

2003
54321531



This is to certify that the
dissertation entitled

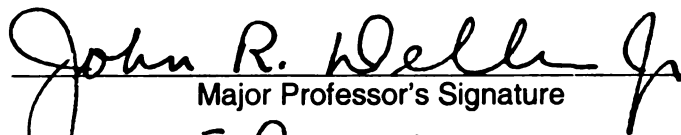
THE QOBE ALGORITHM: STOCHASTIC CONVERGENCE
ANALYSIS AND APPLICATION TO CLASSIFICATION

presented by

GIHAN MANDOUR

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Electrical and Computer
Engineering


Major Professor's Signature

15 August 2003

Date

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

**THE QOBE ALGORITHM: STOCHASTIC CONVERGENCE
ANALYSIS AND APPLICATION TO CLASSIFICATION**

By

Gihan Mandour

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Electrical and Computer Engineering

August 2003

ABSTRACT
THE QOBE ALGORITHM: STOCHASTIC CONVERGENCE
ANALYSIS AND APPLICATION TO CLASSIFICATION

By

Gihan Mandour

The *quasi-optimal bounding ellipsoid* (QOBE) algorithm has potential for application and theoretical development in many applications involving estimation of linear parametric models. QOBE offers superior adaptation, improved accuracy, efficient use of innovation in the data, improved computational efficiency, robustness to measurement noise and to deviation from the assumed noise model, and a set of feasible solutions supplementing the single point estimate.

Deterministic convergence analysis of the QOBE algorithm has been investigated in the literature. However, in many real-time DSP applications the signals to be modeled are approximated as the output from a linear shift-invariant filter whose inputs are stochastic signals. Therefore, there is a need for stochastic convergence analysis of an algorithm used in any system parameter identification or estimation to ensure its proper performance.

This research analyzes the convergence properties of the QOBE algorithm in a stochastic framework and provides insights into its operation in practical applications. For stochastic analysis, the innovations and the bounded noise sequences are modeled as random variables defined on a probability space. Mathematical formulations of theorems proving the almost sure convergence of the central estimator and that the hyperellipsoidal membership set associated with QOBE cannot converge to a point set, are constructed.

Simulation studies are performed to show the performance of the QOBE algorithm, its tracking capabilities, and demonstrate potential uses in classification problems.

**Copyright by
Gihan Mandour
2003**

To my late father: *fulfilling his wish.*

ACKNOWLEDGEMENTS

My gratitude goes first and foremost to God. I feel that He has been watching over me and has blessed me with all the strength to finish up my dissertation.

I would like to deeply thank my dissertation advisor Dr. John R. Deller, Jr. for all his support, help, and guidance throughout this work. The numerous discussions we had together broadened my knowledge of many aspects of this field of research and he had always pointed me towards the correct directions in my research.

I am very indebted to Dr. Hassan Khalil for he has provided me with all kinds of help, valuable advice and moral support.

I would like to extend my gratitude and thanks to Dr. Michael Frazier and Dr. Percy Pierre for all their help and guidance.

I owe a profound sense of gratitude and appreciation to my dear husband Tarek for he has supported me with love, patience, encouragement and strength. To him and to my lovely daughters, Dina and Noora, I am deeply indebted.

Finally, I am very grateful and obliged for all my family and friends who stood by my side and provided me with great comfort and compassion.

Table of Contents

List of Tables	viii
List of Figures	ix
1 Introduction and Background	1
1.1 Introduction	1
1.2 QOBE Algorithm	3
1.3 Review of QOBE Algorithms	6
1.4 Motivation for the Current Research	8
2 Stochastic Convergence Analysis of QOBE Algorithm	10
2.1 Introduction	10
2.2 Definitions	11
2.3 Persistency of Excitation Condition	14
2.4 Lemmas	16
2.5 Almost Sure Stochastic Convergence of QOBE	17
2.6 Convergence of the Central Estimator (Independent Noise Case) . . .	17
2.7 Convergence of the Central Estimator (Mixing Case)	25
2.8 Convergence Analysis of the Membership Set	29
2.9 Convergence Analysis of the Membership Set (Mixing Case)	29
3 Simulation Results	43
3.1 Introduction	43
3.2 Types of Noise Sequences	44
3.2.1 IID Noise Sequence	44
3.2.2 Colored Noise Sequence	52
3.3 Effect of Unsatisfied Conditions	62
3.4 Tracking Performance of QOBE	64
4 Application of QOBE in Classification	70
4.1 Introduction	70
4.2 Classification Using QOBE Algorithms	70
5 Conclusion and Future Work	81
5.1 Conclusions	81
5.2 Contributions and Future Work	82

Appendix	85
Bibliography	87

List of Tables

4.1	The minimum distance between class C and estimates of System 1 and System 2 along the minor axis	78
4.2	The minimum distance between class C and estimates of System 1 and System 2 along the major axis	78
4.3	The minimum distance between class C and estimates of System 1 and System 2 after rotating the major axis of System 2 by 90 degrees . . .	78
4.4	The minimum distance between class C and estimates of System 1 and System 2 after rotating the major axis of System 1 by 90 degrees . . .	79

List of Figures

3.1	Case 1: Estimators of $a_{1*} = 2.0$ convergence using QOBE and OBE algorithms for identification of model (3.1) with $\gamma_n = 0.96$, where 146 points and 223 points were used, respectively, from a total of 3000 points.	45
3.2	Case 1: Estimators of $a_{2*} = -1.48$ convergence using QOBE and OBE algorithms for identification of model (3.1) with $\gamma_n = 0.96$, where 146 points and 223 points were used, respectively, from a total of 3000 points.	45
3.3	Case 1: Estimators of $a_{3*} = 0.34$ convergence using QOBE and OBE algorithms for identification of model (3.1) with $\gamma_n = 0.96$, where 146 points and 223 points were used, respectively, from a total of 3000 points.	46
3.4	Case 1: Volume of the ellipsoids in QOBE and OBE algorithms for identification of model (3.1) with $\gamma_n = 0.96$, where 146 points and 223 points were used, respectively, from a total of 3000 points.	46
3.5	Case 2: Estimators of $a_{1*} = 2.0$ convergence using QOBE and OBE algorithms for identification of model (3.1), $\gamma_n = 1.04$, where 24 points and 916 points were used, respectively, from a total of 3000 points. . .	48
3.6	Case 2: Estimators of $a_{2*} = -1.48$ convergence using QOBE and OBE algorithms for identification of model (3.1), $\gamma_n = 1.04$, where 24 points and 916 points were used, respectively, from a total of 3000 points. . .	48
3.7	Case 2: Estimators of $a_{3*} = 0.34$ convergence using QOBE and OBE algorithms for identification of model (3.1), $\gamma_n = 1.04$, where 24 points and 916 points were used, respectively, from a total of 3000 points. . .	49
3.8	Case 2: Volume of the ellipsoids in QOBE and OBE algorithms for identification of model (3.1) with $\gamma_n = 1.04$, where 24 points and 916 points were used, respectively, from a total of 3000 points.	49
3.9	Case 3: Estimators of $a_{1*} = 2.0$ convergence using QOBE and OBE algorithms for identification of model (3.1), $\gamma_n = 1.04$, where 27 points and 374 points were used, respectively, from a total of 3000 points. . .	50
3.10	Case 3: Estimators of $a_{2*} = -1.48$ convergence using QOBE and OBE algorithms for identification of model (3.1), $\gamma_n = 1.04$, where 27 points and 374 points were used, respectively, from a total of 3000 points. . .	50
3.11	Case 3: Estimators of $a_{3*} = 0.34$ convergence using QOBE and OBE algorithms for identification of model (3.1), $\gamma_n = 1.04$, where 27 points and 374 points were used, respectively, from a total of 3000 points. . .	51
3.12	Case 3: Volume of the ellipsoids in QOBE and OBE algorithms for identification of model (3.1) with $\gamma_n = 1.04$, where 27 points and 374 points were used, respectively, from a total of 3000 points.	51

3.13	Case 4: Estimators of $a_{1*} = -0.1$ and $a_{2*} = 0.56$ convergence using QOBE for identification of model (3.4) with $\gamma_n = 0.96$, where 125 points were used from a total of 3000 points.	53
3.14	Case 4: Volume of the ellipsoid in QOBE algorithm for identification of model (3.4) with $\gamma_n = 0.96$, where 125 points were used from a total of 3000 points.	53
3.15	Case 4: The trajectory of the estimators in model (3.4) with $\gamma_n = 0.96$, where 125 points were used from a total of 3000 points.	53
3.16	Case 5: Estimators of $a_{1*} = 2.0$ convergence using QOBE and OBE algorithms for identification of model (3.1), $\gamma_n = 1.04$, where 26 points and 845 points were used, respectively, from a total of 3000 points. . .	55
3.17	Case 5: Estimators of $a_{2*} = -1.48$ convergence using QOBE and OBE algorithms for identification of model (3.1), $\gamma_n = 1.04$, where 26 points and 845 points were used, respectively, from a total of 3000 points. . .	55
3.18	Case 5: Estimators of $a_{3*} = 0.34$ convergence using QOBE and OBE algorithms for identification of model (3.1), $\gamma_n = 1.04$, where 26 points and 845 points were used, respectively, from a total of 3000 points. . .	56
3.19	Case 5: Volume of the ellipsoids in QOBE and OBE algorithms for identification of model (3.1) with $\gamma_n = 1.04$, where 26 points and 845 points were used, respectively, from a total of 3000 points.	56
3.20	Case 6: Estimators of $a_{1*} = 2.0$ convergence using QOBE and OBE algorithms for identification of model (3.1), $\gamma_n = 1.04$, where 26 points and 207 points were used, respectively, from a total of 3000 points. . .	57
3.21	Case 6: Estimators of $a_{2*} = -1.48$ convergence using QOBE and OBE algorithms for identification of model (3.1), $\gamma_n = 1.04$, where 26 points and 207 points were used, respectively, from a total of 3000 points. . .	57
3.22	Case 6: Estimators of $a_{3*} = 0.34$ convergence using QOBE and OBE algorithms for identification of model (3.1), $\gamma_n = 1.04$, where 26 points and 207 points were used, respectively, from a total of 3000 points. . .	58
3.23	Case 6: Volume of the ellipsoids in QOBE and OBE algorithms for identification of model (3.1) with $\gamma_n = 1.04$, where 26 points and 207 points were used, respectively, from a total of 3000 points.	58
3.24	Case 7: Estimators of $a_{1*} = -0.1$ and $a_{2*} = 0.9175$ and $a_{5*} = 0.2601$ convergence using QOBE for identification of model (3.7) with $\gamma_n = 1.04$, where 115 points were used from a total of 3000 points.	60
3.25	Case 7: Volume of the ellipsoid in QOBE algorithm for identification of model (3.7) with $\gamma_n = 1.04$, where 115 points were used from a total of 3000 points.	60
3.26	Case 8: Estimators of $a_{1*} = -0.1$ and $a_{2*} = 0.56$ convergence using QOBE for identification of model (3.4) with $\gamma_n = 0.96$, where 974 points were used from a total of 3000 points.	61
3.27	Case 8: Volume of the ellipsoid in QOBE algorithm for identification of model (3.4) with $\gamma_n = 0.96$, where 974 points were used from a total of 3000 points.	61

3.28	Case 8: The trajectory of the estimators in model (3.4) with $\gamma_n = 0.96$, where 974 points were used from a total of 3000 points.	61
3.29	Estimators of $a_{1*} = 2.0$, $a_{2*} = -1.48$ and $a_{3*} = 0.34$ using QOBE for identification of model (3.1) when the PE condition is not satisfied. .	62
3.30	Estimators of $a_{1*} = 2.0$, $a_{2*} = -1.48$ and $a_{3*} = 0.34$ using QOBE for identification of model (3.1) when the UCT condition is not satisfied.	63
3.31	Estimators of $a_{1*} = 2.0$, $a_{2*} = -1.48$ and $a_{3*} = 0.34$ using QOBE for identification of model (3.1) when both the PE and the UCT conditions are not satisfied.	63
3.32	Case 9: Estimators of a_{1*} (time-varying parameter with slow abrupt change) convergence using QOBE and OBE algorithms for identification of model (3.8), $\gamma_n = 1.1$	65
3.33	Case 9: Volume of the ellipsoids in QOBE and OBE algorithms for identification of model (3.8), $\gamma_n = 1.1$, and a_{1*} is time-varying parameter with slow abrupt change.	65
3.34	Case 10: Estimators of a_{1*} (time-varying parameter with fast abrupt change) convergence using QOBE and OBE algorithms for identification of model (3.8), $\gamma_n = 1.1$	66
3.35	Case 10: Volume of the ellipsoids in QOBE and OBE algorithms for identification of model (3.8), $\gamma_n = 1.1$, and a_{1*} is time-varying parameter with fast abrupt change.	66
3.36	Case 11: Estimators of a_{1*} (time-varying parameter with slow gradual change) convergence using QOBE and OBE algorithms for identification of model (3.8), $\gamma_n = 1.1$	68
3.37	Case 11: Volume of the ellipsoids in QOBE and OBE algorithms for identification of model (3.8), $\gamma_n = 1.1$, and a_{1*} is time-varying parameter with slow gradual change.	68
3.38	Case 12: Estimators of a_{1*} (time-varying parameter with fast gradual change) convergence using QOBE and OBE algorithms for identification of model (3.8), $\gamma_n = 1.1$	69
3.39	Case 12: Volume of the ellipsoids in QOBE and OBE algorithms for identification of model (3.8), $\gamma_n = 1.1$, and a_{1*} is time-varying parameter with fast gradual change.	69
4.1	QOBE central estimator θ_t of θ_{1*} and θ_{2*} of two AR(2) systems represented by '+' and 'o', respectively.	73
4.2	$\theta_{t,LSE}$ estimator of θ_{1*} and θ_{2*} of two AR(2) systems represented by '+' and 'o', respectively.	73
4.3	QOBE other estimators Ω_t of θ_{1*} and θ_{2*} of two AR(2) systems represented by '+' and 'o', respectively.	74
4.4	The direct 256 points FFT spectrum of the impulse response for AR(2). The upper graph for System1. The lower graph for System2. In each graph: the true parameters (-); estimators along the minor axis Other1 (o) and Other2 (+); and estimators along the major axis Other3 ($\square$) and Other4 (.).	75

4.5	QOBE central estimator θ_t of θ_{1*} and θ_{2*} of two AR(14) systems represented by '+' and 'o', respectively.	76
4.6	$\theta_{t,LSE}$ estimator of θ_{1*} and θ_{2*} of two AR(14) systems represented by '+' and 'o', respectively.	76
4.7	QOBE other estimators Ω_t of θ_{1*} and θ_{2*} of two AR(14) systems represented by '+' and 'o', respectively.	77
4.8	Final ellipsoids of System 1 and System 2 used for associating class C to its membership set	78

Chapter 1

Introduction and Background

1.1 Introduction

Many digital signal processing (DSP) applications require parametric signal models. Linear-in-parameters systems form a broad class of models that has been extensively investigated. Many system identification or parameter estimation techniques have been developed to estimate the unknown parameters in a linear model.

Set membership (SM) algorithms are used to identify linear-in-parameters signal or system models of the form

$$y_n = \theta_*^T x_n + \varepsilon_n^* \quad (1.1)$$

in which $\theta_* \in \mathbb{R}^m$ is the unknown “true” parameter vector to be identified; $\{x_n\}$ is a sequence of measurable vectors of dimension m ; and $\{\varepsilon_n^*\}$ is a noise sequence or an error sequence. SM algorithms have recently begun to find their way into the DSP field because of their tremendous potential for application and theoretical development in virtually any application involving estimation of linear parametric models. The basic idea of the estimation procedure is to provide a set of solutions that are compatible with the observations x_n and the disturbance constraint ε_n^* . Thus the estimate is a set in parameter space rather than a single vector as in the least square (LS) estimation

techniques [1]. In some applications, the center of the parameter set can be used as a point estimate for the model parameters, while in other applications the entire set of parameters might be of interest [2]. A widely researched class of SM algorithms is that involving bounded error (BE) constraints where a pointwise bound on the disturbance sequence is assumed. The *optimal bounding ellipsoid* (OBE) algorithms belong to the class of BE methods. The bounding ellipsoid associated with an OBE algorithm can be calculated recursively in real time using an algorithm of relatively low computational complexity. Like all SM methods, the OBE algorithms seek a membership set in the parameter space which is guaranteed to contain all solutions that are compatible with the model of the underlying process, the assumptions on the noise and the observation data.

OBE identification algorithms are of practical use in many real-time signal-processing applications and have received considerable attention [3]-[16]. With respect to the conventional recursive methods like recursive least squares (RLS) [1],[2],[17]-[22], OBE identifiers offer superior adaptation, improved accuracy, efficient use of innovation in the data, improved computational efficiency, robustness to measurement noise and to deviation from the assumed noise model, and a set of feasible solutions supplementing the single point estimate. The references in [3],[4],[7]-[16],[23]-[38] provide a broad and current overview of such work.

This research is concerned with a specific SM algorithm known as the *quasi optimal bounding ellipsoid* (QOBE) algorithm [39]-[41]. The QOBE algorithm is a member of the class of general OBE algorithms [10] but its particular structure gives it properties that yield special application performance and makes it worthy of separate investigation. QOBE algorithm's convergence properties have been studied in a deterministic framework. The purpose of the present work is to analyze the convergence properties of QOBE in a stochastic framework, and to furnish an understanding of its operation in practice through application studies, such as classification problems.

1.2 QOBE Algorithm

QOBE shares most of the motivating principles and algorithmic structure with the other OBE algorithms; however, it has other geometric and classic least-squares interpretations that promote its use in potential applications. The name QOBE has been used in the literature to indicate both the similarities and differences between QOBE and previous instances of OBE.

The OBE algorithms in general are used to identify a *linear-in-parameters* discrete-time model of the form (1.1). One special case that is used extensively in many signal processing applications and on which we will base our study is the *auto-regressive with exogenous input* (ARX) model [2, 18], in which $x_n = [y_{n-1}, \dots, y_{n-p}, u_n, \dots, u_{n-q}]^T$ is the observed data sequence consisting of samples of an observable input sequence $\{u_n\}$ and output sequence $\{y_n\}$ [13]. In this case $m = p + q + 1$ in (1.1) and $\theta_* = [a_{1*}, \dots, a_{p*}, b_{o*}, b_{1*}, \dots, b_{q*}]^T$.

The OBE algorithm, in general, is based on the assumption that $\{\varepsilon_n^*\}$ has a pointwise energy bound that is known *a priori*

$$\varepsilon_n^{*2} \leq \gamma_n^2 \text{ for all } n \quad (1.2)$$

At each n , these bounds imply two hyperplanes in parameter space

$$H_n^+ = \{\theta \mid y_n = \theta^T x + \gamma_n\} \text{ and } H_n^- = \{\theta \mid y_n = \theta^T x - \gamma_n\} \quad (1.3)$$

between which θ_* must lie. The intersection of these pairs of hyperplanes over time forms a polytope in $\mathbb{R}^m$, which can be shown [10] to be a subset of a hyperellipsoidal set at time n given by

$$\Omega_n \stackrel{\text{def}}{=} \{\theta \mid (\theta - \theta_n)^T C_n (\theta - \theta_n) < \kappa_n\} \quad (1.4)$$

where κ_n is expressed as

$$\kappa_n = \theta_n^T C_n \theta_n + \sum_{s=1}^n \lambda_s (\gamma_s^2 - y_s^2) + \kappa_o, \quad n = 1, 2, \dots \quad (1.5)$$

The ellipsoid center, θ_n , κ_n , and matrix $P_n \stackrel{\text{def}}{=} C_n^{-1}$ are computed recursively using [9, 10, 42]

$$G_n = x_n^T P_{n-1} x_n \quad (1.6)$$

$$\varepsilon_n = y_n - \theta_{n-1}^T x_n \quad (1.7)$$

$$P_n = \frac{1}{\alpha_n} \left[P_{n-1} - \frac{\beta_n P_{n-1} x_n x_n^T P_{n-1}}{\alpha_n + \beta_n G_n} \right] \quad (1.8)$$

$$\theta_n = \theta_{n-1} + \beta_n P_n x_n \varepsilon_n \quad (1.9)$$

$$\kappa_n = \alpha_n \kappa_{n-1} + \beta_n \gamma_n - \frac{\alpha_n \beta_n \varepsilon_n^2}{\alpha_n + \beta_n G_n} \quad (1.10)$$

It is also useful for theoretical purposes to note that C_n , the so-called *covariance matrix*, can be computed recursively as

$$C_n = \alpha_n C_{n-1} + \beta_n x_n x_n^T$$

Recursions (1.6) - (1.10) comprise the basis for a general OBE algorithm. OBE algorithms are distinguished from one another by the choices of the positive weighting sequences $\{\alpha_n\}$ and $\{\beta_n\}$ according to the optimization criterion that minimizes the size of the set Ω_n at each iteration.

