-deconvolution problem [23] has attracted intensive research in recent years in the fields of data communications, adaptive filtering, and signal processing. A typical block diagram for the blind-deconvolution problem is shown in Fig. 6.1. The input $\{v_n\}$ and the noise $\{w_n\}$ in the figure are assumed unobservable. In data communications, the unknown input data are recovered using the output data received from a noisy and distorted channel (e.g., satellite, cellular or optical fiber). Another example of a blind-deconvolution problem is the LP analysis of a speech signal corrupted by unknown measurement or environmental noise.

Several well-known estimation methods have been applied to solve the blind-deconvolution problem. Bayesian estimation [23] (batch method) and LMS (recursive) are two popular methods. However, both of those methods require that the input $\{v_n\}$ and the noise $\{w_n\}$ be white for convergence since they are based on the LSE optimization. This seriously restricts the application scope since input data ($\{v_n\}$) in most engineering applications (e.g., the previous two examples) are not always white.

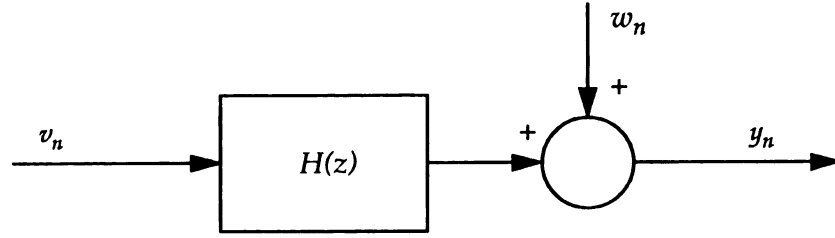


Figure 6.1: Model of the received signal in the blind-deconvolution problem.

The new OBE-ABE algorithms presented in this dissertation are good candidates to solve this difficulty since they converge in colored noise cases as demonstrated in the previous chapters. Although the model of blind deconvolution shown in Fig. 6.1 is not an ARX model, OBE-ABE algorithms show excellent results in this application due to their robustness to the measurement noise.

The block diagram in Fig. 6.2 is one of the application structures for blind deconvolution and is employed in this chapter. The parameters of the inverse filter $\hat{H}^{-1}(z)$ are estimated by OBE-ABE using output data $\{y_n\}$. The detector in the figure has known transfer functions for the recovery of the input data using $\{\tilde{v}_n\}$. The unknown $H(z)$ is simulated as IIR (*infinite impulse response*) and FIR (*finite impulse response*) filters, respectively. The unknown input $\{v_n\}$ is simulated as an iid or a colored digital data stream. The noise $\{w_n\}$ is independent of $\{v_n\}$ and is iid with uniform distribution specified in the following sections.

6.2 IID input and IIR filter

In this section, the IIR filter in Fig. 6.2 is assumed to be:

$$H(z) = 1/(a_{1*}z^{-1} + a_{2*}z^{-2} + a_{3*}z^{-3}) \quad (6.1)$$

where the unknown parameters $a_{1*} = 2, a_{2*} = -1.68, a_{3*} = 0.34$, the input $\{v_n\}$ is iid with equally distributed discrete values $\{-1, -0.5, 0, 0.5, 1\}$, and the noise $\{w_n\}$ is iid and uniformly distributed on $[-0.05, 0.05]$. In this case, the *signal to noise ratio* (S/N) [ignoring the gain of $H(z)$] is 42.5 dB. The order of the inverse filter $\hat{H}^{-1}(z)$ in this case is three. The application results of OBE-ABE and conventional OBE algorithms in this case are shown in Fig. 6.3 and 6.4, respectively. As seen in Fig. 6.4, the ellipsoidal volumes of OBE algorithm ($\gamma_n = 1$) quickly become negative because of the added noise $\{w_n\}$, while the volume of the OBE-ABE algorithm ($\gamma_0 = 10, \epsilon = 0.5$, and $M = 80$) converges to zero. Figure 6.3 shows the estimator of the first parameter both algorithms. As seen in the figure, the OBE estimate diverges due to the negativity of ellipsoidal volumes, while OBE-ABE algorithm converges to the theoretical values ($[a_1, a_2, a_3] = [1.9930, -1.4676, 0.3332]$) calculated by the Wiener optimization method [23]. Simulations show that OBE-ABE algorithm is robust to the noise $\{w_n\}$ while OBE is not.

Simulations also show that, using OBE-ABE in this case, the output of the detector, $\{\hat{v}_n\}$, recovers the unknown input data $\{v_n\}$ with a 99% success rate. Figure 6.5 shows the first 100 samples of v_n , $\tilde{v}_n$, and $\hat{v}_n$. The 1% wrong data can be easily recovered using some error-correction techniques in data communications.

Simulations also show that the performance of OBE-ABE algorithm in the noise-corrupted case is not affected by the order of $H(z)$. However, when the S/N decreases (e.g. increases the bound of the noise $\{w_n\}$), the success rate decreases due to the

biased estimator which can be theoretically calculated by the Wiener optimization method. This can be seen in Figs. 6.6 and 6.7 which are the simulation results of the above model except that now $w_n \sim U(-0.1, 0.1)$ ($S/N = 36.5$ dB). As seen in the figures, the estimator of OBE-ABE is biased (theoretical values of the estimator in this case are $[a_1, a_2, a_3] = [1.9724, -1.4315, 0.3133]$), and the success rate is 91%.

6.3 IID input and FIR filter

In this section, the FIR filter in Fig. 6.2 is assumed to be:

$$H(z) = a_{1*}z^{-1} + a_{2*}z^{-2} + a_{3*}z^{-3}$$

where the unknown parameters $a_{1*} = 0.7, a_{2*} = 0.5, a_{3*} = -0.336$, the input $\{v_n\}$ is iid with equally distributed discrete values $\{-1, -0.5, 0, 0.5, 1\}$, and the noise $\{w_n\}$ is iid and uniformly distributed on $[-0.05, 0.05]$ ($S/N = 42.5$ dB) and $[-0.1, 0.1]$ ($S/N = 36.5$ dB), respectively. In this FIR case, the performance of OBE-ABE algorithm is still excellent if the order of the inverse filter $\hat{H}^{-1}(z)$ increases from 3 to 10. The results are shown in Figs. 6.8 – 6.9. As seen in the figures, OBE-ABE shows excellent performance with 99% success rate in the case of $S/N = 42.5$ dB (Fig. 6.8) and 93% success rate in the case of $S/N = 36.5$ dB (Fig. 6.9).

6.4 Colored Input and IIR Filter

In this section, the IIR filter in Fig. 6.2, $H(z)$, is the same as in (6.1) and the input $\{v_n\}$ is colored as in (4.5), and the noise $\{w_n\}$ is iid and uniformly distributed on $[-0.1, 0.1]$ ($S/N = 36.5$ dB) or $[-0.5, 0.5]$ ($S/N = 22.5$ dB). The application results of OBE-ABE in these cases are shown in Figs. 6.10 and 6.11. As seen in the figures,

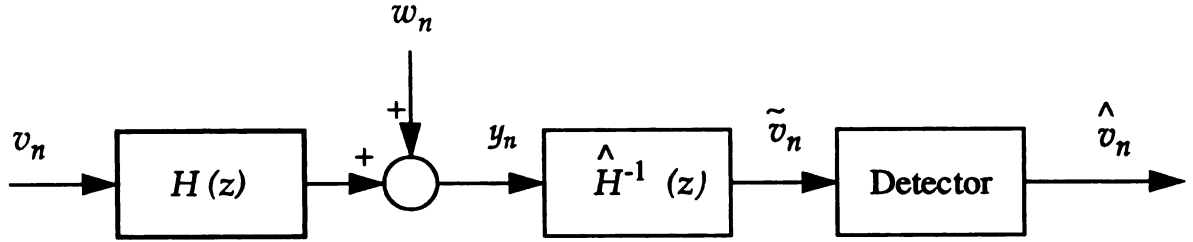


Figure 6.2: Block diagram of the blind-deconvolution problem.

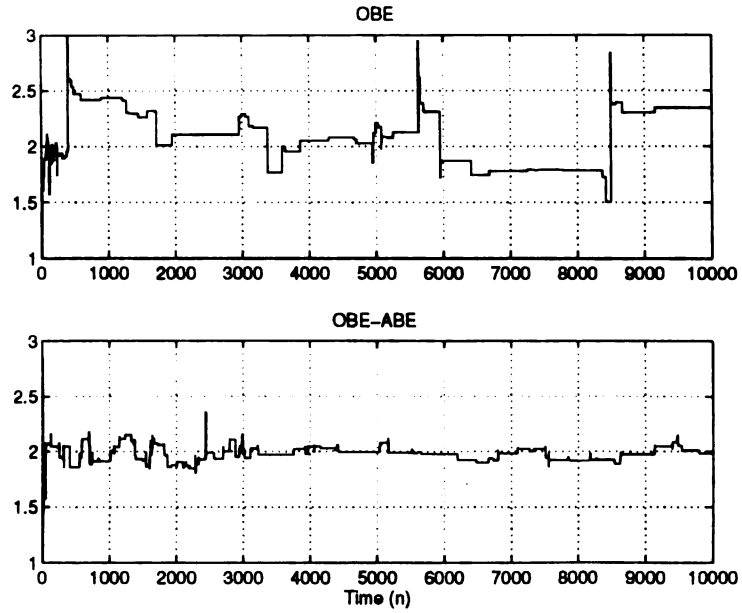


Figure 6.3: Estimates of a_{1*} of OBE ($\gamma_n = 1$) and OBE-ABE ($\gamma_0 = 10$) for Fig. 6.2 (iid-IIR, $S/N = 42$ dB).

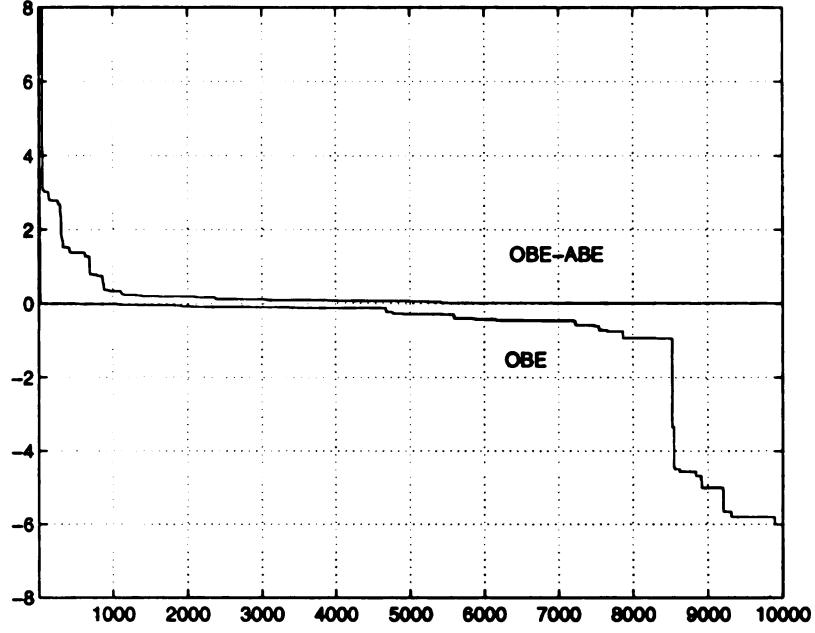


Figure 6.4: Volumes of ellipsoids of OBE ($\gamma_n = 1$) and OBE-ABE ($\gamma_0 = 10$) for Fig. 6.2 (iid-IIR, $S/N = 42$ dB).