In the QOBE algorithm, the scalar κ_n is minimized at each iteration by letting $\beta_n = \lambda_n$ and $\alpha_n = 1$, then seeking the optimal λ_n , a strategy first used by Dasgupta and Huang [4]. Since $\{\alpha_n\}$ and $\{\beta_n\}$ are functions of λ_n , then the optimization criterion is performed with respect to λ_n only. The optimal weight for QOBE is

found by solving $F_\kappa(\lambda) = 0$ where

$$F_\kappa(\lambda) = a_n \lambda^2 + b_n \lambda + c_n \quad (1.11)$$

in which

$$\begin{aligned} a_n &= \gamma_n^2 G_n^2 \\ b_n &= 2\gamma_n^2 G_n \\ c_n &= (\gamma_n^2 - \varepsilon_n^2) \end{aligned} \quad (1.12)$$

Thus yielding

$$\lambda_n = \frac{1}{G_n} \left(\frac{|\varepsilon_n|}{\gamma_n} - 1 \right) = \frac{\delta_n}{G_n \gamma_n} \quad (1.13)$$

where $\delta_t \stackrel{\text{def}}{=} |\varepsilon_t| - \gamma_t$ and the reduced value of κ_n is computed as

$$\kappa_n = \kappa_{n-1} - G_n^{-1} \delta_n^2 \quad (1.14)$$

Note that a necessary and sufficient test for the existence of the positive root is $c_n < 0$.

Therefore, the algorithm updates the estimators only when $c_n < 0$.

QOBE algorithms are distinguishable by their ability to selectively use the incoming data to update the ellipsoid and the central estimator. When the observations at time n contain no innovation in the sense that they cannot be used to reduce the size of the set Ω_{n-1} , then they are discarded. This implies that no valid weights can be found at this instant and thus the effort of updating at this time can be avoided. Therefore, depending on the properties of the sequence $\{\varepsilon_n^{*2}\}$, QOBE algorithms often update only 10 percent of the time or less.

1.3 Review of QOBE Algorithms

This section is based largely on the review in the papers by [7]-[12],[14],[23]-[25],[39],[40],[43]-[46]. In 1968, Scheweppe published the first bounding ellipsoid algorithm [23] for estimating the state parameters of a linear dynamic system using noisy observations. He derived the algorithm under the assumption that both the input and the observation errors are unknown but bounded. His choice of bounding the optimum estimate set by an ellipsoid was due to fact that an ellipsoidal approximation algorithm has certain computational advantages. However, as Scheweppe notes, this new algorithm was presented with no convergence proof and required impractical computational loads.

Following the footsteps of Scheweppe, Witsenhausen[43] , and Bertsekas and Rhodes [24] introduced the SM algorithm using similar assumptions of bounded-input and bounded-noise. However, Bertsekas and Rhodes showed that SM is similar in structure and comparable in simplicity to the Kalman-Bucy filter [48] known to be the best linear estimator for the Gaussian white noise case.

In 1979, Fogel [25], studied for the first time the problem of estimating a membership set of unknown parameters of a dynamic system. The analysis yielded an OBE identification algorithm for the ARX model with bounded-energy constraints on the noise sequence. In a deterministic sense, Fogel has shown that the sequence of bounding hyperellipsoids asymptotically reduces to a point set (the true model parameter vector). Fogel's assertion that the convergence of his algorithm implies the convergence of the OBE algorithm has remained controversial [7, 9].

In 1982, Fogel and Huang [44] formulated system parameter identification as a membership-set estimation problem and proposed the first OBE algorithm (FH/OBE) that employs *selective updating* to ignore redundant data. Selective updating in FH/OBE is achieved by using a weighting strategy that assigns a sequence of weights $\{q_i\}$ to incoming data points. In the FH/OBE algorithm, a set measure defined on $\mathbb{R}^m$

that reflects the geometric size of the set is defined such that only data points which can reduce such size of the set are used. If q_i is zero then the new data point contains no useful information to reduce the set size and thus it can be discarded. The results of FH/OBE algorithm are comparable to the LS estimation method while achieving $O(m^2)$ computations. However, Fogel and Huang expressed the convergence of their algorithm only when the observation error is white noise and with known pointwise upper bounds. This assumption was employed to make use of Fogel's previous assertion [25] and therefore has rendered some argument about the validity of their proof [10]-[12],[14].

A new minimization strategy was introduced by Dasgupta and Huang in 1987 [4]. They considered minimization of κ_n , a scalar, to facilitate convergence. However, later Deller *et al.* [9, 10] questioned the interpretability of the algorithm's optimization criterion since minimization of κ_n had no apparent relation to the hyperellipsoid volume. Dasgupta and Huang proved that their algorithm (DH/OBE) is simple, of $O(m)$ complexity, and its parameter estimates converge asymptotically to a region around the true parameter.

Several attempts have been made to implement the minimization strategy of Dasgupta and Huang within an OBE algorithm that has geometric and least square interpretations [8],[39]-[40],[45]-[46],[48]. In 1996, Gollamudi *et al.* [39] presented a SM state estimation scheme based on the DH/OBE algorithm to recursively estimate the state of a system with bounded inputs and noise. The algorithm showed significant performance and computational savings in terms of the mean-square error in the state estimate. In 1997, the same group of authors [46] introduced a novel formulation of the linear-in-parameters SM filtering problem and presented an adaptive algorithm named "SMART". SMART, the set-membership adaptive recursive technique, was derived using the SM filtering framework of DH/OBE while implementing the same recursions as the normalized least mean square (NLMS) algorithm. The technique offered non-

increasing size of the set estimates in the parameter space. In 1997, Nagaraj *et al.* introduced the “official” QOBE [40] with its asymptotic convergence properties and selective-updating capabilities. The algorithm used the same minimization criterion as the one developed by Dasgupta and Huang and was shown to reduce the percentage of updates and have excellent tracking performance. The convergence analysis of the QOBE algorithm for identification and filtering was presented by Deller *et al.*, in 1997 [41]. They showed that QOBE represents a “hybrid” between the set-membership weighted recursive least square SM-WRLS algorithm and the DH/OBE algorithm where QOBE implements the same weighting technique as the one used in SM-WRLS but uses the minimization technique used by Dasgupta and Huang. However, their convergence analysis was restricted to a deterministic setting. Since then, many papers have been published by this group of authors and others interested in the development of the QOBE algorithm and its implementation in many of the DSP applications [38],[40]-[41],[45]-[47],[49]-[50].

1.4 Motivation for the Current Research

The development of algorithms for recursively identifying the parameters of a time-invariant dynamic system is widely studied in the literature using two approaches, the deterministic and the stochastic approach. In deterministic settings, signals may be defined as those that can be described by a mathematical expression or that can be reproduced repeatedly [1]. On the other hand, stochastic signals are defined as representation of random processes that are not repeatable in a predictable manner and which are identified by the statistical properties of such random processes. In many real-time DSP applications the signals to be modelled are approximated as the output from a linear shift-invariant filter whose inputs are stochastic signals such as stationary white noise. Therefore, there is a need for deriving the stochastic con-

vergence analysis of an algorithm used in any system parameter identification or estimation to ensure its proper performance.

In this research we analyze the convergence properties of the QOBE algorithm in a stochastic framework and provide some insights into its operation in practical applications. For stochastic analysis, y_n , u_n , and ε_n are modelled as random variables defined on a probability space. Complete analysis and details of the stochastic convergence for the QOBE algorithm are introduced in the following chapter.

Chapter 2

Stochastic Convergence Analysis of QOBE Algorithm

2.1 Introduction

Many researchers, especially the group of Nagaraj, Gallamudi, Huang, Kapoor and Deller [39]-[41],[45]-[46],[48]-[49], have focused on the development of the QOBE algorithm, an algorithm that uses a weighting strategy similar to the SM-WRLS [9]-[12] but optimizes over the scalar κ_n as in the DH/OBE algorithm [4]. These groups have proved in a deterministic setting that the parameter estimator asymptotically converges to the “true” parameter vector under certain persistency of excitation (PE) conditions of the observation sequence $\{x_n\}_{n=1}^{\infty}$ and under infinite visitation (IV) conditions of the true disturbance sequence $\{\varepsilon_n^*\}_{n=1}^{\infty}$ of any arbitrary small neighborhood of its true bounds [51]. The purpose of the present work is to analyze the convergence properties of the QOBE algorithm in a stochastic framework and to provide some insights into its operation in practical applications.

QOBE features include highly-selective updating, linear complexity in the number of parameters, and its excellent tracking ability in time-varying environments.

Further, since QOBE does not explicitly depend upon the bounding ellipsoid, the need to assure that θ_* remains inside the ellipsoid, which is critical in OBE processing, is not important in QOBE. This is another important feature of QOBE for tracking performance. Finally, the ellipsoidal set associated with QOBE has interesting convergence behavior *in practice*, which, while seemingly undesirable from a theoretical point of view, may also benefit tracking behavior.

Deterministic approaches to convergence analysis assume signals to be predictable. However, for stochastic analysis, y_n , u_n , and ε_n are modelled as random variables defined on a probability space $(S, \mathfrak{F}, P)$ where S is a sample space, $\mathfrak{F}$ is a Borel σ -algebra of subsets of S , and P is a probability measure on $\mathfrak{F}$. In an ARX model, the observation sequence is of the form $\{x_n\} \stackrel{\text{def}}{=} \{y_{n-1}, \dots, y_{n-p}, u_n, \dots, u_{n-q}\}$. For analytical purposes, we will assume that the model is stable, and, if we let $\mathfrak{F}_n = \sigma(\varepsilon_m, u_{m+1}, m \leq n)$, then x_n is $\mathfrak{F}_{n-1}$ measurable.

Recent results for more conventional OBE algorithms in stochastic settings [13, 14] provide us with guiding principles for our present work on analyzing the stochastic convergence of the QOBE algorithm. They also suggest that the analysis of the QOBE algorithm in a stochastic framework is feasible following similar lines.

2.2 Definitions

In this section we will introduce definitions that will facilitate the statement and analysis of new theorems. Most of these definitions are cited from the literature.

Definition 2.1 [2] *An estimator θ_n , of θ_* , is called p . convergent (convergent in probability) if for all $\epsilon > 0$*

$$\lim_{n \rightarrow \infty} P(\|\theta_n - \theta_*\| > \epsilon) = 0 \quad (2.1)$$

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

It is be noted that convergence in probability implies weak convergence (convergence in distribution).

Definition 2.2 [2] *An estimator θ_n , of θ_* , is called a.s. convergent (almost surely convergent) if*

$$P(\lim_{n \rightarrow \infty} \theta_n = \theta_*) = 1 \quad (2.2)$$

Equation (2.2) is equivalent to

$$P(\|\theta_n - \theta_*\| > \epsilon \text{ i.o.}) = 0 \quad \forall \epsilon > 0 \quad (2.3)$$

where *i.o.* denotes *infinitely often* with respect to n .

From Definition 2.2 it can be seen that *a.s.* convergence implies *p.* convergence. Although *s.* convergence (sure convergence), defined as $\lim_{n \rightarrow \infty} \theta_n = \theta_*$ [52], is theoretically stronger than *a.s.* convergence, the latter is regarded as the strongest type of convergence employed in practice and in most of the literature [13].

Definition 2.3 *A random sequence $\{\varepsilon_n^*\}$ is said to be asymptotically independent if*

$$\lim_{n \rightarrow \infty} |P((\varepsilon_i^* \in A) \cap (\varepsilon_{i+n}^* \in B)) - P(\varepsilon_i^* \in A) P(\varepsilon_{i+n}^* \in B)| = 0$$

$$\forall i \text{ and } \forall A, B \in \mathfrak{F} \quad (2.4)$$

Recall from (1.2) that in SM algorithms, in general, it is assumed that the noise sequence $\{\varepsilon_n^*\}$ has a pointwise energy bound that is known *a priori*. Therefore, for convenience, we will denote ϵ -neighborhoods of the noise bounds as

$$D_\epsilon^+ = [\gamma_n - \epsilon, \gamma_n]$$

$$D_\epsilon^- = [-\gamma_n, -\gamma_n + \epsilon]$$

Definition 2.4 [29] *A random sequence $\{\varepsilon_n^*\}$ is said to be almost surely uniformly*

conditionally heavy tailed (a.s. UCHT) if there exists $C_1 > C_2 > 0$ and an infinite subsequence $\{n_i\} \subset \{n\} \subset \mathbf{N}$ such that, with any sufficiently small $\epsilon > 0$

$$C_1 \epsilon \leq P (\varepsilon_n^* \in (D_\epsilon^+ \cup D_\epsilon^-) \mid \mathfrak{F}_{n-1}) \leq C_2 \epsilon \quad \forall n \in \{n_i\} \quad (2.5)$$

Definition 2.5 A random sequence $\{\varepsilon_n^*\}$ is said to be almost surely uniformly conditionally tailed (a.s. UCT) if given $\epsilon > 0$ there exists a $\delta > 0$ and an infinite subsequence $\{n_i\} \subset \{n\} \subset \mathbf{N}$, such that

$$P (\varepsilon_n^* \in (D_\epsilon^+ \cup D_\epsilon^-) \mid \mathfrak{F}_{n-1}) > \delta \quad \forall n \in \{n_i\} \quad (2.6)$$

Definition 2.6 A random sequence $\{\varepsilon_n^*\}$ is said to be almost surely uniformly tailed (a.s. UT) if given $\epsilon > 0$, there exists a $\delta > 0$ and an infinite subsequence $\{n_i\} \subset \{n\} \subset \mathbf{N}$ such that

$$P (\varepsilon_n^* \in (D_\epsilon^+ \cup D_\epsilon^-)) > \delta \quad \forall n \in \{n_i\} \quad (2.7)$$

It is be noted that UT and UCT conditions are less restrictive than the UCHT condition.

Definition 2.7 A random sequence $\{\varepsilon_n^*\}$ is asymmetrically bounded with a.s. UCT condition if given $\epsilon > 0$, there exists a $\delta > 0$ such that

$$P (\varepsilon_n^* \in D_\epsilon^+ \mid \mathfrak{F}_{n-1}) \geq \delta \quad \text{and} \quad P (\varepsilon_n^* \in D_\epsilon^- \mid \mathfrak{F}_{n-1}) = 0 \quad \forall n$$

or

$$P (\varepsilon_n^* \in D_\epsilon^+ \mid \mathfrak{F}_{n-1}) = 0 \quad \text{and} \quad P (\varepsilon_n^* \in D_\epsilon^- \mid \mathfrak{F}_{n-1}) \geq \delta \quad \forall n$$

Definition 2.8 A random sequence $\{\varepsilon_n^*\}$ is symmetrically bounded with a.s. UCT condition if given $\epsilon > 0$, there exists a $\delta > 0$ such that

$$P (\varepsilon_n^* \in D_\epsilon^+ \mid \mathfrak{F}_{n-1}) \geq \delta_1 \quad \text{and} \quad P (\varepsilon_n^* \in D_\epsilon^- \mid \mathfrak{F}_{n-1}) \geq \delta_2 \quad \forall n$$

Definition 2.9 *The term infinite visitation (IV) refers to the condition in which the true disturbance sequence $\{\varepsilon_n^*\}$ visits arbitrarily small neighborhoods of its bounds infinitely often (i.o.). That is, IV occurs if for any $e > 0$ $|\varepsilon_n^*| > \gamma_n - e$ i.o.*

Definition 2.10 *The term infinite updating (IU) refers to the condition in which the QOBE algorithm does not cease updating in finite time.*

Definition 2.11 *The term infinite nulling (IN) refers to the condition in which any infinite subsequence say $\{\varepsilon_{n_k}^*\}_{k=1}^\infty$ of the true disturbance sequence $\{\varepsilon_n^*\}_{k=1}^\infty$ satisfies the condition that for any $e > 0$, $|\varepsilon_{n_k}^*| < e$ i.o.*

2.3 Persistency of Excitation Condition

A key requirement for the convergence of any recursive estimator is that the inputs be sufficiently uncorrelated or persistently exciting (PE) so as to make the parameters in (1.1) uniquely identifiable. The following definitions of PE appear in the literature:

Definition 2.12 [29] *A sequence of random vectors, $\{x_n\}$, is called persistently exciting (PE), or omni-directional, almost surely (a.s.) if for any nonsingular cone*

$$K = \{x : x = \alpha_1 e_1 + \cdots + \alpha_m e_m, \det(e_1, \cdots, e_m) \neq 0, \alpha_i > 0 \forall i\}$$

there exists ρ_1 and ρ_2 such that

$$\lim_{n \rightarrow \infty} \inf P(x_n \in K \mid \mathfrak{I}_{n-2}) \geq \rho_1 > 0 \quad (2.8)$$

and

$$E(\|x_n\| \mid \mathfrak{I}_{n-2}) \geq \rho_2 > 0 \quad \forall n \quad (2.9)$$

Condition (2.8) means that the orientation of x_n is sufficiently varied in a conditional probability sense, while (2.9) means that the magnitude of x_n cannot be too small on the average.

Definition 2.13 [29] *A sequence of random vectors, $\{x_n\}$, is called PE a.s. if there exists ρ_1 and ρ_2 such that*

$$\lim_{n \rightarrow \infty} \inf E(x_n x_n^T | \mathfrak{S}_{n-2}) \geq \rho_1 I > 0 \quad (2.10)$$

and

$$E(\|x_n\| | \mathfrak{S}_{n-2}) \geq \rho_2 > 0 \quad \forall n \quad (2.11)$$

The fact that $E(E(x_n x_n^T | \mathfrak{S}_{n-2})) = E(x_n x_n^T)$ leads to an equivalent definition of PE:

Definition 2.14 [29] *A sequence of random vectors, $\{x_n\}$, is called PE a.s. if there exists ρ_1 and ρ_2 such that*

$$\lim_{n \rightarrow \infty} \inf E(x_n x_n^T) \geq \rho_1 I > 0 \quad (2.12)$$

and

$$E(\|x_n\| | \mathfrak{S}_{n-2}) \geq \rho_2 > 0 \quad \forall n \quad (2.13)$$

Definition 2.15 [26] *A WSS sequence of random vectors, $\{x_n\}$, is called PE a.s. if there exists ρ_1 and ρ_2 such that*

$$E(x_n x_n^T) \geq \rho_1 I > 0 \quad \forall n \quad (2.14)$$

and

$$E(\|x_n\| | \mathfrak{S}_{n-2}) \geq \rho_2 > 0 \quad \forall n \quad (2.15)$$

It is to be noted that similar definitions exist in the literature for deterministic analysis [2, 18, 26].

2.4 Lemmas

The following lemmas are useful in a convergence proof to follow.

Lemma 2.1 [15] *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.*

Lemma 2.2 [13] *Assume the model (1.1) is a stable ARX model and $\{u_n\}$ and $\{\varepsilon_n^*\}$ are bounded. If both $\{u_n\}$ and $\{\varepsilon_n^*\}$ are asymptotically independent sequences, and $\{u_n\}$ is independent of $\{\varepsilon_n^*\}$, then y_n converges to a continuous random variable as $n \rightarrow \infty$.*

It is to be noted that if $\{u_n\}$ or $\{\varepsilon_n^*\}$ or both are continuously distributed, then the assumptions that $\{u_n\}$ and $\{\varepsilon_n^*\}$ are asymptotically independent sequences and the independence between $\{u_n\}$ and $\{\varepsilon_n^*\}$ are not required in Lemma 2.2.

Lemma 2.3 [13] *Assume that model (1.1) is a stable ARX model, and $\{u_n\}$ and $\{\varepsilon_n^*\}$ are bounded. If PE holds, both $\{u_n\}$ and $\{\varepsilon_n^*\}$ are asymptotically independent sequences, $\{u_n\}$ is independent of $\{\varepsilon_n^*\}$, and the noise bounds are overestimated for QOBE algorithms, i.e., if there exists an $\epsilon_1 > 0$ and an $N \in \mathbf{N}$ such that $\forall n > N$, $\gamma_n^2 - \epsilon_n^{*2} > \epsilon_1$, then the expected updating interval of the estimator θ_n approaches infinity as n approaches infinity.*

It is to be noted that if $\{u_n\}$ or $\{\varepsilon_n^*\}$ or both are continuously distributed, then the assumptions that $\{u_n\}$ and $\{\varepsilon_n^*\}$ are asymptotically independent sequences and the independence between $\{u_n\}$ and $\{\varepsilon_n^*\}$ are not required in Lemma 2.3.

2.5 Almost Sure Stochastic Convergence of QOBE

White noise has been commonly studied in the analysis of many system identification and parameter estimation algorithms as well as model structures. In the following work, we consider cases in which $\{\varepsilon_n^*\}$ is white noise, i.e., $E(\varepsilon_n^*) = 0$ and $E(\varepsilon_s^* \varepsilon_t^*) = \sigma^2 \delta(s - t)$ where δ is the Kronecker sequence.

The theorems of *a.s.* convergence of QOBE algorithm are introduced and proven when both the input sequence $\{u_n\}$ and the noise sequence $\{\varepsilon_n^*\}$ are assumed stationary and $\{u_n\}$ is assumed independent of $\{\varepsilon_n^*\}$. For *p.* convergence, $\{u_n\}$ and $\{\varepsilon_n^*\}$ are not required to be stationary. Furthermore, if $\{u_n\}$, $\{\varepsilon_n^*\}$, or both are continuously distributed, then the conditions for *a.s.* and *p.* convergence can be relaxed.

In [53], it is shown through an illustrative example, how different cross-sequence dependencies can affect the convergence behavior of QOBE. Fortunately, the dependencies needed for “good behavior” can be easily described in a stochastic framework. This will be illustrated in the next theorems of *a.s.* convergence.

2.6 Convergence of the Central Estimator (Independent Noise Case)

Theorem 2.1 *Assume that the stationary sequences $\{u_n\}$ and $\{\varepsilon_n^*\}$ are independent. If PE (2.8) holds, UCT holds, and $\{u_n\}$ and $\{\varepsilon_n^*\}$ are asymptotically independent and ergodic, then the central estimator of the QOBE algorithm is *a.s.* convergent.*

Proof: Since PE (2.8) and UCT (2.6) hold, then there exists an infinite subsequence $\{n_i\} \in \mathbf{Z}$ such that for all $n \in \{n_i\}$, (2.6) and

$$P(x_n \in K \mid \mathfrak{F}_{n-2}) \geq \rho'_1 > 0 \quad a.s.$$

hold, where K is defined in Definition 2.12 and $\rho'_1 = \rho_1 - \epsilon$, for some $\epsilon > 0$. For

convenience we let n replace n_i throughout the proof.

Let $\{\varepsilon_n^* : n \in \mathbf{Z}\}$ denote the noise sequence and $\{u_n : n \in \mathbf{Z}\}$ denote the input sequence of model (1.1). Thus we can rewrite (1.1) as

$$y_n - \sum_{j=1}^p a_j y_{n-j} = \sum_{j=0}^q b_j u_{n-j} + \varepsilon_n^* \quad (2.16)$$

Let

$$w_n \stackrel{\text{def}}{=} \sum_{j=0}^q b_j u_{n-j} \quad (2.17)$$

and $v_n = w_n + \varepsilon_n^*$. Then (2.16) becomes

$$y_n = \sum_{-\infty}^n h_{n-j} v_j \quad (2.18)$$

where h_n is the infinite impulse response, y_n , if v_n is set to a Kronecker delta function.

By Lemma A.1 $\{w_n\}$ is stationary and ergodic. Also, by Lemma A.2 $\{v_n\}$ is stationary and ergodic.

Let $\{I_k\}$ be a sequence of time intervals of length M over which the algorithm does not update its estimator. This will be helpful in asserting a contradiction later on.

Therefore, 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 . From (1.7) we have

$$\begin{aligned} \varepsilon_n &= y_n - \theta_{n-1}^T x_n \\ &= y_n - \sum_{i=1}^m \theta(i) y_{n-i} \\ \varepsilon_n &= - \sum_{i=0}^m \theta(i) y_{n-i}, \text{ where } \theta(0) \stackrel{\text{def}}{=} -1 \end{aligned} \quad (2.19)$$

From (2.18) and (2.19) we have

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

Changing the order of summation, we get $\forall n > m$

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

Let $p = n - j$, then

$$\begin{aligned} \varepsilon_n &= - \sum_{p=0}^{m-1} v_{n-p} \sum_{i=0}^p \theta(i) h_{p-i} - \sum_{p=m}^{\infty} v_{n-p} \sum_{i=0}^m \theta(i) h_{p-i} \\ \varepsilon_n &= - \sum_{p=0}^{m-1} v_{n-p} g_1 - \sum_{p=m}^{\infty} v_{n-p} g_2 \end{aligned}$$

where g_1 and g_2 are mapping functions independent of n , hence,

$$\varepsilon_n = g(v_n, v_{n-1}, v_{n-2}, \dots) \quad (2.20)$$

where $g : \mathbb{R}^k \rightarrow \mathbb{R}$ is measurable and independent of n . Thus, by Lemma A.1 $\{\varepsilon_n\}$ is stationary and ergodic as $M \rightarrow \infty$.

For each $k = 1, 2, \dots, K$, it follows from (1.12)

$$c_n = \gamma_n^2 - \varepsilon_n^2 > 0 \quad \forall n \in I_k$$

and from (1.7)

$$\varepsilon_n = y_n - \theta_{n-1}^T x_n$$

Since

$$\varepsilon_n^* = y_n - \theta_*^T x_n$$

then we get by letting $\tilde{\theta}_n \stackrel{\text{def}}{=} \theta_* - \theta_n$

$$\varepsilon_n - \varepsilon_n^* = \tilde{\theta}_{n-1}^T x_n$$

and thus

$$\varepsilon_n = \varepsilon_n^* + \tilde{\theta}_{n-1}^T x_n \quad (2.21)$$

Let A_n^- be the event that $\varepsilon_n^* \in D_\epsilon^-$ and $\tilde{\theta}_{n-1}^T x_n \geq 0$.

Let A_n^+ be the event that $\varepsilon_n^* \in D_\epsilon^+$ and $\tilde{\theta}_{n-1}^T x_n < 0$.

Let $A_n = A_n^+ \cup A_n^-$. Note that $\tilde{\theta}_{n-1}^T x_n$ is $\mathfrak{F}_{n-1}$ measurable.

Also, let B_k denote the event that A_n occurs at least once in I_k . Then for all $n \in I_{k+1}$ we have $B_k \in \mathfrak{F}_{n-2}$. Thus since

$$P(x_n \in K \mid \mathfrak{F}_{n-2}) \geq \rho'_1 > 0 \text{ a.s.}$$

then

$$P\left(\tilde{\theta}_{n-1}^T x_n \geq 0 \mid B_k\right) \geq \rho'_1 > \rho > 0 \quad (2.22)$$

and

$$P\left(\tilde{\theta}_{n-1}^T x_n < 0 \mid B_k\right) \geq \rho'_2 > \rho > 0 \quad (2.23)$$

where $\rho'_2 = \rho_2 - \epsilon$ for some $\epsilon > 0$. Since $B_k \in \mathfrak{F}_{n-2} \subset \mathfrak{F}_{n-1}$ and $\tilde{\theta}_{n-1}^T x_n$ is $\mathfrak{F}_{n-1}$ measurable, it follows from (2.6) that

$$P\left(\varepsilon_n^* \in (D_\epsilon^+ \cup D_\epsilon^-) \mid \tilde{\theta}_{n-1}^T x_n \geq 0, B_k\right) \geq \delta_1 > \frac{\delta}{\rho} > 0 \quad (2.24)$$

and

$$P\left(\varepsilon_n^* \in (D_\epsilon^+ \cup D_\epsilon^-) \mid \tilde{\theta}_{n-1}^T x_n < 0, B_k\right) \geq \delta_2 > \frac{\delta}{\rho} > 0 \quad (2.25)$$

Since A_n^+ and A_n^- are mutually exclusive for sufficiently small ϵ , it follows from (2.22)

- (2.25) and the definition of symmetrically bounded sequence with UCT in Defintion 2.7, we have that

$$\begin{aligned}
P(A_n | B_k) &= P(A_n^- | B_k) + P(A_n^+ | B_k) \\
&= P\left(\varepsilon_n^* \in D_\epsilon^- \mid \tilde{\theta}_{n-1}^T x_n \geq 0, B_k\right) P\left(\tilde{\theta}_{n-1}^T x_n \geq 0 \mid B_k\right) \\
&\quad + P\left(\varepsilon_n^* \in D_\epsilon^+ \mid \tilde{\theta}_{n-1}^T x_n < 0, B_k\right) P\left(\tilde{\theta}_{n-1}^T x_n < 0 \mid B_k\right)
\end{aligned}$$

Thus,

$$\begin{aligned}
P(A_n | B_k) &> P\left(\varepsilon_n^* \in D_\epsilon^- \mid \tilde{\theta}_{n-1}^T x_n \geq 0, B_k\right) \rho \\
&\quad + P\left(\varepsilon_n^* \in D_\epsilon^+ \mid \tilde{\theta}_{n-1}^T x_n < 0, B_k\right) \rho \\
&> \delta > 0
\end{aligned} \tag{2.26}$$

From (2.26) and the definition of A_n , we have

$$P(|\varepsilon_n| > \gamma_n - \epsilon \mid B_k) \geq P(A_n | B_k) > \delta > 0 \tag{2.27}$$

Thus by Theorem A.1 and (2.27) we have

$$P(|\varepsilon_n| > \gamma_n - \epsilon \text{ i.o.} \mid B_k) = 1 \text{ as } M \rightarrow \infty \tag{2.28}$$

Therefore, $\{\varepsilon_n\}$ exhibits IV and thus from (2.21) we have

$$\left| \varepsilon_n^* + \tilde{\theta}_{n-1}^T x_n \right| > \gamma_n - \epsilon$$

and so we get

$$|\varepsilon_n^*| + \left| \tilde{\theta}_{n-1}^T x_n \right| > \gamma_n - \epsilon$$

Since $\{x_n\}$ is PE, we consider the following two cases:

1. If $\lim_{n \rightarrow \infty} \theta_n = \theta^*$, then there exists some K_ϵ such that $\left| \tilde{\theta}_{n-1}^T x_n \right| < \epsilon$ for all $k \geq K_\epsilon$ for arbitrary $\epsilon > 0$. Therefore $|\varepsilon_n^*| > \gamma_n - \epsilon$ for all $k \geq K_\epsilon$ and $\{\varepsilon_n^*\}$ exhibits IV
2. If $\lim_{n \rightarrow \infty} \theta_n \neq \theta^*$ then from (2.21) and the PE condition

$$\left| \tilde{\theta}_{n-1}^T x_n \right| = |\varepsilon_n - \varepsilon_n^*| > \eta \text{ i.o. for some } \eta > 0$$

If we define $\delta_n = |\varepsilon_n| - \gamma_n$ and $\delta_n^* = \gamma_n - |\varepsilon_n^*|$ so that $\delta_n + \delta_n^* = |\varepsilon_n| - |\varepsilon_n^*|$ and if $|\delta_n^*| < \frac{\eta}{2}$ i.o., then $|\varepsilon_n^*| > \gamma_n - \frac{\eta}{2}$ i.o. and $\{\varepsilon_n^*\}$ exhibits IV. Thus contradiction occurs and the assumption here is not true. From this argument and from (2.28) and (2.2) we conclude *a.s.* convergence of the QOBE estimator. $\diamond$

Theorem 2.2 *Assume that the stationary sequences $\{u_n\}$ and $\{\varepsilon_n^*\}$ are independent. If $\{u_n\}$, or $\{\varepsilon_n^*\}$, or both, are continuously distributed random sequences, and if PE (2.8) holds, UCT holds, and $\{u_n\}$ and $\{\varepsilon_n^*\}$ are ergodic, then the estimator of the QOBE algorithm is *a.s.* convergent.*

Proof: Continuously distributed $\{u_n\}$ or $\{\varepsilon_n^*\}$ or both implies that Lemmas 2.2 and 2.3 hold without the assumption that $\{u_n\}$ and $\{\varepsilon_n^*\}$ are asymptotically independent sequences. The theorem can therefore be proved following the same steps as in the proof of Theorem 2.1. $\diamond$

Corollary 2.1 *Assume that the stationary sequences $\{u_n\}$ and $\{\varepsilon_n^*\}$ are independent. If PE (2.8) holds, UT holds, $\{u_n\}$ is asymptotically independent and $\{\varepsilon_n^*\}$ is i.i.d., then the estimator of the QOBE algorithm is *a.s.* convergent.*

Proof: Since UT and UCT imply each other with the i.i.d. assumption, and i.i.d. implies asymptotic independence, then *a.s.* convergence follows immediately from Theorem 2.1. $\diamond$

Corollary 2.2 *Assume that the stationary sequences $\{u_n\}$ and $\{\varepsilon_n^*\}$ are independent. If $\{u_n\}$, or $\{\varepsilon_n^*\}$, or both, are continuously distributed random sequences, and if PE (2.8) holds, UT holds, $\{u_n\}$ is ergodic and $\{\varepsilon_n^*\}$ is i.i.d., then the estimator of QOBE algorithm is a.s. convergent.*

Proof: Since UT and UCT imply each other with the i.i.d. assumption, and i.i.d. implies ergodic, then a.s. convergence follows immediately from Theorem 2.2. $\diamond$

The following corollary asserts a.s. convergence when the probability distribution of the noise sequence $\{\varepsilon_n^*\}$ is known.

Corollary 2.3 *Assume that the stationary sequences $\{u_n\}$ and $\{\varepsilon_n^*\}$ are mutually independent. If PE (2.8) holds, $\{u_n\}$ is ergodic and $\{\varepsilon_n^*\}$ is i.i.d. with uniform distribution then the estimator of the QOBE algorithm is a.s. convergent.*

Proof: For a uniform distribution,

$$P(\varepsilon_n^* \in (D_\epsilon^+ \cup D_\epsilon^-)) \geq \frac{\epsilon}{2\gamma_n} > 0$$

Thus UT holds. By Corollary 2.3, the estimator of the QOBE algorithm is a.s. convergent. $\diamond$

Corollary 2.4 *Assume that the sequences $\{u_n\}$ and $\{\varepsilon_n^*\}$ are stationary and $\{u_n\}$ is independent. If PE (2.8) holds, UCT holds, $\{u_n\}$ is asymptotically independent and $\{\varepsilon_n^*\}$ exhibits IV, then the estimator of QOBE algorithm is a.s. convergent.*

Proof: The statement falls immediately from the proof of Theorem 2.1. $\diamond$

Definition 2.16 *Let $(S, \mathfrak{F}, P)$ be a probability space as previously defined. Let $T : (S, \mathfrak{F}) \rightarrow (S, \mathfrak{F})$ be a measurable transformation on $(S, \mathfrak{F}, P)$, then T is said to be a “measure preserving” if and only if $P(T^{-1}A) = P(A)$ for all $A \in \mathfrak{F}$*

We will define the measure preserving transformation (*m.p.t.*) to be “one-sided shifter”

$T(v_0, v_1, \dots) = (v_1, v_2, \dots)$ where v_i , for all i , is an event defined on the sample space $S_1 = (a, b)$ for some real numbers a and b .

$$P(T^{-1}A) = P(A) \text{ implies that } P(T^{-k}A) = P(A) \text{ for } k = 1, 2, \dots$$

Thus with $S = S_1^n$ we have

$$P\{(v_0, v_1, \dots, v_{n-1}) \in \beta_n\} = P\{(v_k, v_{k+1}, \dots, v_{n+k-1}) \in \beta_n\}$$

for all n and k and n -dimensional Borel sets β_n .

Definition 2.17 Let T be measure preserving on $(S, \mathfrak{S}, P)$. T is said to be independent if for all $A, B \in \mathfrak{S}$, $P(A \cap T^{-k}B) = P(A)P(B)$ for all $k \in N$. T is said to be “mixing” if for all $A, B \in \mathfrak{S}$, $\lim_{k \rightarrow \infty} P(A \cap T^{-k}B) = P(A)P(B)$.

From Definition 2.14 and since $E(x_n x_n^T)$ is positive definite, an upper bound is guaranteed on $E(x_n x_n^T)$ by the stability assumption of the model in (1.1). Also we note that if $\{x_n\}$ is mixing then all subsequences of $\{x_n\}$ are also mixing. Thus if $E(x_n x_n^T)$ is bounded below at the points where the updates take place then the PE condition (2.10) with x_n replaced by x_{n_k} is satisfied.

A key point in the convergence proof is the independence between $\{\varepsilon_n^*\}$ and $\left\{ \begin{smallmatrix} \sim T \\ \theta_{n-1} \end{smallmatrix} x_n \right\}$. In order to extend the convergence property from independence to mixing condition we need to modify the definition of x_n in the ARX model since x_n contains y_{n-1} , the output at time $n-1$. Thus, independency of $\{\varepsilon_n^*\}$ is required to remove the dependency of ε_n^* and $\begin{smallmatrix} \sim T \\ \theta_{n-1} \end{smallmatrix} x_n$. Otherwise, x_n can remain unchanged unless an update takes place and therefore we can apply the mixing condition to show convergence, provided that for large n the time difference between two updates is large. We proved convergence before in case of $\{\varepsilon_n^*\}$ is independent. We will now prove convergence for the mixing condition case. So we will use a modified model (1.1) with

$$x_n = [y_{t_i-1} \cdots y_{t_i-p} u_{t_i} \cdots u_{t_i-q}] \quad (2.29)$$

where $\{y_n\}$ and $\{u_n\}$ are the same as in (1.1), and $\{t_i\}$ is a subsequence such that $c_{t_i} < 0$ and $t_i \leq n < t_{i+1}$.

2.7 Convergence of the Central Estimator (Mixing Case)

Theorem 2.3 *Assume $\{x_n\}$ is either ergodic or mixing. Let $\{\varepsilon_n^*\}$ be a noise sequence and let T be mixing on $\{S, \mathfrak{S}, P\}$. If the modified model of (1.1) with (2.29) is employed, PE (2.8) condition holds, UCT condition holds, then the central estimator of the QOBE algorithm is a.s. convergent.*

Proof: Since PE and UCT conditions hold, then there exists an infinite subsequence $\{t_i\} \in Z$ such that for all $n \in \{t_i\}$ (2.6) and $P(x_n \in K \mid \mathfrak{S}_{n-2}) \geq \rho_1 > 0$ a.s. hold, where K is defined in Definition 2.12.

Following the same argument as in Theorem 2.1, we conclude that $\{\varepsilon_n\}$ is stationary and ergodic as $M \rightarrow \infty$, where M is defined in the theorem statement. For convenience we will replace $\{t_i\}$ with $\{n\}$ throughout the proof.

Let $\{I_k\}$ be the sequence of time intervals of length $M > N$ over which the QOBE algorithm does not update its estimator. This will be helpful in asserting a contradiction later on.

From (2.21) we have

$$\varepsilon_n^* + \tilde{\theta}_{n-1}^T x_n = \varepsilon_n$$

Let A_n^- be the event that $\varepsilon_n^* \in D_\epsilon^-$ and $\tilde{\theta}_{n-1}^T x_n \geq 0$

Let A_n^+ be the event that $\varepsilon_n^* \in D_\epsilon^+$ and $\tilde{\theta}_{n-1}^T x_n < 0$

Let $A_n \stackrel{\text{def}}{=} A_n^- \cup A_n^+$

Note that $\tilde{\theta}_{n-1}^T x_n$ is $\mathfrak{S}_{n-1}$ measurable.

Let B_k denote the event that A_n occurs at least once in I_k . Then $\forall n \in I_{k+1}$, $B_k \in \mathfrak{S}_{n-2}$.