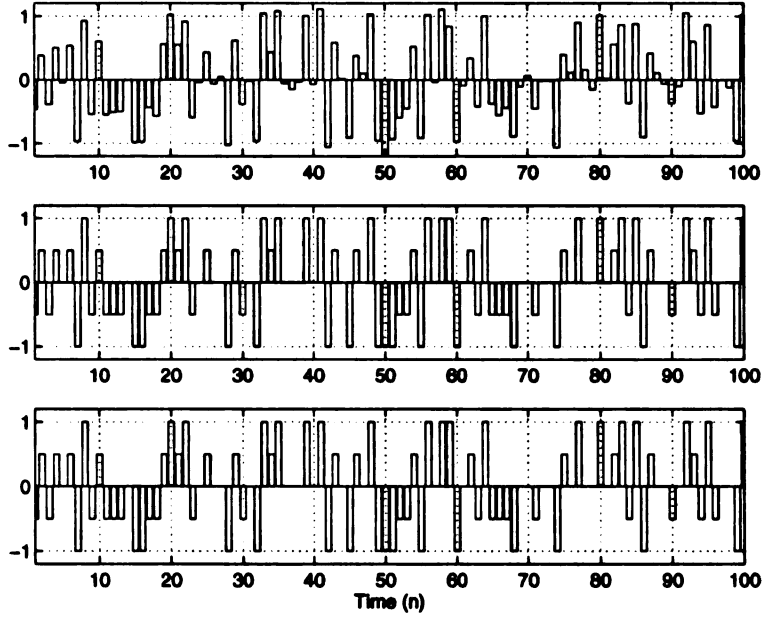


Figure 6.5: First 100 samples of $\tilde{v}_n$ (top), $\hat{v}_n$ (middle), and v_n (bottom) using OBE-ABE ($\gamma_0 = 10$) in Fig. 6.2 (iid-IIR, $S/N = 42.5$ dB).

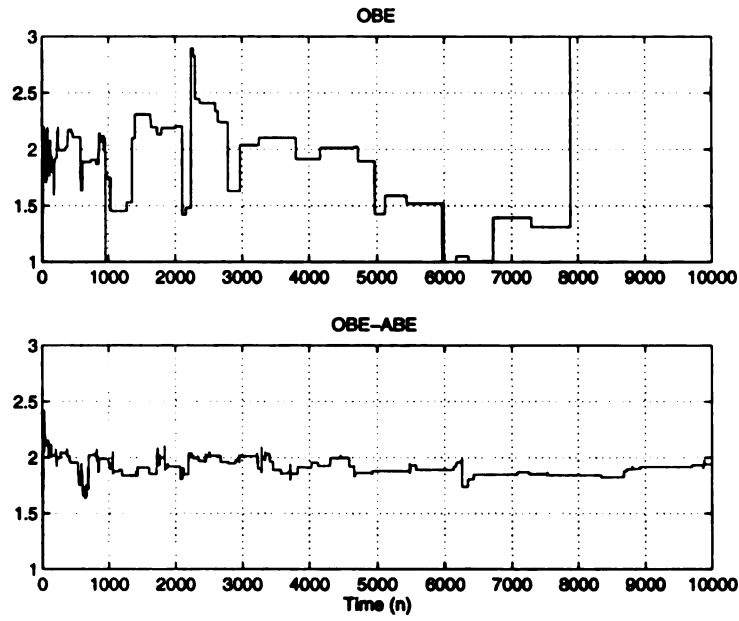


Figure 6.6: Estimates of a_{1*} of OBE ($\gamma_n = 1$) and OBE-ABE ($\gamma_0 = 10$) for Fig. 6.2 (iid-IIR, $S/N = 36.5$ dB).

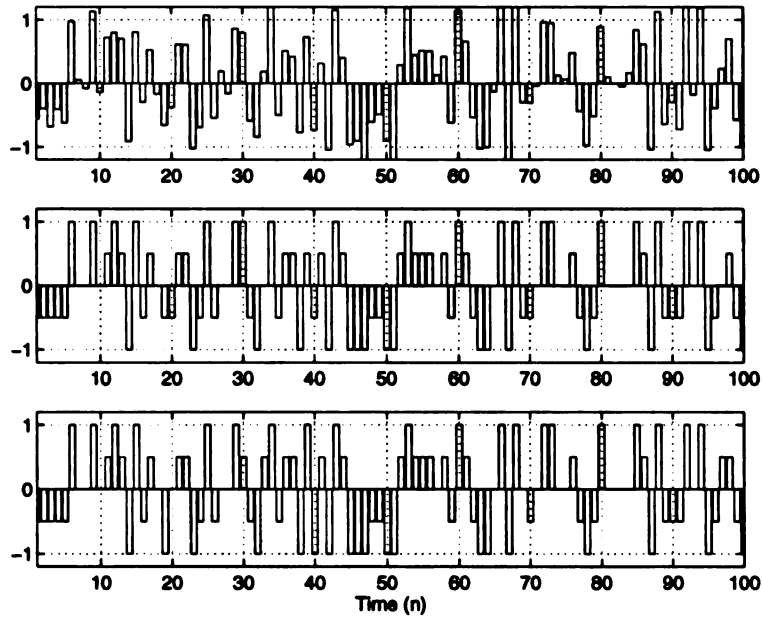


Figure 6.7: First 100 samples of $\tilde{v}_n$ (top), $\hat{v}_n$ (middle), and v_n (bottom) using OBE-ABE ($\gamma_0 = 10$) in Fig. 6.2 (iid-IIR, $S/N = 36.5$ dB).

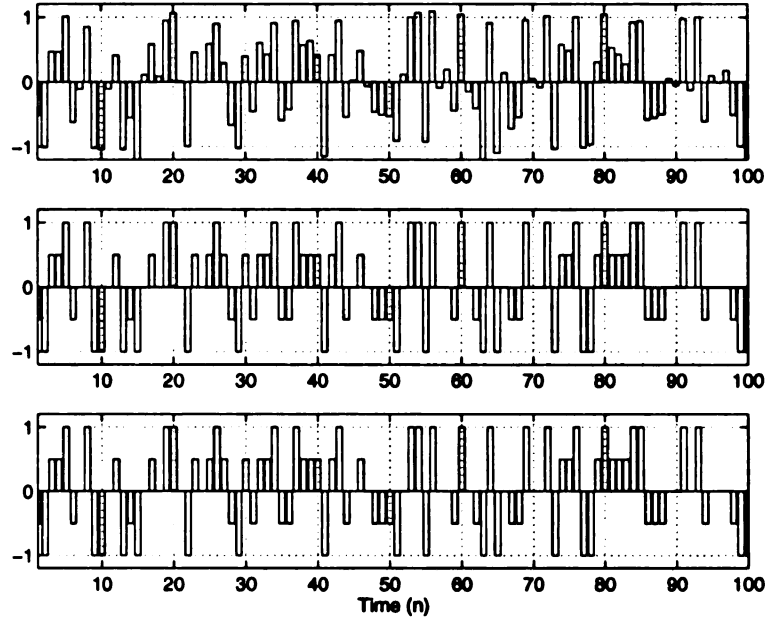


Figure 6.8: First 100 samples of $\tilde{v}_n$ (top), $\hat{v}_n$ (middle), and v_n (bottom) using OBE-ABE ($\gamma_0 = 10$) in Fig. 6.2 (iid-FIR, $S/N = 42.5$ dB).

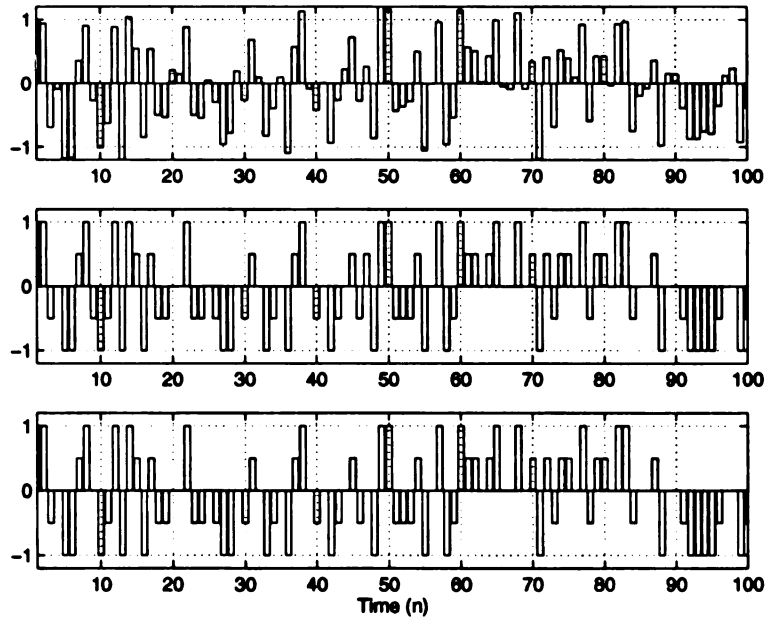


Figure 6.9: First 100 samples of $\tilde{v}_n$ (top), $\hat{v}_n$ (middle), and v_n (bottom) using OBE-ABE ($\gamma_0 = 10$) in Fig. 6.2 (IID-FIR, $S/N = 36.5$ dB).

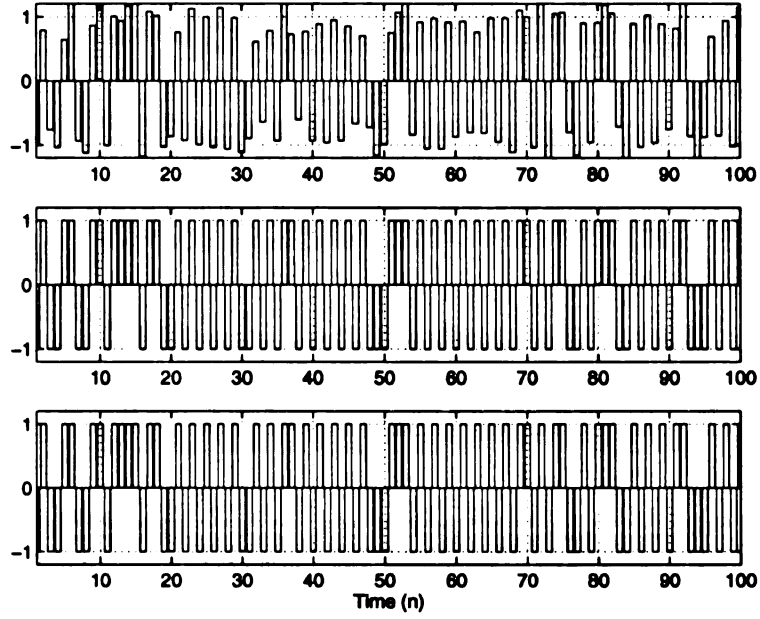


Figure 6.10: First 100 samples of $\tilde{v}_n$ (top), $\hat{v}_n$ (middle), and v_n (bottom) using OBE-ABE ($\gamma_0 = 10$) in Fig. 6.2 (Colored-IIR, $S/N = 36.5$ dB).