Since

$$P(x_n \in \kappa \mid \mathfrak{S}_{n-2}) \geq \rho'_1 > 0 \text{ a.s.}$$

then

$$P(\tilde{\theta}_{n-1}^T x_n \geq 0 \mid B_k) \geq \rho'_1 > \rho > 0 \quad (2.30)$$

and

$$P(\tilde{\theta}_{n-1}^T x_n < 0 \mid B_k) \geq \rho'_2 > \rho > 0 \quad (2.31)$$

Since $B_k \in \mathfrak{S}_{n-2} \subset \mathfrak{S}_{n-1}$ and $\tilde{\theta}_{n-1}^T x_n$ is $\mathfrak{S}_{n-1}$ measurable, it follows from (2.6) that

$$P\left(\varepsilon_n^* \in (D_\epsilon^- \cup D_\epsilon^+) \mid \tilde{\theta}_{n-1}^T x_n \geq 0, B_k\right) \geq \delta_1 > \frac{\delta}{\rho} > 0 \quad (2.32)$$

and

$$P\left(\varepsilon_n^* \in (D_\epsilon^- \cup D_\epsilon^+) \mid \tilde{\theta}_{n-1}^T x_n < 0, B_k\right) \geq \delta_2 > \frac{\delta}{\rho} > 0 \quad (2.33)$$

Since A_n^+ and A_n^- are mutually exclusive events for sufficiently small ϵ , it follows from (2.30)- (2.33) and asymmetrical bounded condition with UCT that

$$\begin{aligned} P(A_n \mid B_k) &= P(A_n^- \mid B_k) + P(A_n^+ \mid B_k) \\ &= P\left(\varepsilon_n^* \in D_\epsilon^- \mid \tilde{\theta}_{n-1}^T x_n \geq 0, B_k\right) \\ &\quad P\left(\tilde{\theta}_{n-1}^T x_n \geq 0 \mid B_k\right) \\ &\quad + P\left(\varepsilon_n^* \in D_\epsilon^+ \mid \tilde{\theta}_{n-1}^T x_n < 0, B_k\right) \\ &\quad P\left(\tilde{\theta}_{n-1}^T x_n < 0 \mid B_k\right) \end{aligned} \quad (2.34)$$

Let $k_i = t_i - t_{i-1}$ (recall the replacement of t_i by n previously) and define

$$W_{t_{i-1}} \stackrel{\text{def}}{=} \left\{ \tilde{\theta}_{t_{i-1}}^T x_{t_i} \geq 0 \right\} \text{ and } V_{t_i} \stackrel{\text{def}}{=} \{ \varepsilon_{t_i}^* \in D_\epsilon^- \}$$

Then V_{t_i} and $W_{t_{i-1}} \in \mathfrak{S}$ and $P(V_{t_i}) = \delta$

By the definition of mixing in Definition 2.17,

$$\lim_{k_i \rightarrow \infty} P(V_{t_i} \cap W_{t_i - k_i}) = P(V_{t_i}) \lim_{k_i \rightarrow \infty} P(W_{t_i - k_i}) \quad (2.35)$$

Suppose $P(W_{t_i - k_i}) \neq 0$, then (2.35) can be written as

$$\begin{aligned} \lim_{k_i \rightarrow \infty} \frac{P(V_{t_i} \cap W_{t_i - k_i})}{P(W_{t_i - k_i})} &= \lim_{k_i \rightarrow \infty} P(V_{t_i} \mid W_{t_i - k_i}) \\ &= P(V_{t_i}) \end{aligned}$$

Let $0 < \xi < \delta$ and there exists an N such that

$$P(V_{t_i}) - \xi < P(V_{t_i} \mid W_{t_i - N}) < P(V_{t_i}) + \xi \quad (2.36)$$

This inequality holds if no updates take place on the interval $[t_i - N, t_i]$.

Now replace $\{t_i\}$ again with $\{n\}$, from (2.34) we have

$$\begin{aligned}
P(A_n | B_k) &= P\left(\varepsilon_n^* \in D_\epsilon^- \mid \tilde{\theta}_{n-1}^T x_n \geq 0, B_k\right) P\left(\tilde{\theta}_{n-1}^T x_n \geq 0 \mid B_k\right) \\
&\quad + P\left(\varepsilon_n^* \in D_\epsilon^+ \mid \tilde{\theta}_{n-1}^T x_n < 0, B_k\right) P\left(\tilde{\theta}_{n-1}^T x_n < 0 \mid B_k\right) \\
&\geq [P(\varepsilon_n^* \in D_\epsilon^-, B_k) - \xi] P\left(\tilde{\theta}_{n-1}^T x_n \geq 0 \mid B_k\right) \\
&\quad + [P(\varepsilon_n^* \in D_\epsilon^+, B_k) - \xi] P\left(\tilde{\theta}_{n-1}^T x_n < 0 \mid B_k\right) \\
&\geq [P(\varepsilon_n^* \in D_\epsilon^-, B_k) - \xi] \rho + [P(\varepsilon_n^* \in D_\epsilon^+, B_k) - \xi] \rho \\
&\geq [P(\varepsilon_n^* \in D_\epsilon^- \cup D_\epsilon^+, B_k) - \xi] \rho \\
&\geq \left(\frac{\delta}{\rho} - \xi\right) \rho > 0
\end{aligned}$$

From the definition of A_n

$$P(|\varepsilon_n| > \gamma_n - \epsilon \mid B_k) \geq P(A_n \mid B_k) > \delta - \xi \rho > 0$$

Thus by Theorem A.1 we have

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

Thus $|\varepsilon_n|$ exhibits IV, and thus from (2.21) we have

$$\left| \varepsilon_n^* + \tilde{\theta}_{n-1}^T x_n \right| > \gamma_n - \epsilon$$

and so we get

$$|\varepsilon_n^*| + \left| \tilde{\theta}_{n-1}^T x_n \right| > \gamma_n - \epsilon$$

Since $\{x_n\}$ is PE, we consider the following two cases:

1. If $\lim_{n \rightarrow \infty} \theta_n = \theta_*$, then there exists some K_ϵ such that $\left| \tilde{\theta}_{n-1}^T x_n \right| < \epsilon$ for all $k \geq K_\epsilon$ for arbitrary $\epsilon > 0$. Therefore $|\varepsilon_n^*| > \gamma_n - \epsilon$ for all $k \geq K_\epsilon$ and $\{\varepsilon_n^*\}$ exhibits IV

2. If $\lim_{n \rightarrow \infty} \theta_n \neq \theta_*$ then from (2.21) and the PE condition

$$\left| \tilde{\theta}_{n-1}^T x_n \right| = |\varepsilon_n - \varepsilon_n^*| > \eta \text{ i.o. for some } \eta > 0$$

If we define $\delta_n = |\varepsilon_n| - \gamma_n$ and $\delta_n^* = \gamma_n - |\varepsilon_n^*|$ so that $\delta_n + \delta_n^* = |\varepsilon_n| - |\varepsilon_n^*|$, and $|\delta_n^*| < \frac{\eta}{2}$ i.o. then $|\varepsilon_n^*| > \gamma_n - \frac{\eta}{2}$ i.o. and $\{\varepsilon_n^*\}$ exhibits IV. Therefore we have a contradiction, and the assumption here is not true. From this argument and from (2.28) and (2.2) we conclude *a.s.* convergence of the QOBE central estimator. $\diamond$

2.8 Convergence Analysis of the Membership Set

In optimizing the parameter κ_n , QOBE does not place formal emphasis on the contraction of the hyperellipsoidal membership set, Ω_n , in the selection and weighting of data. As a consequence of minimizing κ_n , the volume is automatically decreased at each update, but generally not optimally, unlike OBE algorithm, which optimizes over the volume of Ω_n and under proper data conditions, $\theta_n \rightarrow \theta_*$ asymptotically as the set Ω_n converges to a point. In QOBE, therefore, κ_n parameterizes the ellipsoid and has no clear connection to the ellipsoid center θ_n . However, QOBE provides excellent parameter convergence and faster tracking accompanied by larger ellipsoid volumes than its OBE counterpart. This is again due to the fact that QOBE becomes extremely selective of data as its estimator approaches the true parameter vector.

2.9 Convergence Analysis of the Membership Set (Mixing Case)

Theorem 2.4 *Even under the conditions of Theorem 2.3 (leading to the central estimator convergence), the volume of the ellipsoid of the QOBE algorithm does not*

converge to zero (or in other words the sequence of hyperellipsoidal membership sets associated with QOBE cannot converge to a point set) in probability.

Proof: Let $\mathcal{N}' \stackrel{\text{def}}{=} \{n \in \mathcal{N} : c_n > 0\}$ so to consider only $n \in \mathcal{N}'$ that implies no update takes place (i.e. $\gamma_n^2 - \varepsilon_n^2 > 0$ over a sequence of infinite number of intervals of length M)

$$(\gamma_n - |\varepsilon_n|)(\gamma_n + |\varepsilon_n|) > 0$$

Let us define the sequence $\{t_i\}_{i=1}^{\infty}$ to be the sequence over which no update takes place. This will be helpful in asserting a contradiction later on.

Since $\varepsilon_{t_i} = \varepsilon_{t_i}^* + \tilde{\theta}_{t_i-1}^T x_{t_i}$ we have

$$\begin{aligned} \gamma_{t_i}^2 - \varepsilon_{t_i}^2 &= (\gamma_{t_i} + |\varepsilon_{t_i}|)(\gamma_{t_i} - |\varepsilon_{t_i}|) \\ &= (\gamma_{t_i} + |\varepsilon_{t_i}|) \left(\gamma_{t_i} - \left| \varepsilon_{t_i}^* + \tilde{\theta}_{t_i-1}^T x_{t_i} \right| \right) > 0 \end{aligned}$$

Therefore

$$\gamma_{t_i} - \left| \varepsilon_{t_i}^* + \tilde{\theta}_{t_i-1}^T x_{t_i} \right| > 0 \quad (2.37)$$

Thus

$$\gamma_{t_i}^2 - \varepsilon_{t_i}^2 > 0 \text{ for all } \{t_i\} \text{ if and only if}$$

$$\left| \varepsilon_{t_i}^* + \tilde{\theta}_{t_i-1}^T x_{t_i} \right| < \gamma_{t_i} \quad (2.38)$$

Let us assume that the volume of the ellipsoid of QOBE does not converge to zero and try to assert that this assumption is true. If the volume of the ellipsoid does not converge to zero then condition (2.38) must hold. For convenience, we will denote

ϵ' – neighborhoods of noise bounds as

$$\begin{aligned} D_{\epsilon'}^+ &= [\gamma_n - \epsilon', \gamma_n] \\ D_{\epsilon'}^- &= [-\gamma_n, -\gamma_n + \epsilon'] \end{aligned}$$

At no update we have

$$\begin{aligned} \varepsilon_n^2 &< \gamma_n^2 \\ \left| \varepsilon_n^* + \tilde{\theta}_{n-1}^T x_n \right| &< \gamma_n \end{aligned} \tag{2.39}$$

If $\tilde{\theta}_{n-1}^T x_n \geq 0$ and $\varepsilon_n^* \in D_{\epsilon'}^-$, i.e. $(-\gamma_n < \varepsilon_n^* < -\gamma_n + \epsilon')$ then since ε_n^* is uniformly distributed, we have

$$\gamma_n - \epsilon' < |\varepsilon_n^*| < \gamma_n$$

$$\begin{aligned} |\varepsilon_n^*| + \left| \tilde{\theta}_{n-1}^T x_n \right| &< \gamma_n + \left| \tilde{\theta}_{n-1}^T x_n \right| \\ \left| \varepsilon_n^* + \tilde{\theta}_{n-1}^T x_n \right| &< \gamma_n \end{aligned}$$

Same as defined in (2.39), therefore $\varepsilon_n^* \in D_{\epsilon'}^-$. Similar analysis can be repeated to show that $\varepsilon_n^* \in D_{\epsilon'}^+$ when $\tilde{\theta}_{n-1}^T x_n < 0$.

Given ϵ' , assume $\{\varepsilon_{t_i}^*\}$ is symmetrically bounded with UCT, which means there exists $\delta_{t_i} > 0$, $\rho_{t_i} > 0$ such that for all t_i

$$\begin{aligned} P(\varepsilon_{t_i}^* \in D_{\epsilon'}^+ \mid \mathfrak{S}_{n-1}) &> \delta_{t_i}/\rho_{t_i} \text{ and} \\ P(\varepsilon_{t_i}^* \in D_{\epsilon'}^- \mid \mathfrak{S}_{n-1}) &> \delta_{t_i}/\rho_{t_i} \text{ a.s.} \end{aligned} \tag{2.40}$$

Let there exists $M > N$, and let $\{I_{k_i}\}$ be the sequence of infinite number of intervals $[n-1, n-M]$ such that no updates take place.

Let $A_{t_i}^-$ be the event that $\varepsilon_{t_i}^* \in D_{\epsilon'}^-$ and $\tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0$

Let $A_{t_i}^+$ be the event that $\varepsilon_{t_i}^* \in D_{\epsilon'}^+$ and $\tilde{\theta}_{t_i-1}^T x_{t_i} < 0$

Let $A_{t_i} \stackrel{\text{def}}{=} A_{t_i}^+ \cup A_{t_i}^-$

Note that $\tilde{\theta}_{t_i-1}^T x_{t_i}$ is $\mathfrak{F}_{n-1}$ measurable.

Since $A_{t_i}^+$ and $A_{t_i}^-$ are mutually exclusive events for a sufficiently small ϵ' , then

$$P(A_{t_i}) = P(A_{t_i}^-) + P(A_{t_i}^+)$$

$$\begin{aligned} P(A_{t_i}) &= P\left(\varepsilon_{t_i}^* \in D_{\epsilon'}^- \mid \tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0\right) P\left(\tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0\right) \\ &\quad + P\left(\varepsilon_{t_i}^* \in D_{\epsilon'}^+ \mid \tilde{\theta}_{t_i-1}^T x_{t_i} < 0\right) P\left(\tilde{\theta}_{t_i-1}^T x_{t_i} < 0\right) \end{aligned}$$

Let $k_i = t_i - t_{i-1}$ and define $W_{t_i-1} \stackrel{\text{def}}{=} \left\{ \tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0 \right\}$ and $V_{t_i} = \{\varepsilon_{t_i}^* \in D_{\epsilon'}^-\}$

Then V_{t_i} and $W_{t_i-1} \in \mathfrak{F}$ and $P(V_{t_i}) = \delta_{t_i}$. By the definition of the mixing condition in Definition 2.17

$$\lim_{k_i \rightarrow \infty} P(V_{t_i} \cap W_{t_i-1}) = P(V_{t_i}) \lim_{k_i \rightarrow \infty} P(W_{t_i-k_i}) \quad (2.41)$$

Suppose $P(W_{t_i-k_i})$ is not zero. Then (2.41) can be written as

$$\begin{aligned} \lim_{k_i \rightarrow \infty} \frac{P(V_{t_i} \cap W_{t_i-k_i})}{P(W_{t_i-k_i})} &= \lim_{k_i \rightarrow \infty} P(V_{t_i} \mid W_{t_i-k_i}) \\ &= P(V_{t_i}) \end{aligned}$$

Let $0 < \xi_{t_i} < \delta_{t_i}$ and there exists an N such that

$$|P(V_{t_i} \mid W_{t_i-k_i}) - P(V_{t_i})| < \xi_{t_i}$$

Therefore

$$P(V_{t_i}) - \xi_{t_i} < P(V_{t_i} \mid W_{t_i-N}) < P(V_{t_i}) + \xi_{t_i} \quad (2.42)$$

Note that the inequality of (2.42) holds only if no updates take place on the interval $[t_t - N, t_i]$.

Recall the definition of A_{t_i} , and thus we can write

$$\begin{aligned} P(A_{t_i}) &= P\left(\varepsilon_{t_i}^* \in D_{\epsilon'}^- \mid \tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0\right) P\left(\tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0\right) \\ &\quad + P\left(\varepsilon_{t_i}^* \in D_{\epsilon'}^+ \mid \tilde{\theta}_{t_i-1}^T x_{t_i} < 0\right) P\left(\tilde{\theta}_{t_i-1}^T x_{t_i} < 0\right) \end{aligned}$$

From (2.42) and recalling the definition of V_{t_i} and W_{t_i-N} , we get

$$\begin{aligned} P(A_{t_i}) &\geq [P(\varepsilon_{t_i}^* \in D_{\epsilon'}^-) - \xi_{t_i}] P\left(\tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0\right) \\ &\quad + [P(\varepsilon_{t_i}^* \in D_{\epsilon'}^+) - \xi_{t_i}] P\left(\tilde{\theta}_{t_i-1}^T x_{t_i} < 0\right) \end{aligned}$$

From PE condition and from (2.30) and (2.31)

$$P\left(\tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0\right) \geq \rho_{t_i} > 0$$

and

$$P\left(\tilde{\theta}_{t_i-1}^T x_{t_i} < 0\right) \geq \rho_{t_i} > 0$$

Also, from (2.32) and (2.33) we have

$$P\left(\varepsilon_{t_i}^* \in (D_{\epsilon'}^- \cup D_{\epsilon'}^+) \mid \tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0\right) \geq \delta_{t_i}/\rho_{t_i} > 0$$

and

$$P\left(\varepsilon_{t_i}^* \in (D_{\epsilon'}^- \cup D_{\epsilon'}^+) \mid \tilde{\theta}_{t_i-1}^T x_{t_i} < 0\right) \geq \delta_{t_i}/\rho_{t_i} > 0$$

Therefore, we have

$$\begin{aligned}
P(A_{t_i}) &\geq [P(\varepsilon_{t_i}^* \in D_{\epsilon'}^-) - \xi_{t_i}] \rho_{t_i} + [P(\varepsilon_{t_i}^* \in D_{\epsilon'}^+) - \xi_{t_i}] \rho_{t_i} \\
&\geq [P(\varepsilon_{t_i}^* \in (D_{\epsilon'}^- \cup D_{\epsilon'}^+)) - \xi_{t_i}] \rho_{t_i} \\
&\geq (\delta_{t_i} / \rho_{t_i} - \xi_{t_i}) \rho_{t_i} \\
&\geq \delta_{t_i} - \xi_{t_i} \rho_{t_i} \\
&> 0
\end{aligned}$$

and

$$\begin{aligned}
P(A_{t_i} \mid A_{t_i-N}) &\geq P\left[\varepsilon_{t_i}^* \in D_{\epsilon'}^- \mid \left\{\tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0\right\} \cap A_{t_i-N}\right] \\
&\quad P\left[\left\{\tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0\right\} \mid A_{t_i-N}\right] \\
&\quad + P\left[\varepsilon_{t_i}^* \in D_{\epsilon'}^+ \mid \left\{\tilde{\theta}_{t_i-1}^T x_{t_i} < 0\right\} \cap A_{t_i-N}\right] \\
&\quad P\left[\left\{\tilde{\theta}_{t_i-1}^T x_{t_i} < 0\right\} \mid A_{t_i-N}\right]
\end{aligned}$$

$$\begin{aligned}
&\geq P \left[\varepsilon_{t_i}^* \in D_{\epsilon'}^- \mid \left\{ \tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0 \right\} \mid A_{t_i-N} \right] \\
&\quad P \left[\tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0 \right] P \left[\left\{ \tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0 \right\} \mid A_{t_i-N} \right] \\
&\quad + P \left[\varepsilon_{t_i}^* \in D_{\epsilon'}^+ \mid \left\{ \tilde{\theta}_{t_i-1}^T x_{t_i} < 0 \right\} \mid A_{t_i-N} \right] \\
&\quad P \left[\tilde{\theta}_{t_i-1}^T x_{t_i} < 0 \right] P \left[\left\{ \tilde{\theta}_{t_i-1}^T x_{t_i} < 0 \right\} \mid A_{t_i-N} \right] \\
&\geq [P(\varepsilon_{t_i}^* \in D_{\epsilon'}^-) - \xi_{t_i}] \left\{ P \left[\tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0 \right] \right\}^2 \\
&\quad + [P(\varepsilon_{t_i}^* \in D_{\epsilon'}^+) - \xi_{t_i}] \left\{ P \left[\tilde{\theta}_{t_i-1}^T x_{t_i} < 0 \right] \right\}^2 \\
&\geq \delta_{t_i} - \xi_{t_i} \rho_{t_i} \\
&> 0
\end{aligned}$$

Thus,

$$\begin{aligned}
P(A_{t_i} \cap A_{t_i-N}) &= P(A_{t_i} \mid A_{t_i-N}) P(A_{t_i-N}) \\
&\geq (\delta_{t_i} - \xi_{t_i} \rho_{t_i}) (\delta_{t_i+N} - \xi_{t_i+N} \rho_{t_i+N}) \\
&> 0
\end{aligned}$$

$$\begin{aligned}
P(A_{t_i} \cap A_{t_i-1} \cap \dots \cap A_{t_i-N}) &\geq P(A_{t_i} \cap A_{t_i-N}) \\
&\geq (\delta_{t_i} - \xi_{t_i} \rho_{t_i}) (\delta_{t_i+N} - \xi_{t_i+N} \rho_{t_i+N}) \\
&> 0
\end{aligned} \tag{2.43}$$

$$\begin{aligned}
\lim_{k \rightarrow \infty} P(A_{t_i+kN} \cap A_{t_i+kN-1} \cap \dots \cap A_{t_i}) &\geq \lim_{k \rightarrow \infty} P(A_{t_i+kN} \cap A_{t_i+(k-1)N} \cap \dots \cap A_{t_i}) \\
&\geq \lim_{k \rightarrow \infty} \prod_{l=1}^k (\delta_{t_i+lN} - \xi_{t_i+lN} \rho_{t_i+lN}) \\
&> 0
\end{aligned}$$

In probability, this asserts our assumption that the volume of the ellipsoid does not converge to zero. Therefore, the sequence of hyperellipsoidal membership sets associated with QOBE cannot converge to a point set. $\diamond$

Although IU condition is a necessary and sufficient condition for the convergence of the central estimator in the i.i.d. case, it has not been explicitly proven in the mixing case. The following theorem is an alternative theorem to Theorem 2.4 asserting that even if IU is appended as an explicit condition, the volume of the ellipsoid of the QOBE algorithm does not converge to zero.