OBE-ABE shows excellent performance, achieving 100% success rate in the case of $S/N = 36.5$ dB (Fig. 6.10) and 90% success rate in the case of $S/N = 22.5$ dB (Fig. 6.11). This demonstrates the excellent blind-deconvolution performance of OBE-ABE algorithm in the colored-input case.

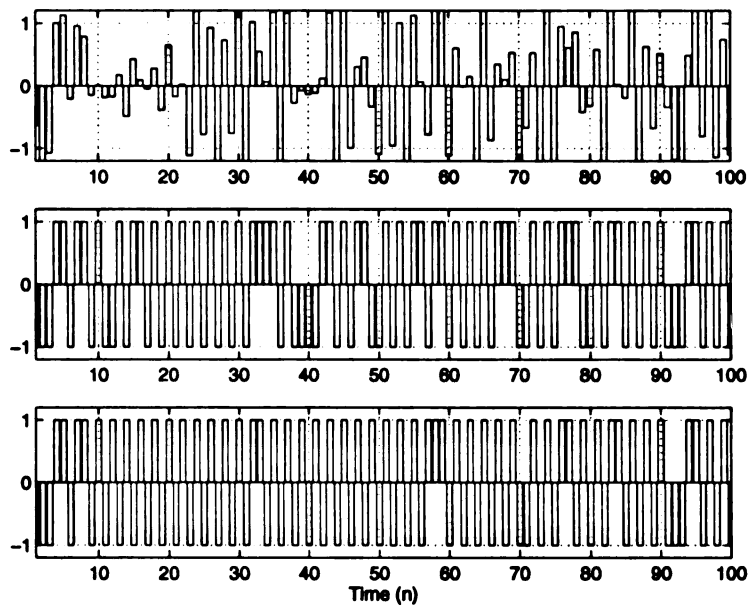


Figure 6.11: First 100 samples of $\tilde{v}_n$ (top), $\hat{v}_n$ (middle), and v_n (bottom) using OBE-ABE ($\gamma_0 = 10$) in Fig. 6.2 (Colored-IIR, $S/N = 22.5$ dB).

Chapter 7

Conclusions

7.1 Concluding Remarks

This research has been concerned with an innovative parameter-estimation technique, the OBE algorithm, which is based on a bounded-noise assumption. The OBE algorithms employ selective updating to improve computational efficiency (using suboptimal check) while having excellent speed of convergence. However, this advantage is offset by the impractical assumption of known noise bounds. Hence, despite those superiorities, OBE algorithms have rarely been found in real-world applications.

The new algorithms, OBE-ABE and their variants, introduced in this dissertation do not require the knowledge of the noise bounds, while preserving or improving, virtually without cost, the inherently superior performance of OBE algorithms with regard to convergence, speed of convergence, computational efficiency, and tracking capability. The new algorithms also converge in colored noise and non-stationary noise cases – a feat that is theoretically impossible for other popular techniques like RLS, LMS, Kalman-Bucy filter, etc. Hence, the popularity of the new algorithms in real-world applications is promising. A summary of the contributions of this dissertation appears in the next section.

7.2 Contributions and further work

The major contributions of this dissertation are summarized as follows: This research has

1. Produced an algorithm (OBE-ABE) which theoretically relaxes and practically releases the unavailable prerequisite (known noise bounds) of all SM (and OBE) algorithms while preserving and improving their inherently superior performance.
2. Produced a computationally efficient version of the new algorithm (Sub-OBE-ABE) without any sacrifice of performance. This efficient Sub-OBE-ABE algorithm has similar performance to the original OBE-ABE with respect to convergence, speed of convergence, and tracking capability.
3. Produced an adaptive version of Sub-OBE-ABE (and OBE-ABE) to track time-varying parameters. The tracking performance of the algorithm is excellent while its computation is very efficient.
4. Proved the a.s. convergence and the p. convergence of OBE-ABE and Sub-OBE-ABE in iid, mixing, ergodic, and non-stationary noise cases by employing the analytical techniques from probability theory and measure theory. Several theorems (with very general conditions) provide the theoretical foundations for the analysis and applications of OBE-ABE (and hence all SM) algorithms.
5. Provided extensive simulations which reveal some not-clearly-known behaviors or characteristics of OBE-type (hence SM) algorithms.
6. Applied the new algorithms successfully (due to their intensionally-devised flexibility and the resulting robustness) to real-world problems of speech analysis

and blind deconvolution. The application results have demonstrated the potential of the new algorithms.

As seen in the theorems and in simulation studies, the excellent convergence performance of all SM algorithms depends on UCT (3.9) or UT (3.10) condition of the noise sequence. Hence , an interesting open research problem is as follows:

- Devise an algorithm which has similar performance and *a priori* knowledge to the OBE-ABE's, and still converges consistently and quickly when UT (3.10) or UCT (3.9) is not satisfied.

APPENDIX

Appendix A

The following lemmas from the literature are essential for the proofs of new theorems. Original proofs of some of those lemmas are listed here for better understanding those lemmas and the proofs of the new theorems in this dissertation. Please refer to Chap. 3 for the notation.