Theorem 2.5 *With the explicit condition that there exists a sequence $\{t_k\}_{k=1}^{\infty}$ such that $|\varepsilon_{t_k}| > \gamma_{t_k}$ for some positive sequence $\{\gamma_{t_k}\}$ appended to those of Theorem 2.4, the volume of the ellipsoid of the QOBE algorithm does not converge to zero.*

Proof: Let $\mathcal{N}' \stackrel{\text{def}}{=} \{n \in \mathcal{N} : c_n < 0\}$ and consider only $n \in \mathcal{N}'$ that implies that an update will take place; $\gamma_n^2 - \varepsilon_n^2 < 0$. Let us define $\{t_k\}_{k=1}^{\infty}$ to be the sequence over which IU (infinite update) take place. This will be helpful in asserting a contradiction later on.

Since $\varepsilon_{t_i} = \varepsilon_{t_i}^* + \tilde{\theta}_{t_i-1}^T x_{t_i}$ we have

$$\begin{aligned}
\gamma_{t_i}^2 - \varepsilon_{t_i}^2 &= (\gamma_{t_i} + |\varepsilon_{t_i}|)(\gamma_{t_i} - |\varepsilon_{t_i}|) \\
&= (\gamma_{t_i} + |\varepsilon_{t_i}|) \left(\gamma_{t_i} - \left| \varepsilon_{t_i}^* + \tilde{\theta}_{t_i-1}^T x_{t_i} \right| \right) < 0
\end{aligned}$$

Thus

$$\gamma_{t_i} - \left| \varepsilon_{t_i}^* + \tilde{\theta}_{t_i-1}^T x_{t_i} \right| < 0 \quad (2.44)$$

Therefore

$$\gamma_{t_i}^2 - \varepsilon_{t_i}^2 < 0 \text{ for all } \{t_i\} \text{ if and only if}$$

$$\left| \varepsilon_{t_i}^* + \tilde{\theta}_{t_i-1}^T x_{t_i} \right| > \gamma_{t_i} + \epsilon \text{ for sufficiently small } \epsilon > 0 \quad (2.45)$$

Let us assume that the volume of the ellipsoid of QOBE converges to zero and try to show that this assumption is not true. If the volume of the ellipsoid of QOBE converges to zero then condition (2.44) must hold. For convenience, we will denote ϵ' - neighborhoods of noise bounds as

$$\begin{aligned} D_{\epsilon'}^+ &= [-\gamma_n - \epsilon' < \varepsilon_n^* < -\gamma_n] \\ D_{\epsilon'}^- &= [\gamma_n < \varepsilon_n^* < \gamma_n + \epsilon'] \end{aligned}$$

At IU we have

$$\begin{aligned} \varepsilon_n^2 &> \gamma_n^2 \\ \left| \varepsilon_n^* + \tilde{\theta}_{n-1}^T x_n \right| &> \gamma_n + \epsilon' \text{ for small } \epsilon' \end{aligned} \quad (2.46)$$

If $\tilde{\theta}_{n-1}^T x_n \geq 0$ and $\varepsilon_n^* \in D_{\epsilon'}^-$, i.e. $(-\gamma_n - \epsilon' < \varepsilon_n^* < -\gamma_n)$ then since ε_n^* is uniformly distributed, we have

$$\gamma_n + \epsilon' < |\varepsilon_n^*| < \gamma_n$$

$$\begin{aligned} |\varepsilon_n^*| + \left| \tilde{\theta}_{n-1}^T x_n \right| &> \gamma_n + \epsilon' + \left| \tilde{\theta}_{n-1}^T x_n \right| \\ \left| \varepsilon_n^* + \tilde{\theta}_{n-1}^T x_n \right| &> \gamma_n + \epsilon' \end{aligned}$$

Same as defined in (2.46) , therefore $\varepsilon_n^* \in D_{\epsilon'}^-$. Similar analysis can be repeated to show that $\varepsilon_n^* \in D_{\epsilon'}^+$ when $\tilde{\theta}_{n-1}^T x_n < 0$.

Given ϵ' , assume $\{\varepsilon_{t_i}^*\}$ is symmetrically bounded with UCT, where there exists $\delta_{t_i} > 0$, $\rho_{t_i} > 0$ such that for all t_i

$$P(\varepsilon_{t_i}^* \in D_{\epsilon'}^+ \mid \mathfrak{F}_{n-1}) > \delta_{t_i}/\rho_{t_i} \quad \text{and} \quad (2.47)$$

$$P(\varepsilon_{t_i}^* \in D_{\epsilon'}^- \mid \mathfrak{F}_{n-1}) > \delta_{t_i}/\rho_{t_i} \quad a.s.$$

Let $A_{t_i}^-$ be the event that $\varepsilon_{t_i}^* \in D_{\epsilon'}^-$ and $\tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0$

Let $A_{t_i}^+$ be the event that $\varepsilon_{t_i}^* \in D_{\epsilon'}^+$ and $\tilde{\theta}_{t_i-1}^T x_{t_i} < 0$

Let $A_{t_i} \stackrel{\text{def}}{=} A_{t_i}^+ \cup A_{t_i}^-$

Now let us define A_{t_i} to be the event that (2.44) holds at time t_i , then

$P(A_{t_i}) = P(A_{t_i}^-) + P(A_{t_i}^+)$, since $A_{t_i}^+$ and $A_{t_i}^-$ are mutually exclusive events for a sufficiently small ϵ'

$$\begin{aligned} P(A_{t_i}) &= P\left(\varepsilon_{t_i}^* \in D_{\epsilon'}^- \mid \tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0\right) P\left(\tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0\right) \\ &\quad + P\left(\varepsilon_{t_i}^* \in D_{\epsilon'}^+ \mid \tilde{\theta}_{t_i-1}^T x_{t_i} < 0\right) P\left(\tilde{\theta}_{t_i-1}^T x_{t_i} < 0\right) \end{aligned}$$

Let $k_i = t_i - t_{i-1}$ and define $W_{t_{i-1}} \stackrel{\text{def}}{=} \left\{ \tilde{\theta}_{t_{i-1}}^T x_{t_i} \geq 0 \right\}$ and $V_{t_i} = \{\varepsilon_{t_i}^* \in D_{\epsilon'}^-\}$

Then V_{t_i} and $W_{t_{i-1}} \in \mathfrak{F}$ and $P(V_{t_i}) = \delta_{t_i}$. By the definition of the mixing condition in Definition (2.17)

$$\lim_{k_i \rightarrow \infty} P(V_{t_i} \cap W_{t_{i-1}}) = P(V_{t_i}) \lim_{k_i \rightarrow \infty} P(W_{t_{i-1}}) \quad (2.48)$$

Suppose $P(W_{t_i-k_i})$ is not zero. Then (2.48) can be written as

$$\begin{aligned} \lim_{k_i \rightarrow \infty} \frac{P(V_{t_i} \cap W_{t_i-k_i})}{P(W_{t_i-k_i})} &= \lim_{k_i \rightarrow \infty} P(V_{t_i} | W_{t_i-k_i}) \\ &= P(V_{t_i}) \end{aligned}$$

Let $0 < \frac{\xi_{t_i}}{\rho_{t_i}} < \delta_{t_i}$ such that

$$|P(V_{t_i} | W_{t_i-k_i}) - P(V_{t_i})| < \xi_{t_i}/\rho_{t_i}$$

Therefore

$$P(V_{t_i}) - \frac{\xi_{t_i}}{\rho_{t_i}} < P(V_{t_i} | W_{t_i-k_i}) < P(V_{t_i}) + \frac{\xi_{t_i}}{\rho_{t_i}} \quad (2.49)$$

The inequality of (2.49) holds for IU condition.

Therefore we can write

$$\begin{aligned} P(A_{t_i}) &= P\left(\varepsilon_{t_i}^* \in D_{\epsilon'}^- \mid \tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0\right) P\left(\tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0\right) \\ &\quad + P\left(\varepsilon_{t_i}^* \in D_{\epsilon'}^+ \mid \tilde{\theta}_{t_i-1}^T x_{t_i} < 0\right) P\left(\tilde{\theta}_{t_i-1}^T x_{t_i} < 0\right) \end{aligned}$$

From (2.49) and recalling the definition of V_{t_i} and $W_{t_i-k_i}$, we get

$$\begin{aligned} P(A_{t_i}) &\geq \left[P(\varepsilon_{t_i}^* \in D_{\epsilon'}^-) - \frac{\xi_{t_i}}{\rho_{t_i}} \right] P\left(\tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0\right) \\ &\quad + \left[P(\varepsilon_{t_i}^* \in D_{\epsilon'}^+) - \frac{\xi_{t_i}}{\rho_{t_i}} \right] P\left(\tilde{\theta}_{t_i-1}^T x_{t_i} < 0\right) \end{aligned}$$

From PE condition and from (2.30) and (2.31)

$$P\left(\tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0\right) \geq \rho_{t_i} > 0$$

and

$$P\left(\tilde{\theta}_{t_i-1}^T x_{t_i} < 0\right) \geq \rho_{t_i} > 0$$

Also, from (2.32) and (2.33) we have

$$P\left(\varepsilon_{t_i}^* \in (D_{\epsilon'}^- \cup D_{\epsilon'}^+) \mid \tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0\right) \geq \delta_{t_i}/\rho_{t_i} > 0$$

and

$$P\left(\varepsilon_{t_i}^* \in (D_{\epsilon'}^- \cup D_{\epsilon'}^+) \mid \tilde{\theta}_{t_i-1}^T x_{t_i} < 0\right) \geq \delta_{t_i}/\rho_{t_i} > 0$$

Therefore, we have

$$\begin{aligned} P(A_{t_i}) &\geq \left[P(\varepsilon_{t_i}^* \in D_{\epsilon'}^-) - \frac{\xi_{t_i}}{\rho_{t_i}} \right] \rho_{t_i} + \left[P(\varepsilon_{t_i}^* \in D_{\epsilon'}^+) - \frac{\xi_{t_i}}{\rho_{t_i}} \right] \rho_{t_i} \\ &\geq \left[P(\varepsilon_{t_i}^* \in (D_{\epsilon'}^- \cup D_{\epsilon'}^+)) - \frac{\xi_{t_i}}{\rho_{t_i}} \right] \rho_{t_i} \\ &\geq \left(\delta_{t_i}/\rho_{t_i} - \frac{\xi_{t_i}}{\rho_{t_i}} \right) \rho_{t_i} \\ &\geq \delta_{t_i} - \xi_{t_i} \\ &> 0 \end{aligned}$$

Let there exists $M > N$ and $M \rightarrow \infty$ and $\{I_{t_i}\}$ be the infinite sequence of intervals, each of length M over which IU occur. Then

$$\begin{aligned} P(A_{t_i} \mid A_{t_i-N}) &\geq P\left[\varepsilon_{t_i}^* \in D_{\epsilon'}^- \mid \left\{ \tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0 \right\} \cap A_{t_i-N}\right] \\ &\quad P\left[\left\{ \tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0 \right\} \mid A_{t_i-N}\right] \\ &\quad + P\left[\varepsilon_{t_i}^* \in D_{\epsilon'}^+ \mid \left\{ \tilde{\theta}_{t_i-1}^T x_{t_i} < 0 \right\} \cap A_{t_i-N}\right] \\ &\quad P\left[\left\{ \tilde{\theta}_{t_i-1}^T x_{t_i} < 0 \right\} \mid A_{t_i-N}\right] \end{aligned}$$

$$\begin{aligned}
&\geq P \left[\varepsilon_{t_i}^* \in D_{\epsilon'}^- \mid \left\{ \tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0 \right\} \mid A_{t_i-N} \right] \\
&\quad P \left[\tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0 \right] P \left[\left\{ \tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0 \right\} \mid A_{t_i-N} \right] \\
&\quad + P \left[\varepsilon_{t_i}^* \in D_{\epsilon'}^+ \mid \left\{ \tilde{\theta}_{t_i-1}^T x_{t_i} < 0 \right\} \mid A_{t_i-N} \right] \\
&\quad P \left[\tilde{\theta}_{t_i-1}^T x_{t_i} < 0 \right] P \left[\left\{ \tilde{\theta}_{t_i-1}^T x_{t_i} < 0 \right\} \mid A_{t_i-N} \right] \\
&\geq \left[P(\varepsilon_{t_i}^* \in D_{\epsilon'}^-) - \frac{\xi_{t_i}}{\rho_{t_i}} \right] \left\{ P \left[\tilde{\theta}_{t_i-1}^T x_{t_i} \geq 0 \right] \right\}^2 \\
&\quad + \left[P(\varepsilon_{t_i}^* \in D_{\epsilon'}^+) - \frac{\xi_{t_i}}{\rho_{t_i}} \right] \left\{ P \left[\tilde{\theta}_{t_i-1}^T x_{t_i} < 0 \right] \right\}^2 \\
&\geq \delta_{t_i} - \xi_{t_i} \\
&> 0
\end{aligned}$$

Thus,

$$\begin{aligned}
P(A_{t_i} \cap A_{t_i-N}) &= P(A_{t_i} \mid A_{t_i-N}) P(A_{t_i-N}) \\
&\geq (\delta_{t_i} - \xi_{t_i})(\delta_{t_i+N} - \xi_{t_i+N}) \\
&> 0
\end{aligned}$$

$$\begin{aligned}
P(A_{t_i} \cap A_{t_i-1} \cap \dots \cap A_{t_i-N}) &\geq P(A_{t_i} \cap A_{t_i-N}) \\
&\geq (\delta_{t_i} - \xi_{t_i})(\delta_{t_i+N} - \xi_{t_i+N}) \\
&> 0
\end{aligned} \tag{2.50}$$

$$\begin{aligned}
\lim_{k \rightarrow \infty} P(A_{t_i+kN} \cap A_{t_i+kN-1} \cap \dots \cap A_{t_i}) &\geq \lim_{k \rightarrow \infty} P(A_{t_i+kN} \cap A_{t_i+(k-1)N} \cap \dots \cap A_{t_i}) \\
&\geq \lim_{k \rightarrow \infty} \prod_{l=1}^k (\delta_{t_i+lN} - \xi_{t_i+lN}) = 0
\end{aligned}$$

In probability, this is contrary to our assumption that the volume of the ellipsoid converges to zero. Therefore, the volume of the ellipsoid of QOBE does not converge to zero under IU condition. $\diamond$

Chapter 3

Simulation Results

3.1 Introduction

Simulation studies provide a visual perspective of any mathematical analysis. In working with adaptive filtering techniques, the algorithm that models the filter's performance can be better understood and evaluated through experimenting with data and simulating the results on a digital computer. For example, simulations can illustrate improvements in the speed of convergence of the algorithm, represent the effect of different factors on the filter's estimated parameters, show efficient use of innovation in the data, and indicate order of computations.

A process x_n is said to be deterministic if it can be predicted without error in terms of a linear combination of its previous values. However, in many signal processing applications where linear shift-invariant filters are used, the inputs to these filters are random processes. In the simulation studies in this chapter we will investigate the performance of the QOBE algorithm in a stochastic setting and verify its behavior indicated by theorems that have been proven in the previous chapter. Different types of noise sequences will be used.

3.2 Types of Noise Sequences

In this section and its subsections we will illustrate the behavior and performance of the QOBE algorithm in iid and colored noise cases and when the noise bounds are either symmetrical or asymmetrical. For our purposes of simulation, we will use an AR(3) model of the form:

$$y_n = a_{1*} y_{n-1} + a_{2*} y_{n-2} + a_{3*} y_{n-3} + \varepsilon_n^* \quad (3.1)$$

where $a_{1*} = 2$, $a_{2*} = -1.48$, $a_{3*} = 0.34$ are unknown parameters to be identified. Some other model forms will also be used.

3.2.1 IID Noise Sequence

In most of the simulations presented in this section and others to follow, we will show the performance and characteristics of the QOBE algorithm in identification problem by way of comparison with those of the OBE algorithm under the same system settings. Since the performance of the OBE algorithm has been investigated with respect to other identification methods [7],[9],and [44] in the literature, such as recursive least squares (RLS) and least squares (LS), we will restrict our comparison to OBE in order to show the difference in behavior between the minimization criteria used in OBE and QOBE. We have used the set membership stochastic approximation (SMSA) version of OBE for it is used in the literature when stochastic settings are under investigation.

In Cases 1 to 3, we will consider an AR(3) model as in (3.1) where the noise sequences $\{\varepsilon_n^*\}$ used to generate the observable output sequence $\{y_n\}$ are specified within.

Case 1: $\varepsilon_n^* \sim U(-1, 1)$ is an i.i.d. sequence with zero mean and is uniformly distributed on $[-1, 1]$.

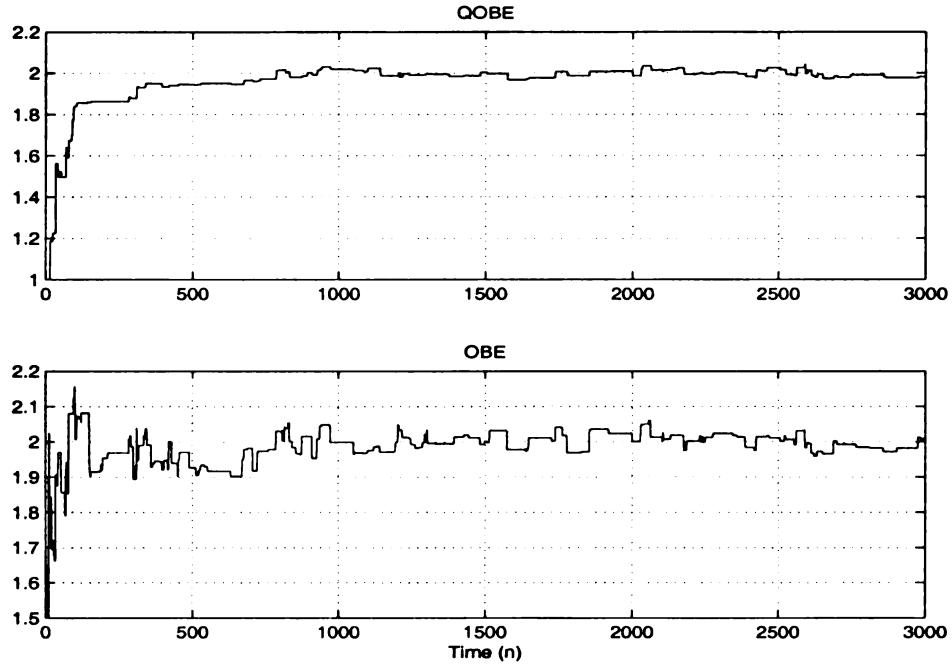


Figure 3.1: Case 1: Estimators of $a_{1*} = 2.0$ convergence using QOBE and OBE algorithms for identification of model (3.1) with $\gamma_n = 0.96$, where 146 points and 223 points were used, respectively, from a total of 3000 points.

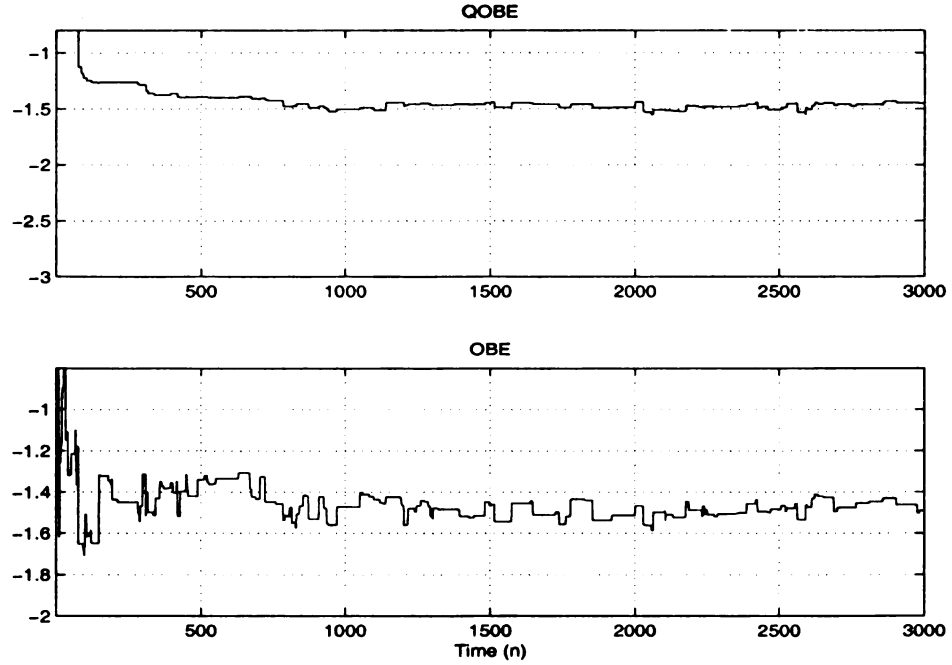


Figure 3.2: Case 1: Estimators of $a_{2*} = -1.48$ convergence using QOBE and OBE algorithms for identification of model (3.1) with $\gamma_n = 0.96$, where 146 points and 223 points were used, respectively, from a total of 3000 points.

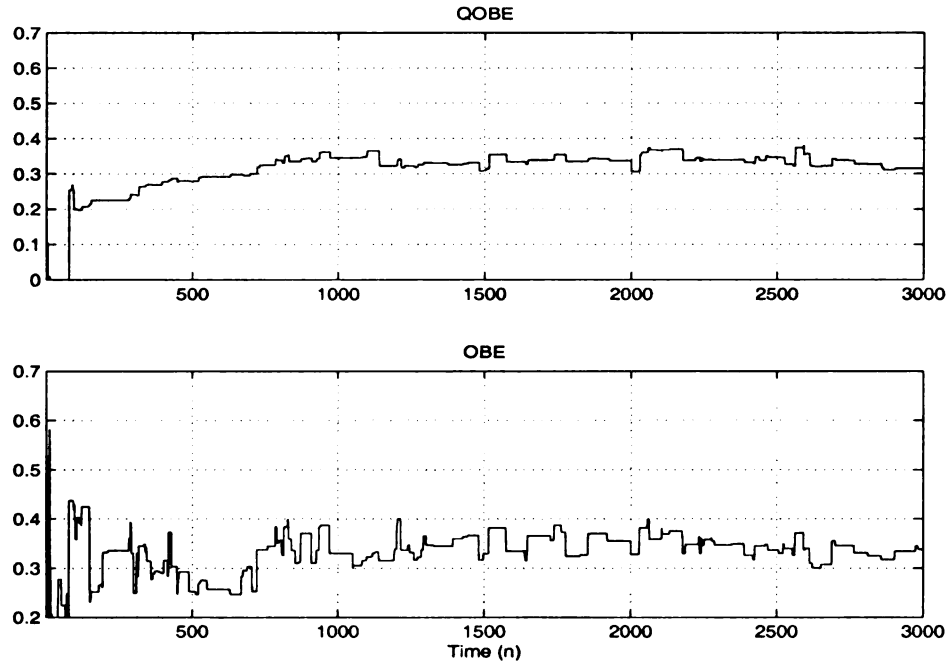


Figure 3.3: Case 1: Estimators of $a_{3*} = 0.34$ convergence using QOBE and OBE algorithms for identification of model (3.1) with $\gamma_n = 0.96$, where 146 points and 223 points were used, respectively, from a total of 3000 points.

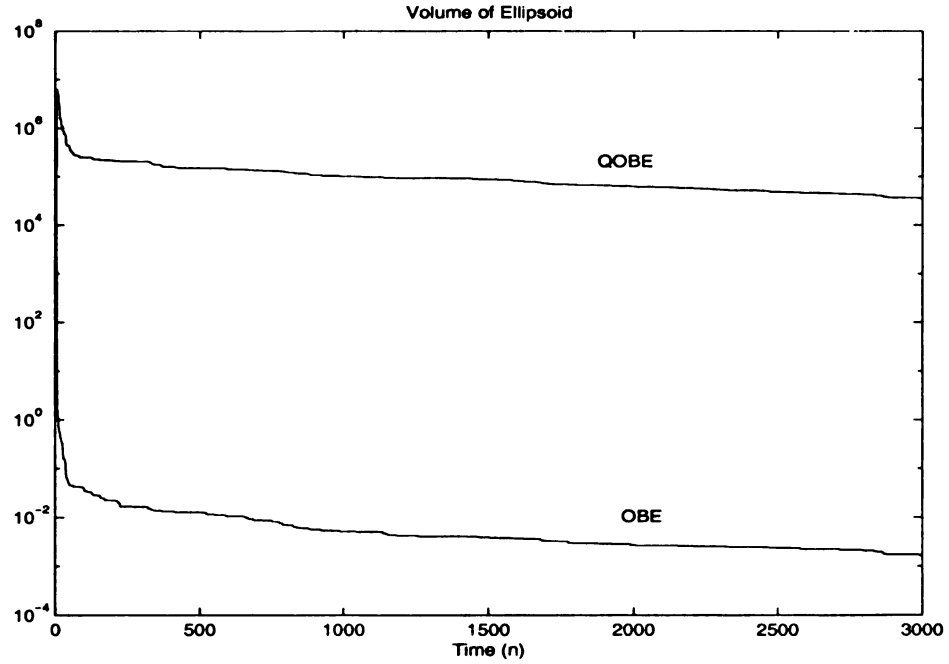


Figure 3.4: Case 1: Volume of the ellipsoids in QOBE and OBE algorithms for identification of model (3.1) with $\gamma_n = 0.96$, where 146 points and 223 points were used, respectively, from a total of 3000 points.

The simulation results of QOBE under the specified conditions and $\gamma_n = 0.96$ are shown in Figures 3.1 - 3.4 with relative comparison to results obtained from OBE identification. From the figures it is noticed that the estimators of a_{1*} , a_{2*} , and a_{3*} are converging to the true parameters in both algorithms with QOBE using fewer data points as expected. This supports the assertions in Theorems 2.1 and 2.2 in Chapter 2. The volume of the ellipsoid in QOBE, Figure 3.4, is obvious to be larger than that of OBE, however, it is converging to a certain size.

Case 2: This is a case of symmetrical noise bounds where $\varepsilon_n^* \sim B(-1, 1)$ is an i.i.d. Bernoulli sequence with zero mean and a binary distribution specified as

$$\begin{aligned}\varepsilon_n^* &= 1 \text{ with probability } 0.7 \\ \varepsilon_n^* &= -1 \text{ with probability } 0.3\end{aligned}\tag{3.2}$$

The simulation results of both QOBE and OBE under the specified conditions are shown in Figures 3.5 - 3.8 with $\gamma_n = 1.04$. It is seen that the QOBE algorithm has a faster convergence rate for the estimated parameters to their true counterparts although the speed of both algorithms are not very much affected by the ϵ - tail probability, δ , defined in UCT (2.6) or UT (2.7) conditions.

Case 3: This is a case of asymmetrical noise bounds where $\varepsilon_n^* \sim B(-0.5, 1)$ is an i.i.d. Bernoulli sequence with non-zero mean and is asymmetrically bounded with binary distribution specified as

$$\begin{aligned}\varepsilon_n^* &= 1 \text{ with probability } 0.7 \\ \varepsilon_n^* &= -0.5 \text{ with probability } 0.3\end{aligned}\tag{3.3}$$

The simulation results for both QOBE and OBE under the specified conditions and with $\gamma_n = 1.04$ are shown in Figures 3.9 - 3.12. It is evident that the OBE algorithm is affected by the asymmetrical noise bounds as defined in Definition 2.7

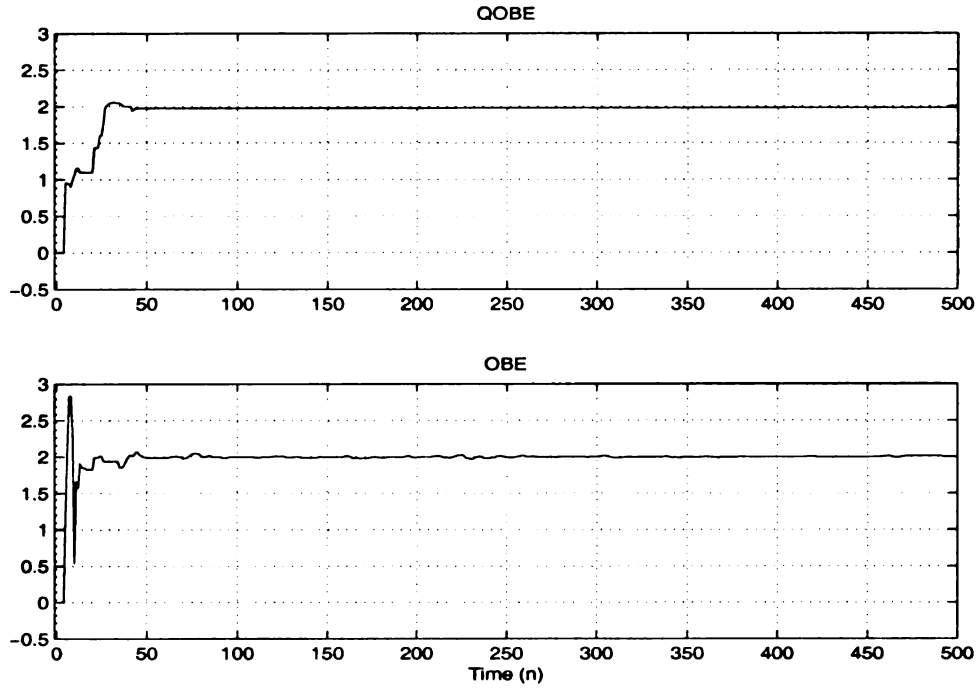


Figure 3.5: Case 2: Estimators of $a_{1*} = 2.0$ convergence using QOBE and OBE algorithms for identification of model (3.1), $\gamma_n = 1.04$, where 24 points and 916 points were used, respectively, from a total of 3000 points.

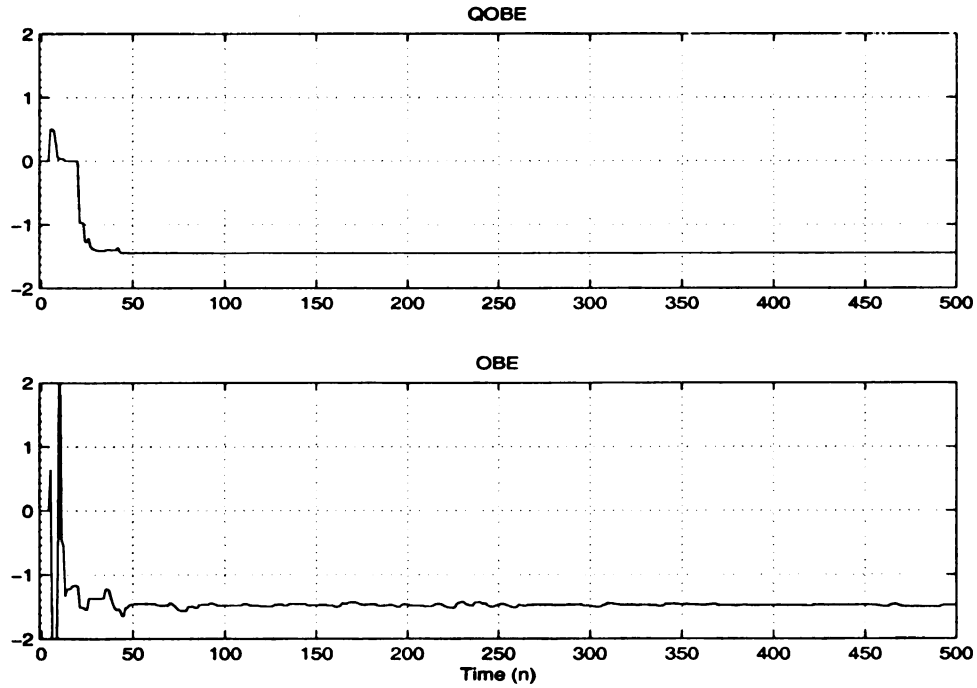


Figure 3.6: Case 2: Estimators of $a_{2*} = -1.48$ convergence using QOBE and OBE algorithms for identification of model (3.1), $\gamma_n = 1.04$, where 24 points and 916 points were used, respectively, from a total of 3000 points.

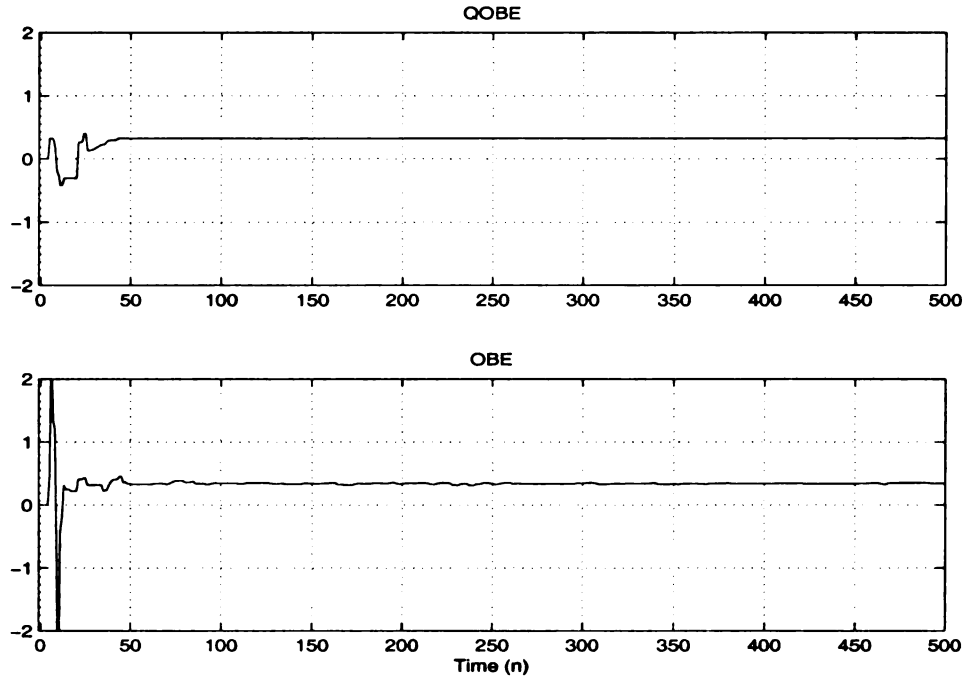


Figure 3.7: Case 2: Estimators of $a_{3*} = 0.34$ convergence using QOBE and OBE algorithms for identification of model (3.1), $\gamma_n = 1.04$, where 24 points and 916 points were used, respectively, from a total of 3000 points.

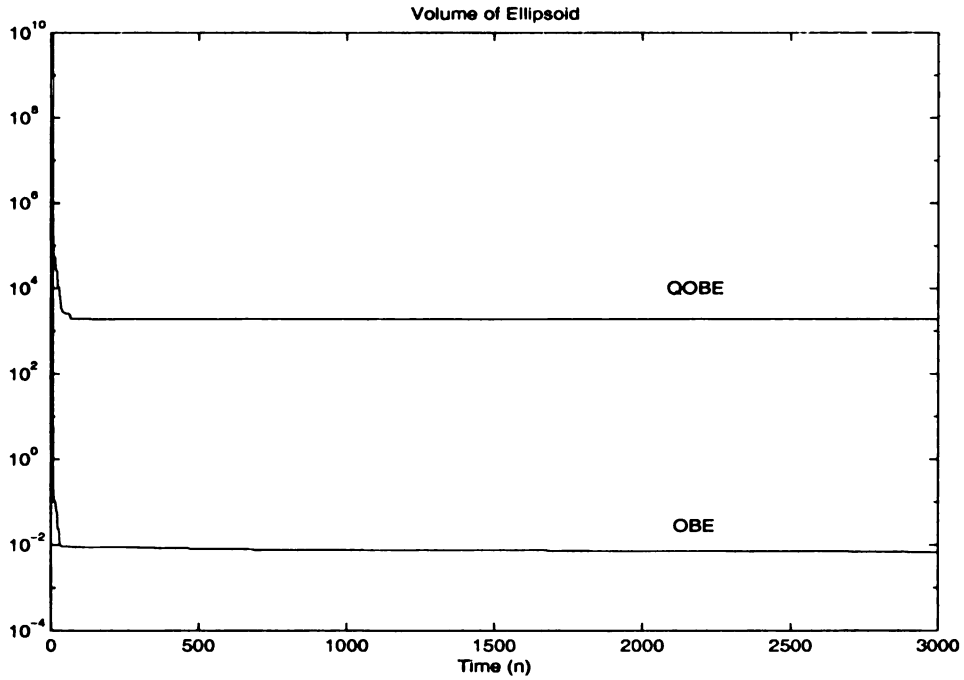


Figure 3.8: Case 2: Volume of the ellipsoids in QOBE and OBE algorithms for identification of model (3.1) with $\gamma_n = 1.04$, where 24 points and 916 points were used, respectively, from a total of 3000 points.

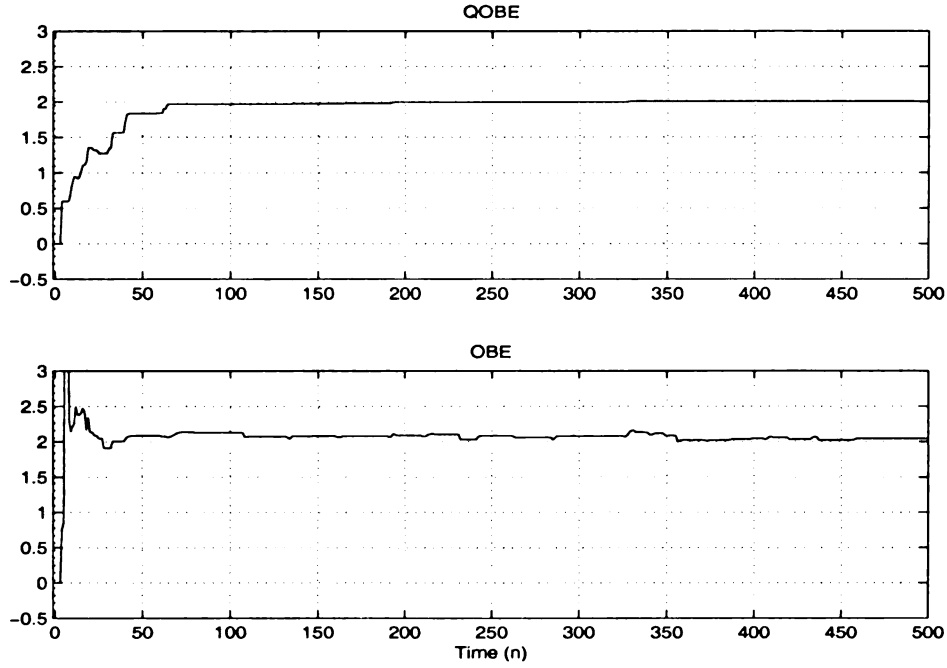


Figure 3.9: Case 3: Estimators of $a_{1*} = 2.0$ convergence using QOBE and OBE algorithms for identification of model (3.1), $\gamma_n = 1.04$, where 27 points and 374 points were used, respectively, from a total of 3000 points.

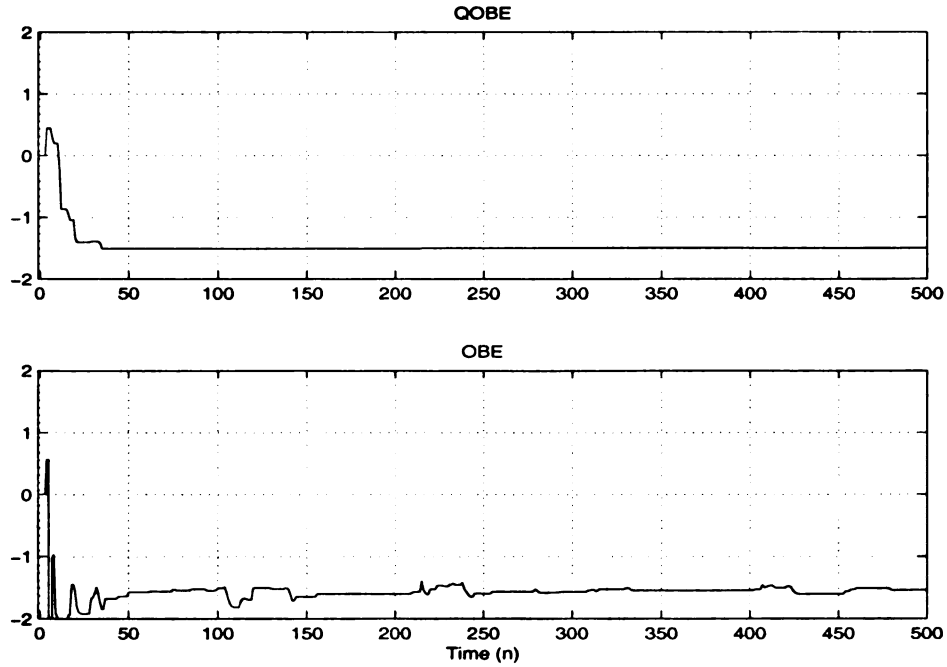


Figure 3.10: Case 3: Estimators of $a_{2*} = -1.48$ convergence using QOBE and OBE algorithms for identification of model (3.1), $\gamma_n = 1.04$, where 27 points and 374 points were used, respectively, from a total of 3000 points.