Lemma A.1 (Proof: [27, 35]). *Assume that condition PE (3.12) holds. If there exists an $\epsilon > 0$ and $N \in \mathbb{N}$, such that $\gamma_n - v_n^2 > \epsilon$, $\forall n > N$, then the ellipsoids of OBE algorithms do not asymptotically shrink to a point.*

Lemma A.2 (Proof: [27]). *Let c_n^- and c_n^+ be the negative and positive parts of c_n , respectively, where c_n is defined in (2.10). That is $c_n^- = \{c_n, \text{ if } c_n < 0; 0, \text{ otherwise}\}$, and $c_n^+ = \{c_n, \text{ if } c_n \geq 0; 0, \text{ otherwise}\}$. If PE holds, then $\lim_{n \rightarrow \infty} c_n^- = 0$.*

Lemma A.3 [46, p.120], [16, p.35]. *A discrete random variable has an almost periodic characteristic function, and the converse is also true.*

Lemma A.4 [17]. *If $\{X_k : k \in \mathbb{K}\}$ is stationary and $g : \mathbb{R}^{\mathbb{K}} \rightarrow \mathbb{R}$ is measurable, then $\{Y_k = g(X_k, X_{k-1}, \dots) : k \in \mathbb{K}\}$ is also stationary.*

Proof: Let $x \in \mathbb{R}^{\mathbb{K}}$, $B \in \mathcal{B}^{\mathbb{K}}$, and $g_k = g(x_k, x_{k-1}, \dots)$. Also, let

$$A = \{x : (g_0(x), g_{-1}(x), \dots) \in B\}.$$

Then, stationarity of $\{X_k\}$ implies

$$\begin{aligned} P(\omega : (Y_0, Y_{-1}, \dots) \in B) &= P(\omega : (X_0, X_{-1}, \dots) \in A) = P(\omega : (X_k, X_{k-1}, \dots) \in A) \\ &= P(\omega : (Y_k, Y_{k-1}, \dots) \in B). \end{aligned}$$

Hence, by Definition 3.8, $\{Y_k\}$ is stationary. $\square$

Lemma A.5 [17]. *Let $X_k \stackrel{\text{def}}{=} X \circ T^{k-1}$, $\forall k \in \mathbb{K}$ where X is a r.v. on $(\Omega, \mathcal{F}, P)$. If T is an m.p.t., then, $\{X_k : k \in \mathbb{K}\}$ is a stationary sequence.*

Proof: The lemma is proved by observing that, for $B \in \mathcal{B}^{\mathbb{K}}$ and $A \stackrel{\text{def}}{=} \{\omega : (\dots, X_1, X_2, \dots) \in B\}$,

$$\begin{aligned} P((\dots, X_k, X_{k+1}, \dots) \in B) &= P(T^{k-1}(\omega) \in A) = P(\omega \in A) \\ &= P((\dots, X_1, X_2, \dots) \in B). \end{aligned}$$

$\square$

Lemma A.6 [17]. *If m.p.t. T is mixing, then T is ergodic.*

Proof: Let $A \in \mathcal{I}$. Then, $T^{-1}(A) = A, \dots, T^{-k}(A) = A, \forall k$. Hence, by definition 3.13,

$$P(A) = P(A \cap A) = P(A \cap T^{-k}(A)) \rightarrow P(A) \cdot P(A) \quad \text{as } n \rightarrow \infty.$$

So, $P(A) = [P(A)]^2$. Thus, $P(A) \in \{0, 1\}$. Hence, $\mathcal{I}$ is trivial. Therefore, T is ergodic. $\square$

Lemma A.7 [17]. *If $\{X_k : k \in \mathbb{K}\}$ is an ergodic stationary sequence, and $g : \mathbb{R}^{\mathbb{K}} \rightarrow \mathbb{R}$ is measurable, then $\{Y_k = g(X_k, X_{k-1}, \dots) : k \in \mathbb{K}\}$ is also stationary and ergodic.*

Proof: Let $X_0, X_{-1}, \dots$ be defined on sequence space $\Omega = \mathbb{R}^{\mathbb{K}}$ with $X_k(\omega) = \omega_k$, where $\omega = (\dots \omega_{-1}, \omega_0, \omega_1, \dots) \in \Omega$. Since $\{Y_k\}$ is stationary by Lemma A.4, let $B \in \mathcal{B}^{\mathbb{K}}$ be such that

$$\{\omega : (Y_0, Y_{-1}, \dots) \in B\} = \{\omega : (Y_k, Y_{k-1}, \dots) \in B\}.$$

then, by Definition 3.10, $A \stackrel{\text{def}}{=} \{\omega : (Y_0, Y_{-1}, \dots) \in B\}$ is shift invariant. So, by the ergodicity of $\{X_k\}$,

$$P(A) = P(\{\omega : (X_0, X_{-1}, \dots) \in A\}) \in \{0, 1\}.$$

Therefore, $\{Y_k\}$ is ergodic by Definitions 3.11 and 3.12. $\square$

Note that $\{X_k\}$ and $\{Y_k\}$ in the above lemma can be one-sided infinite sequences. Also, the mapping g which is independent of index k can also be any *finite*-length moving function.

Theorem A.1 (Poincaré Recurrence Theorem) [17]. *Let $T : \Omega \rightarrow \Omega$ be an m.p.t. and $A \in \mathcal{F}$. Define $\tau_A = \inf\{n \geq 1 : T^n(\omega) \in A\}$. Then,*

(i) $\tau_A < \infty$ a.s. on A , i.e., $P(\omega \in A : \tau_A(\omega) = \infty) = 0$.

(ii) $A \subset \{\omega : T^n(\omega) \in A \text{ i.o.}\}$.

(iii) If T is ergodic and $P(A) > 0$, then $P(\omega : T^n(\omega) \in A \text{ i.o.}) = 1$.

Proof:

(i) Let $B = \{\omega \in A : \tau_A(\omega) = \infty\}$. If $\omega \in T^{-1}(B)$, then $T(\omega) \in A$, $T^2(\omega) \notin A$, $T^3(\omega) \notin A, \dots$.

If $\omega \in T^{-2}(B)$, then $T^2(\omega) \in A$, $T^3(\omega) \notin A$, $T^4(\omega) \notin A, \dots$.

Hence, $T^{-1}(B)$, $T^{-2}(B)$, $T^{-3}(B)$, $\dots$ are pairwise disjoint. Since T is m.p.t., it follows that $P(B) = P(T^{-1}(B)) = P(T^{-2}(B)) = \dots$. Thus, $P(B) = 0$. (Otherwise, $P(\Omega) \geq \sum_{k=0}^{\infty} P(T^{-k}(B)) = \infty$.)

(ii) For any $k \in \mathbb{N}$, since T is m.p.t., it follows from (i) that

$$0 = P(\omega \in A : T^{nk}(\omega) \notin A, \forall n \geq 1) \geq P(\omega \in A : T^m(\omega) \notin A, \forall m \geq k).$$

This inequality implies that the last probability is 0 for all k . Hence, $P(\omega \in A : T^m(\omega) \in A \text{ i.o.}) = 1$. Thus, $A \subset \{\omega : T^n(\omega) \text{ i.o.}\}$.

(iii) By Definition 3.10, The set $C \stackrel{\text{def}}{=} \{\omega : T^n(\omega) \in A \text{ i.o.}\}$ is an invariant set. Since T is ergodic, it follows from Definition 3.12 that $P(C) \in \{0, 1\}$. Also, by (ii), $C \supset A$. Hence, $P(C) \geq P(A) > 0$ by assumption. Therefore, $P(C) = 1$. $\square$

BIBLIOGRAPHY

Bibliography

- [1] Amerio, L., Prouse, G., *Almost Periodic Functions and Functional Equations*, Van Norstrand Reinhold Company, New York, 1971.
- [2] Atal, B.S., J.R. Remde, "A new model of LPC excitation for producing natural-sounding speech at low bit rates," *Proc. ICASSP*, Paris, pp. 614-617, 1982.
- [3] Bányász, Cs. and Keviczky, L. (editors), *Proc. 9th IFAC Symp. on Identification and System Parameter Estimation (SYSID '91)* (Special sessions on bounded-error methods), vols. 1 & 2, Budapest, July, 1991.
- [4] Bertsekas, D. P., and I.B. Rhodes, "Recursive state estimation for a set-membership description of uncertainty," *IEEE Trans. on Automatic Control*, vol. AC-16, pp. 117-128, April, 1971.
- [5] Bitmead, R. R., "Persistence of excitation conditions and the convergence of adaptive schemes," *IEEE Trans. on Information Theory*, vol. IT-30, No. 2, pp. 183-191, March, 1984.
- [6] Bohr, H., *Almost Periodic Functions*, Chelsea Publishing, New York, 1951.
- [7] Cheung, M. F., *On Optimal Algorithms for Parameter Set Estimation* (Ph.D. dissertation), Ohio State University, 1991.
- [8] Dasgupta, S., and Y. F. Huang, "Asymptotically convergent modified recursive least squares with data dependent updating and forgetting factor for systems with bounded noise," *IEEE Trans. Info. Theory*, vol. 33, pp. 383-392, 1987.
- [9] Deller Jr., J. R., T. M. Lin, and M. Nayeri, "Linear prediction using set-membership identification with blind error-bound estimation," *Proc. 1997 IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, (in press).
- [10] Deller Jr., J. R., and M. Nayeri (session chairs), *Proc. IEEE Intl. Symp. on Circuits and System* (Special session on set-membership-based identification for signal processing and control), vol. 1, Chicago, May, 1993.
- [11] Deller Jr., J. R., M. Nayeri and S. F. Odeh, "Least-square identification with error bounds for real-time signal processing and control," *Proc. IEEE*, vol. 81, pp. 815-849, 1993.

- [12] Deller, Jr., J. R., J. G. Proakis, J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*, Macmillan, New York, 1993.
- [13] Deller Jr., J. R. and S. F. Odeh, "Adaptive set-membership identification in $\mathcal{O}(m)$ time for linear-in-parameters models," *IEEE Trans. on Signal Processing*, vol. 41, No. 5, May, 1993.
- [14] Deller Jr., J. R., M. Nayeri, and M. S. Liu, "Unifying the landmark developments in optimal bounding ellipsoid identification," *Int. J. of Adaptive Control and Signal Processing*, vol. 8, pp. 43–60, 1994.
- [15] Deller, J.R., and T.C. Luk, "Linear prediction analysis of speech based on set-membership theory," *Computer Speech and Language*, Oct., vol. 3, pp. 301–327, 1989
- [16] Douglas, J. B., *Analysis with Standard Contagious Distributions*, vol. 4, International Co-operative Publishing House, Burtonsville, Maryland, 1980.
- [17] Durrett, R., *Probability: Theory and Examples*, Wadsworth & Brooks/Cole Publishing Co., California, 1993.
- [18] Fisher, W. M., G. R. Doddington, and K. M. Goudie-Marshall, "The DARPA speech recognition research database," *Proc. DARPA Speech Recognition Workshop*, pp. 93–99, 1986.
- [19] Fogel, E., "System identification via membership set constraints with energy constrained noise" *IEEE Trans. on Automatic Control*, vol. 24, pp. 752–758, 1979.
- [20] Fogel, E., and Y. F. Huang, "On the value of information in system identification – bounded noise case," *Automatica*, vol. 18, pp. 229–238, 1982.
- [21] Goldberg, J. L., *Matrix Theory with Applications.*, MacGraw-Hill Co., 1991.
- [22] Graupe, D., *Time Series Analysis, Identification, and Adaptive Systems*, Krieger, Malabar, Florida, 1989.
- [23] Haykin, S. S., *Adaptive filter theory*, 3rd ed., Prentice Hall, Englewood Cliffs, New Jersey, 1996.
- [24] Kosut, R.L. (editor), *IEEE Trans. Automatic Control* (Special section on error bounding methods for robust control), vol. 37, July, 1992.
- [25] Lin, T. M., Nayeri, M., and J. R. Deller Jr., "Performance of Blind OBE in a Correlated-Error Environment," *30th Asilomar Conf. on Signals, Systems, and Computers.*, Pacific Grove, CA, Nov. 2 – 5, 1996 (in press).
- [26] Lin, T. M., Nayeri, M., and J. R. Deller Jr., "A Consistently Convergent OBE algorithm with Automatic Bound Estimation," *Int. J. of Adaptive Control and Signal Processing*, (in review).
- [27] Liu, M. S., *Development and Analysis of an Optimal Bounding Ellipsoid algorithm using Stochastic Approximation* (Ph.D. dissertation), Michigan State University, 1993.