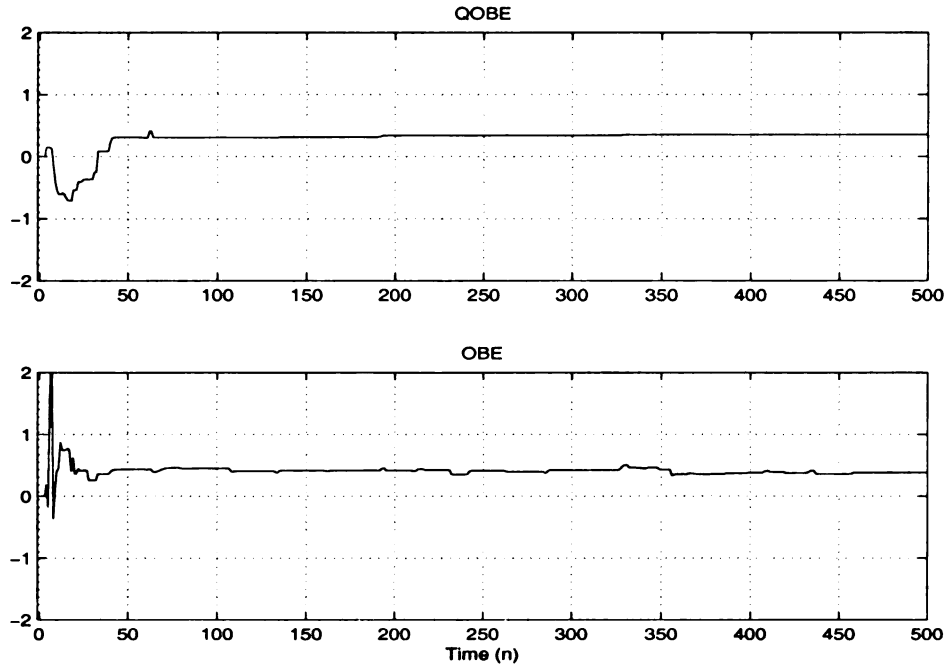


Figure 3.11: Case 3: Estimators of $a_{3*} = 0.34$ convergence using QOBE and OBE algorithms for identification of model (3.1), $\gamma_n = 1.04$, where 27 points and 374 points were used, respectively, from a total of 3000 points.

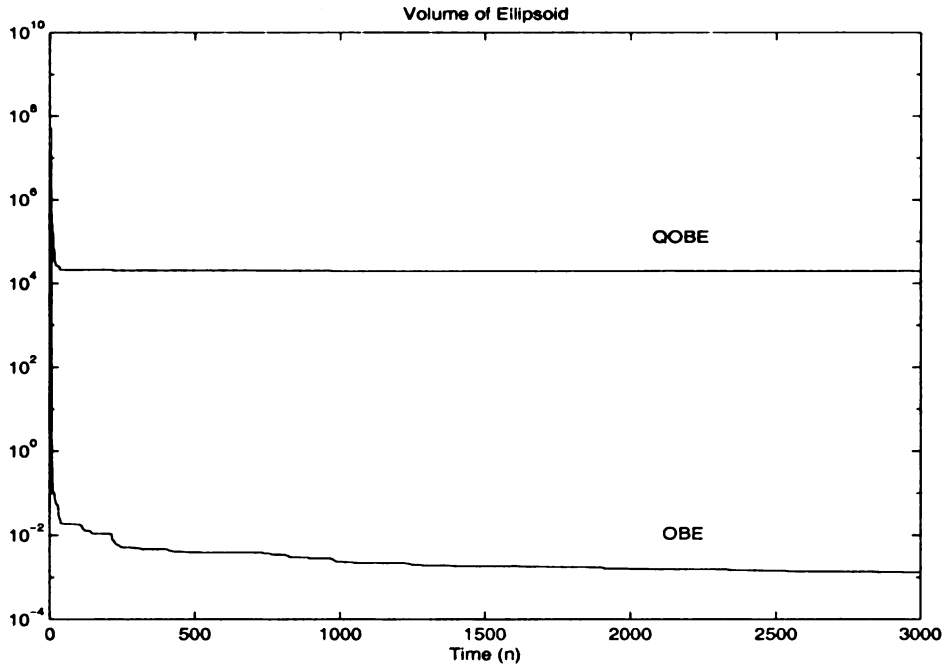


Figure 3.12: Case 3: Volume of the ellipsoids in QOBE and OBE algorithms for identification of model (3.1) with $\gamma_n = 1.04$, where 27 points and 374 points were used, respectively, from a total of 3000 points.

while QOBE converges consistently and complies with Theorems 2.1 and 2.2.

Case 4: We will use an AR(2) model of the form

$$y_n = -0.1 y_{n-1} + 0.56 y_{n-2} + \varepsilon_n^* \quad (3.4)$$

with ε_n^* as in Case 1, in order to illustrate the trajectory of the estimates and the final ellipsoid of the QOBE algorithm.

The simulations were performed for identification of an AR(2) model and are shown in Figures 3.13 - 3.15. The estimators of a_{1*} and a_{2*} are converging to the true parameters and the volume is decreasing but to a certain size that is not as small (or to a point) as it would be in OBE. This is also clear from the trajectory of the estimators in Figure 3.15 which conforms to the positiveness of P_n and κ_n under the PE assumptions.

3.2.2 Colored Noise Sequence

In this section we will consider the AR(3) model as in (3.1) in Cases 5 and 6 where the types of colored noise sequences used are described within. We will also compare the performance of QOBE and OBE when the noise is colored with both symmetrical and asymmetrical bounds.

Case 5: This is a case of symmetrical noise bounds where ε_n^* is a colored non-zero mean noise sequence that is related to a colored sequence $\{\omega_n\}$ by:

$$\begin{aligned} \varepsilon_n^* &= 1, \text{ if } \omega_n > -1 \\ \varepsilon_n^* &= -1, \text{ otherwise} \end{aligned} \quad (3.5)$$

where the colored sequence $\{\omega_n\}$ is generated by a uniformly distributed i.i.d. white

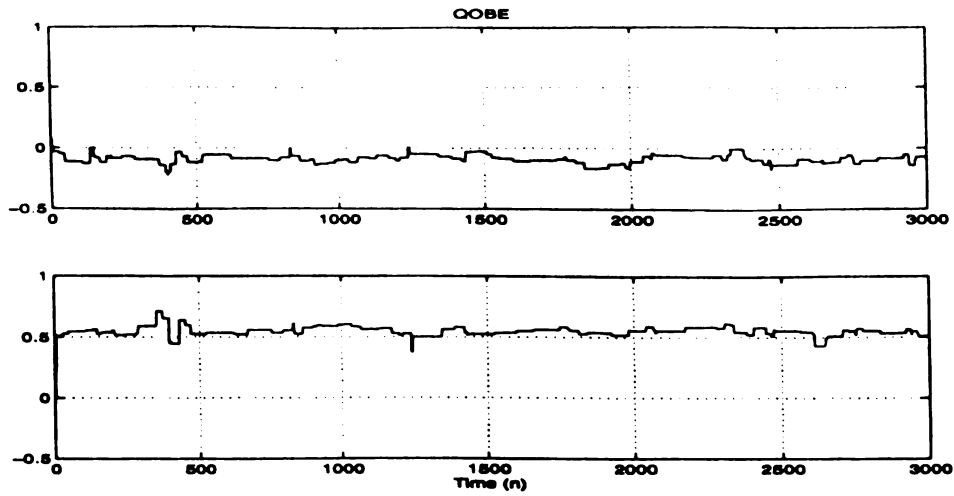


Figure 3.13: Case 4: Estimators of $a_{1*} = -0.1$ and $a_{2*} = 0.56$ convergence using QOBE for identification of model (3.4) with $\gamma_n = 0.96$, where 125 points were used from a total of 3000 points.

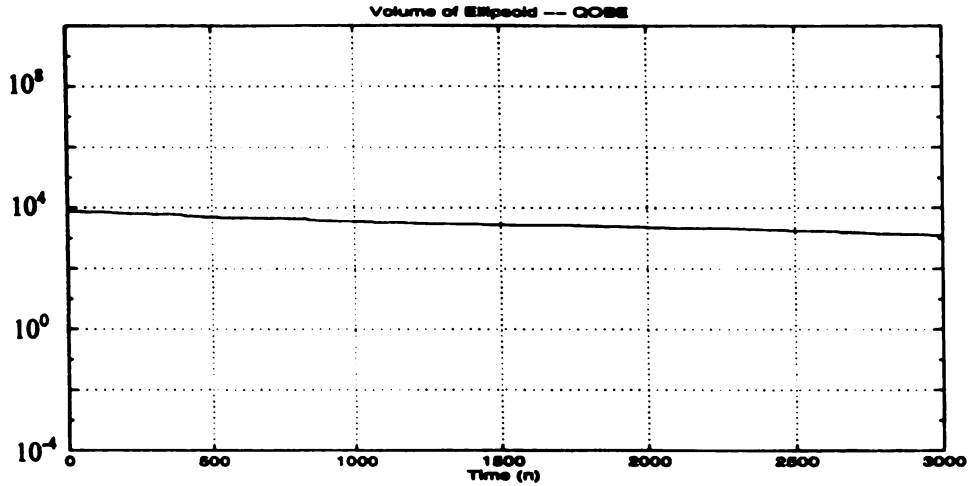


Figure 3.14: Case 4: Volume of the ellipsoid in QOBE algorithm for identification of model (3.4) with $\gamma_n = 0.96$, where 125 points were used from a total of 3000 points.

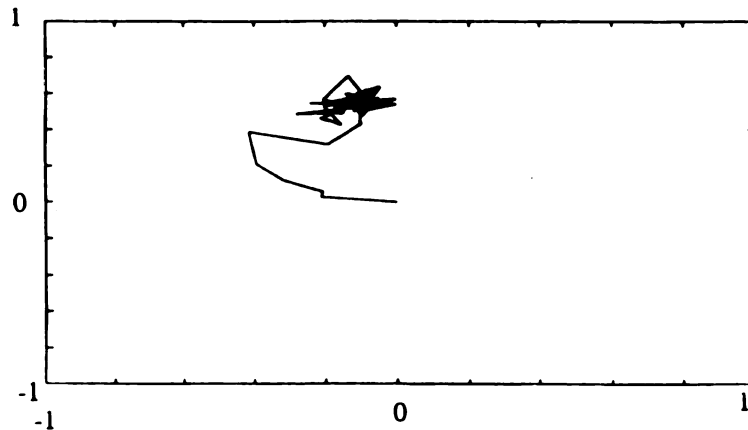


Figure 3.15: Case 4: The trajectory of the estimators in model (3.4) with $\gamma_n = 0.96$, where 125 points were used from a total of 3000 points.

noise sequence $z_n \sim U(-1, 1)$ as

$$\omega_n = -0.8\omega_{n-1} + z_n$$

The simulation results of QOBE under the specified conditions and $\gamma_n = 1.04$ are shown in Figures 3.16 - 3.19 with relative comparison to results obtained from OBE identification. From the figures it is noticed that the estimators of a_{1*} , a_{2*} , and a_{3*} are converging to the true parameters in both algorithms with neither of them being affected by the color of ε_n^* . This supports the assertions in Theorem 2.3 and others in Chapter 2.

Case 6: This is a case of asymmetrical noise bounds where ε_n^* is a colored non-zero mean noise sequence with asymmetrical bounds generated by colored sequence $\{\omega_n\}$ as follows:

$$\begin{aligned}\varepsilon_n^* &= 1, \text{ if } \omega_n > -0.5 \\ \varepsilon_n^* &= -0.5, \text{ otherwise}\end{aligned}\tag{3.6}$$

and where the colored sequence $\{\omega_n\}$ is generated as in Case 4.

The simulation results of both QOBE and OBE under the specified conditions are shown in Figures 3.20 - 3.23 with $\gamma_n = 1.04$. It is seen from these figures that OBE suffers from the condition of asymmetrical bounds of noise while QOBE converges rather consistently and faster as well to the true parameters. Again the volume of the ellipsoid in QOBE is obvious to be larger than that of OBE, however, it reaches its limited size faster than OBE.

Case 7: In this case we will use the same noise sequence as in Case 5 while we increase the order of the system to 12. Therefore, the stable AR(12) model

$$y_n = a_{1*} y_{n-1} + a_{2*} y_{n-2} + \cdots + a_{12*} y_{n-12} + \varepsilon_n^*\tag{3.7}$$

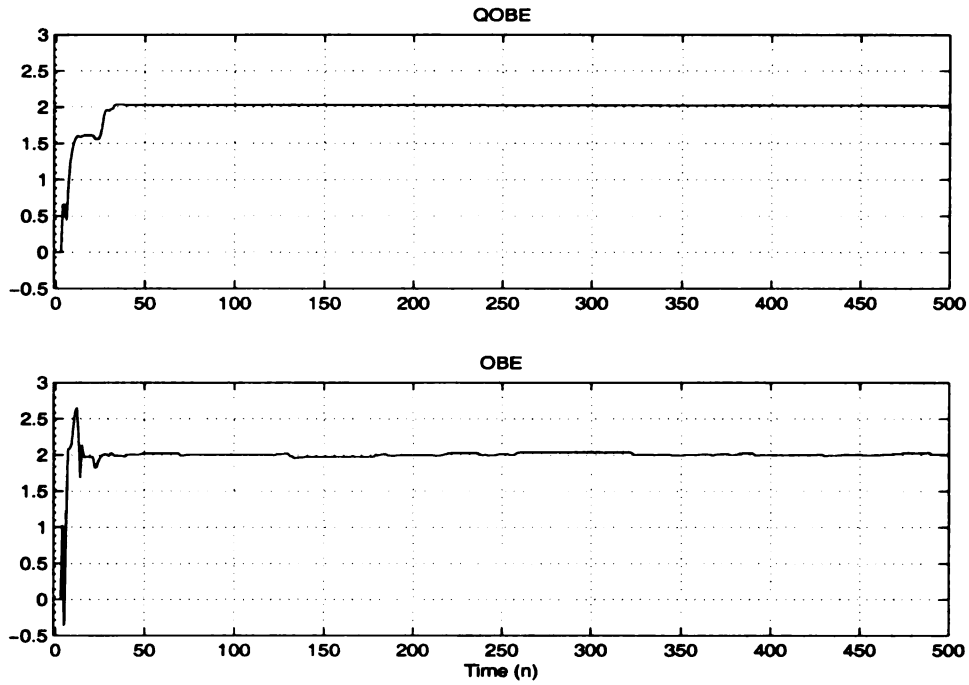


Figure 3.16: Case 5: Estimators of $a_{1*} = 2.0$ convergence using QOBE and OBE algorithms for identification of model (3.1), $\gamma_n = 1.04$, where 26 points and 845 points were used, respectively, from a total of 3000 points.

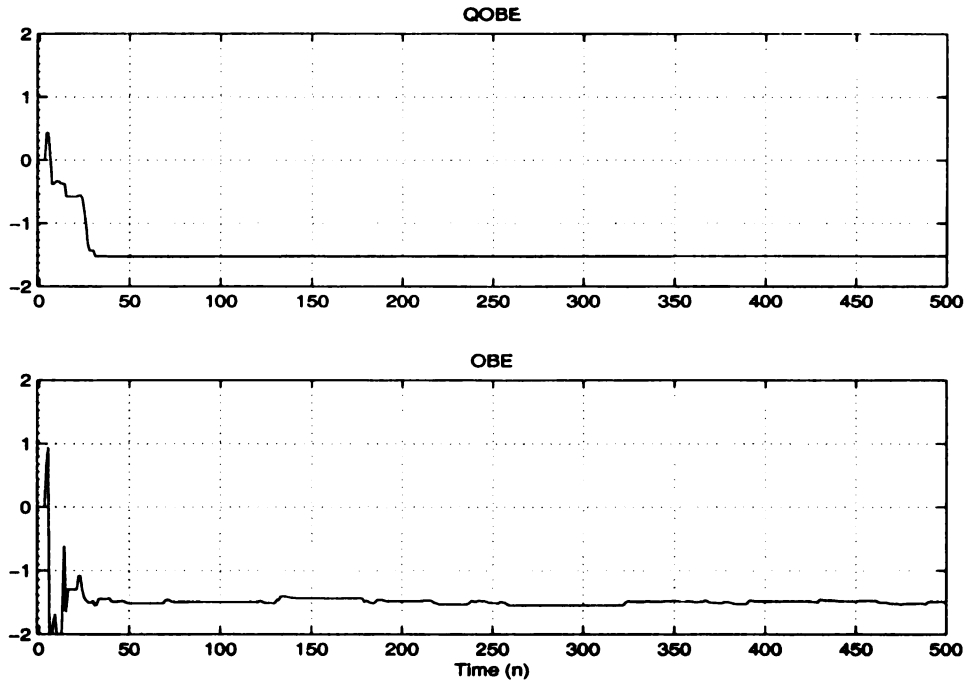


Figure 3.17: Case 5: Estimators of $a_{2*} = -1.48$ convergence using QOBE and OBE algorithms for identification of model (3.1), $\gamma_n = 1.04$, where 26 points and 845 points were used, respectively, from a total of 3000 points.

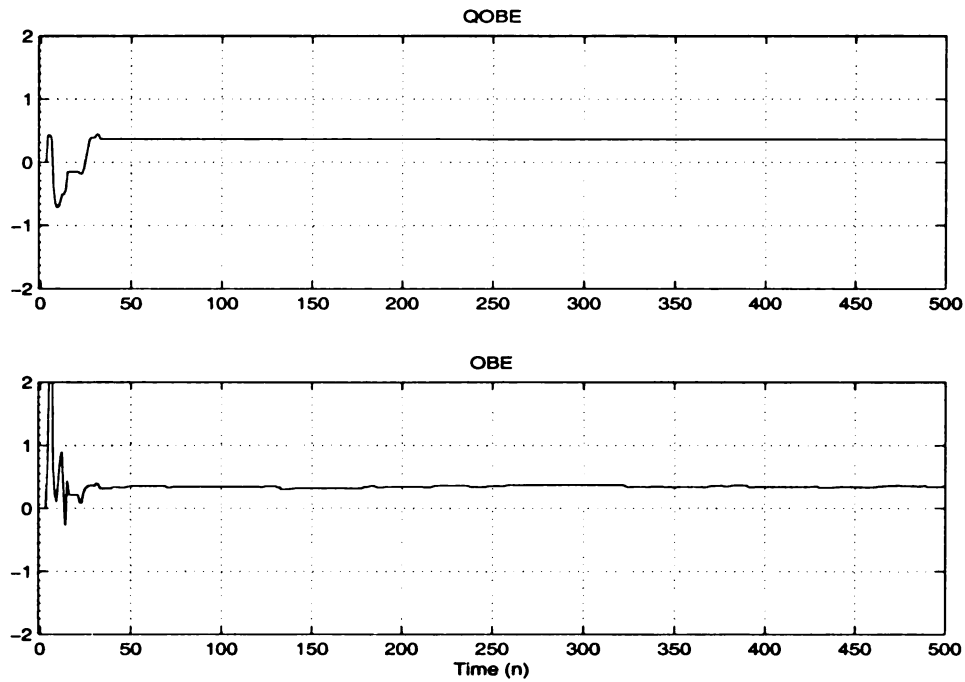


Figure 3.18: Case 5: Estimators of $a_{3*} = 0.34$ convergence using QOBE and OBE algorithms for identification of model (3.1), $\gamma_n = 1.04$, where 26 points and 845 points were used, respectively, from a total of 3000 points.

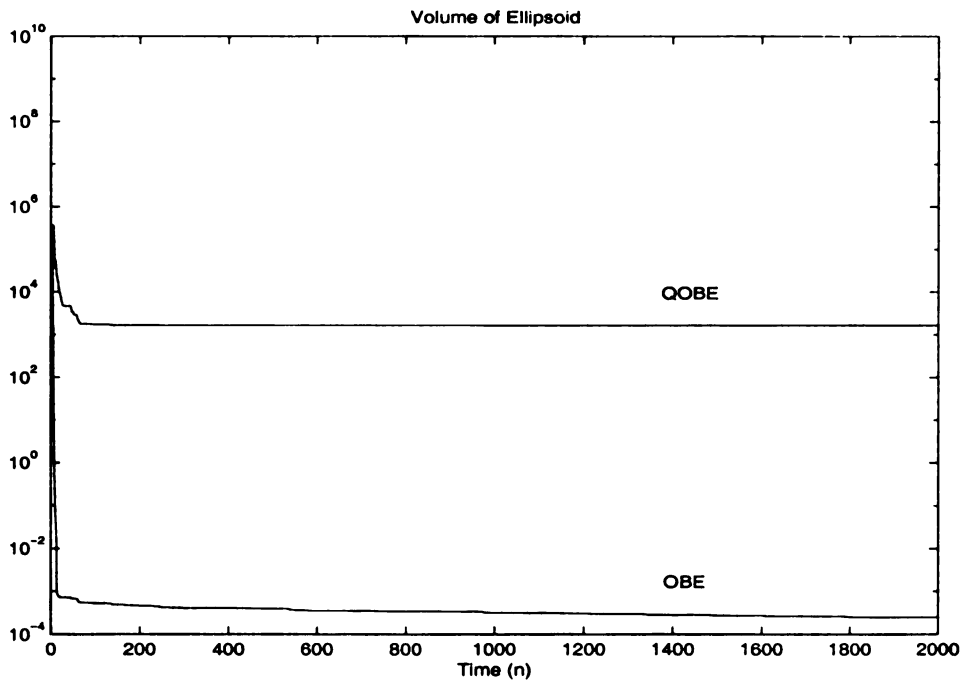


Figure 3.19: Case 5: Volume of the ellipsoids in QOBE and OBE algorithms for identification of model (3.1) with $\gamma_n = 1.04$, where 26 points and 845 points were used, respectively, from a total of 3000 points.