- [28] Liu, M.S., M. Nayeri, and J.R. Deller Jr., "Do interpretable optimal bounding ellipsoid algorithms converge? Part II - OBE vs. RLS: Clearing the smoke," *Proc. 10th IFAC Symp. on System Identification (SYSID '94)*, vol. 3, pp. 395-400, Copenhagen, Denmark, July, 1994.
- [29] Milanese, M., H. Piet-Lahanier, J. Norton, and E. Walter (editors), *Bounding Approaches to System Identification*, Plenum, London, 1996.
- [30] Milanese, M., and A. Vicino, "Estimation theory for dynamic systems with unknown but bounded uncertainty: an overview," *Automatica*, vol. 27, pp. 997-1009, 1991.
- [31] Moore, J.B., "Persistence of excitation in extended least squares," *IEEE Trans. on Automatic Control*, vol. AC-28, No. 2, pp. 60 – 68, January, 1983.
- [32] Nayeri, M., J. R. Deller Jr., and M. M. Krunk, "Convergence and colored noise issues in bounding ellipsoid identification," *Proc. IEEE ICASSP '92*, San Francisco, vol. 5, pp. 337-340, March, 1992.
- [33] Nayeri, M., T. M. Lin, and J. R. Deller Jr., "Blind error-bound estimation for OBE identification," *Proc. 34th Annual Allerton Conf. on Communications, Control and Computing*, University of Illinois, Urbana-Champaign, Oct 2-4, 1996 (in press).
- [34] Nayeri, M., T. M. Lin, and J. R. Deller Jr., "Some novel variants of blind OBE algorithm," *Proc. 1997 IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, (in press).
- [35] Nayeri, M., M. S. Liu, and J. R. Deller Jr., "Do interpretable optimal bounding ellipsoid algorithms converge? Part I - The long-awaited set-convergence proof," *Proc. 10th IFAC Symp. on System Identification (SYSID '94)*, July, vol. 3, pp. 389-394, Copenhagen, Denmark, 1994.
- [36] Norton, J.P. (editor), *Int. J. Automatic Control and Signal Process* (Special issues on bounded-error estimation), Jan.-Feb., 1994-95.
- [37] Odeh, S. F., *Algorithm and Architectures For Adaptive Set Membership-based Signal Processing* (Ph.D. dissertation), Michigan State University, 1990.
- [38] Rao, A. K., and Y. F. Huang, "Tracking characteristics of an OBE parameter estimation algorithm," *IEEE Trans. Signal Processing*, vol. 41, pp. 1140-1148, 1993.
- [39] Rao, A. K., Y. F. Huang, and Dasgupta, "ARMA parameter estimation using a novel recursive estimation algorithm with selecting updating," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. 38, pp. 447-457, 1990.
- [40] Rabiner, L. R., and B.-H. Juang, *Fundamentals of Speech Recognition*, Prentice-Hall, Englewood Cliffs, New Jersey, 1993.
- [41] Schweppe F. C., "Recursive state estimation: Unknown but bounded errors and system inputs," *IEEE Trans. on Automatic Control*, vol. AC-13, pp. 22-28, Feb., 1968.

- [42] Söderström, T., and P. Stoica, *System Identification*, Prentice Hall, Englewood Cliffs, New Jersey, 1989.
- [43] Stark, H., J. W. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*, Prentice Hall, Englewood Cliffs, New Jersey, 1986.
- [44] Veres, S.M., and J.P. Norton, "Structure selection for bounded-parameter models: Consistency conditions and selection criterion," *IEEE Trans. on Automatic Control*, vol. 36, pp. 474-481, 1991.
- [45] Walter, E., and H. Piet-Lahanier, "Estimation of parameter bounds from bounded-error data: A survey," *Mathematics and Computers in Simulation*, vol. 32, pp. 449-468, 1990.
- [46] Wiener, N., *Generalized Harmonic Analysis and Tauberian Theorems*, MIT press, Cambridge, Massachusetts, 1964.
- [47] Witsenhausen, H. S., "Sets of possible state of linear systems given perturbed observation," *IEEE Trans. on Automatic Control*, vol. AC-13, pp. 556-568, Oct., 1968.

MICHIGAN STATE UNIV. LIBRARIES



31293015644614