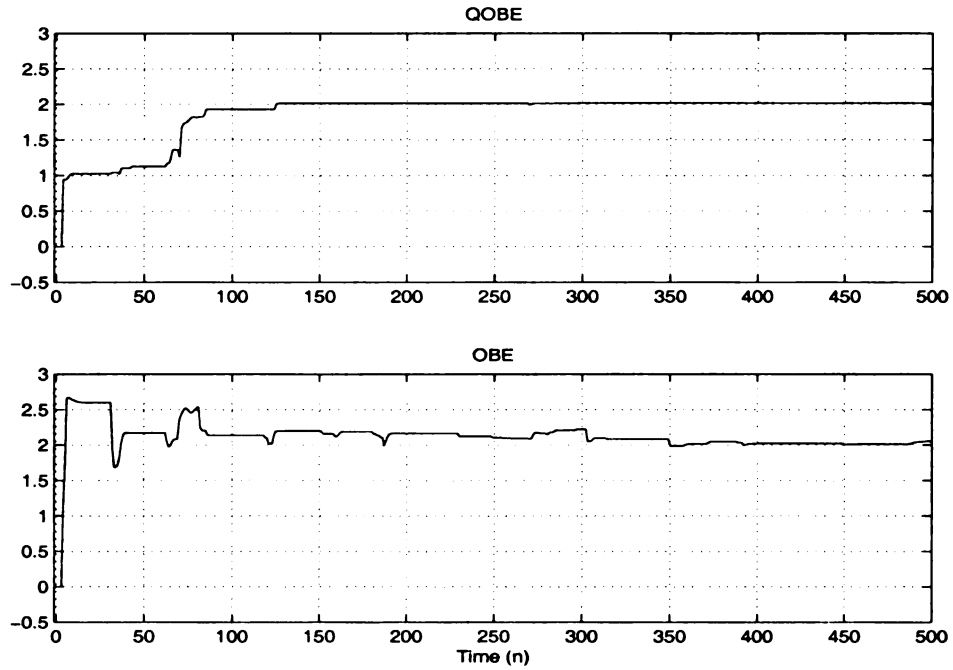


Figure 3.20: Case 6: Estimators of $a_{1*} = 2.0$ convergence using QOBE and OBE algorithms for identification of model (3.1), $\gamma_n = 1.04$, where 26 points and 207 points were used, respectively, from a total of 3000 points.

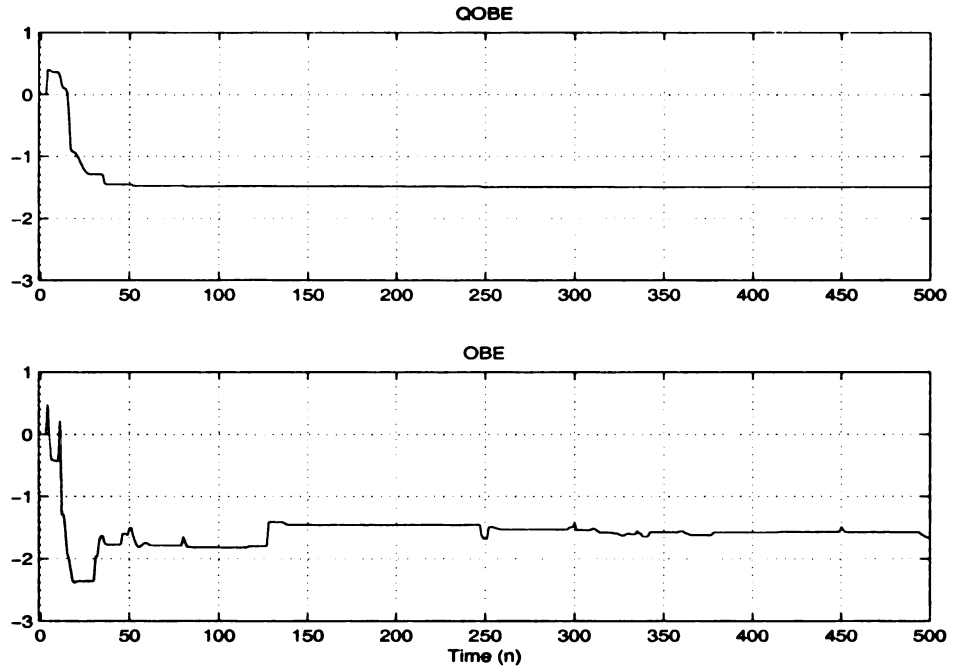


Figure 3.21: Case 6: Estimators of $a_{2*} = -1.48$ convergence using QOBE and OBE algorithms for identification of model (3.1), $\gamma_n = 1.04$, where 26 points and 207 points were used, respectively, from a total of 3000 points.

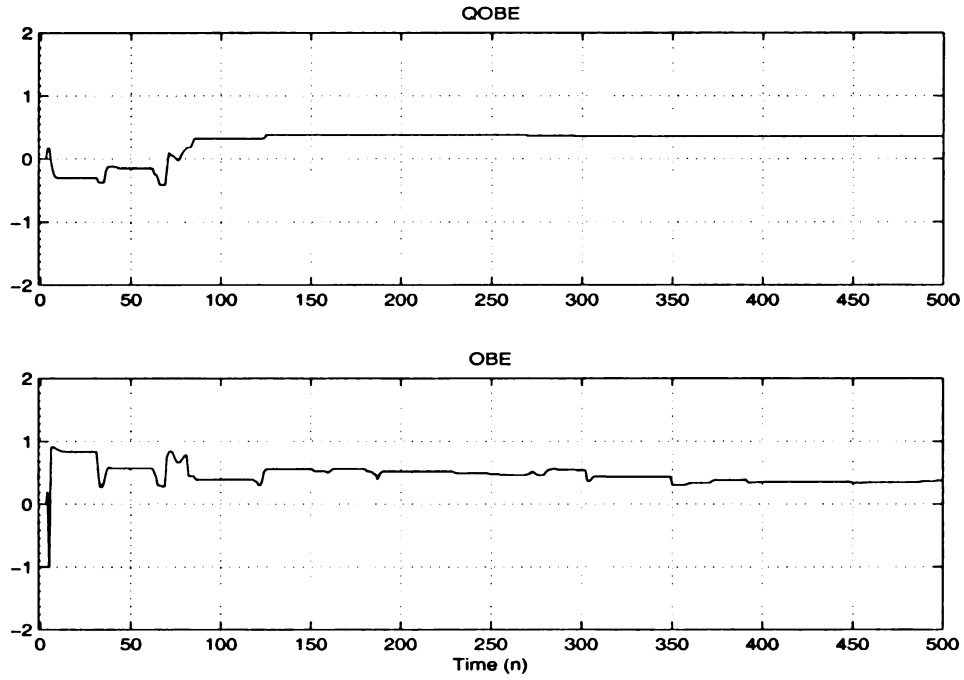


Figure 3.22: Case 6: Estimators of $a_{3*} = 0.34$ convergence using QOBE and OBE algorithms for identification of model (3.1), $\gamma_n = 1.04$, where 26 points and 207 points were used, respectively, from a total of 3000 points.

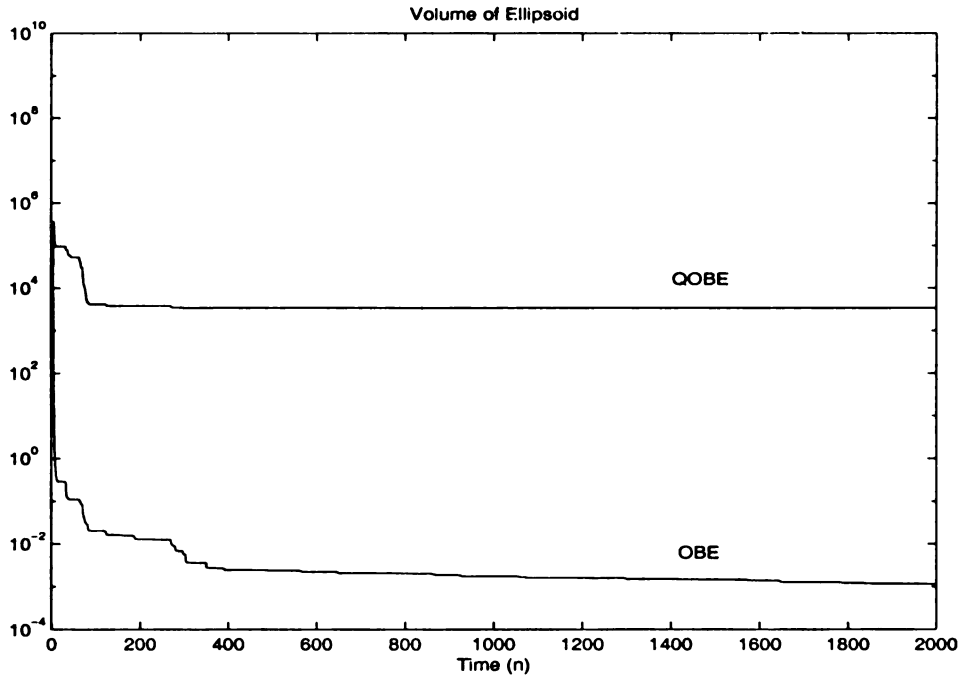


Figure 3.23: Case 6: Volume of the ellipsoids in QOBE and OBE algorithms for identification of model (3.1) with $\gamma_n = 1.04$, where 26 points and 207 points were used, respectively, from a total of 3000 points.

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

The simulations were performed for identification of an AR(12) model and are shown in Figures 3.24 - 3.25. The estimators of a_{1*} , a_{2*} and a_{5*} are shown as an example of the performance of the QOBE algorithm for all the other estimators as well. It is seen that the estimators are converging to the true parameters with a colored noise sequence and the volume of the ellipsoid is decreasing in size. These simulations show that the performance of the QOBE algorithm is not affected by a change in the model order.

Case 8: We will use the same AR(2) model as in Case 4 while employing the noise sequence in Case 5.

The simulations were performed for identification of an AR(2) model and are shown in Figures 3.26 - 3.28. The estimators of a_{1*} and a_{2*} are converging to the true parameters and the volume is decreasing but to a certain size that is not as small (or to a point) as it would be in OBE. This is also clear from the trajectory of the estimators in Figure 3.28 which conforms to the positiveness of P_n and κ_n under the PE assumptions.

In all of the above simulations it is shown that the number of usable data points are much less in the QOBE algorithms than in OBE and thus the number of updates for the estimators in QOBE is smaller than its counterpart in OBE. This leads to the fact that the QOBE algorithm has an improved computational efficiency due to its simple check test and its focus on minimizing the scalar κ_n rather than minimizing the volume as is the case in OBE algorithm. Therefore we can summarize the performance of the QOBE algorithm so far in that it retains a large volume, uses fewer points, exhibits faster parameter estimator convergence to the true parameter, is robust to additive noise, and its performance in colored noise is improved.

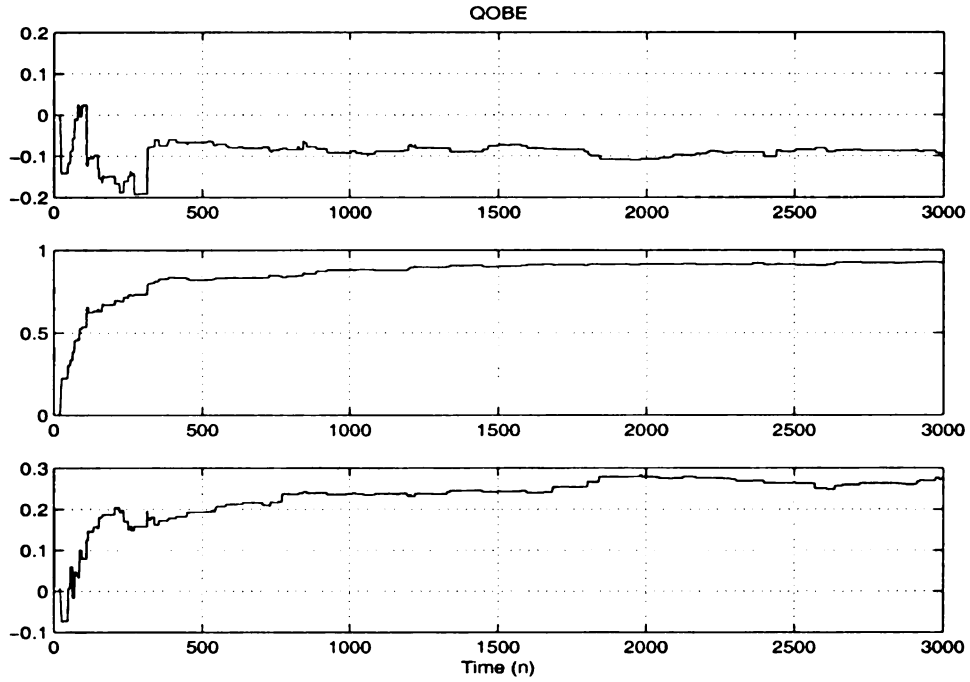


Figure 3.24: Case 7: Estimators of $a_{1*} = -0.1$ and $a_{2*} = 0.9175$ and $a_{5*} = 0.2601$ convergence using QOBE for identification of model (3.7) with $\gamma_n = 1.04$, where 115 points were used from a total of 3000 points.

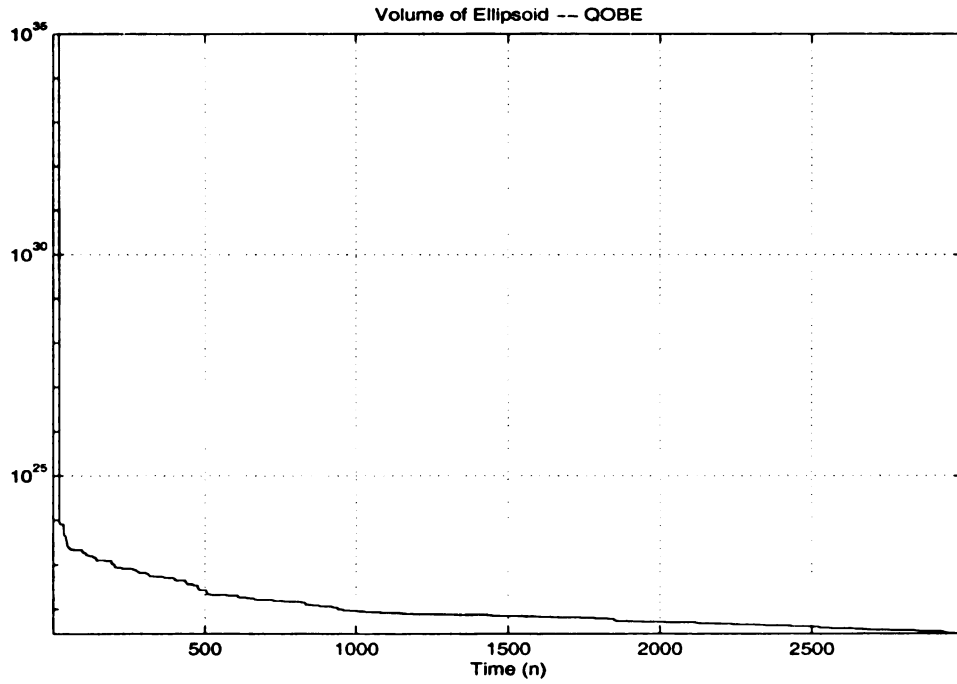


Figure 3.25: Case 7: Volume of the ellipsoid in QOBE algorithm for identification of model (3.7) with $\gamma_n = 1.04$, where 115 points were used from a total of 3000 points.

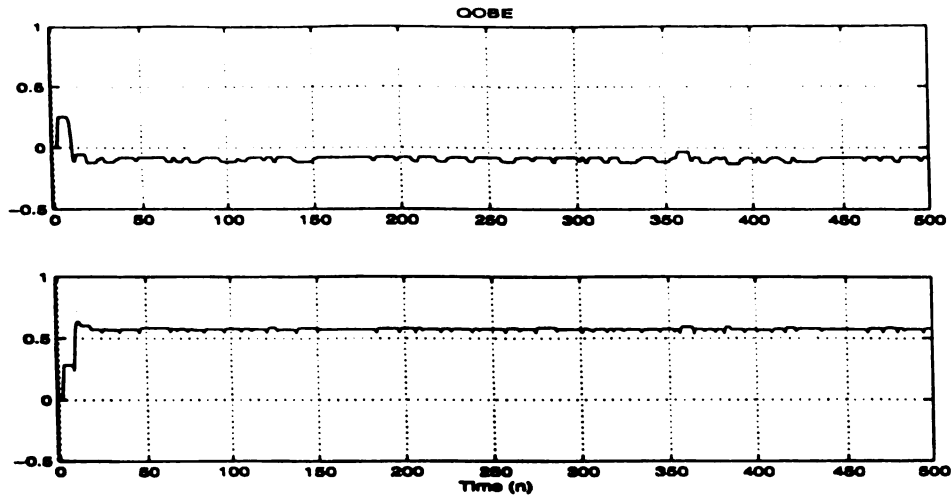


Figure 3.26: Case 8: Estimators of $a_{1*} = -0.1$ and $a_{2*} = 0.56$ convergence using QOBE for identification of model (3.4) with $\gamma_n = 0.96$, where 974 points were used from a total of 3000 points.

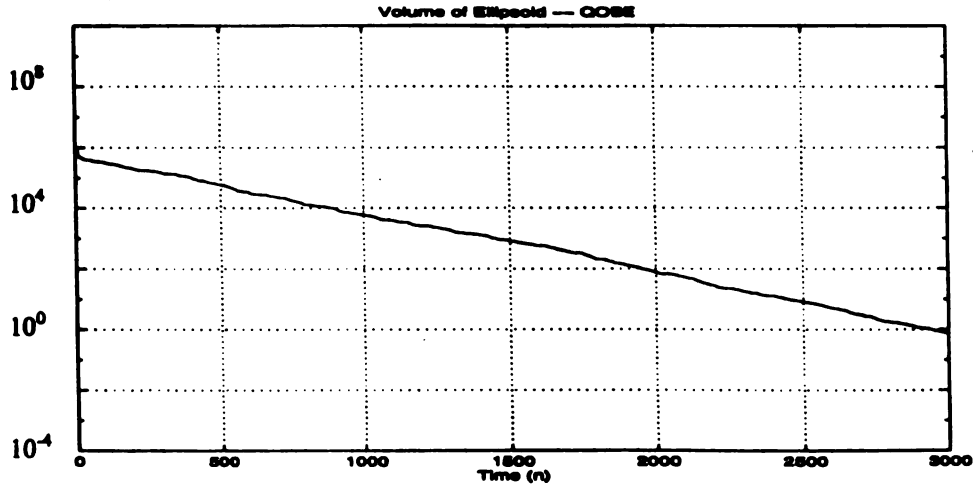


Figure 3.27: Case 8: Volume of the ellipsoid in QOBE algorithm for identification of model (3.4) with $\gamma_n = 0.96$, where 974 points were used from a total of 3000 points.

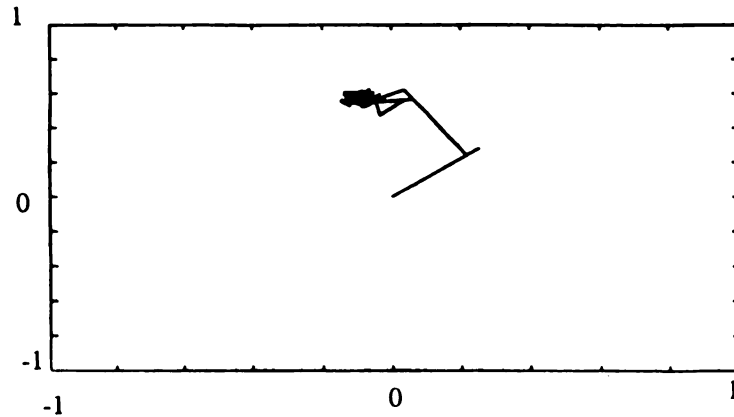


Figure 3.28: Case 8: The trajectory of the estimators in model (3.4) with $\gamma_n = 0.96$, where 974 points were used from a total of 3000 points.

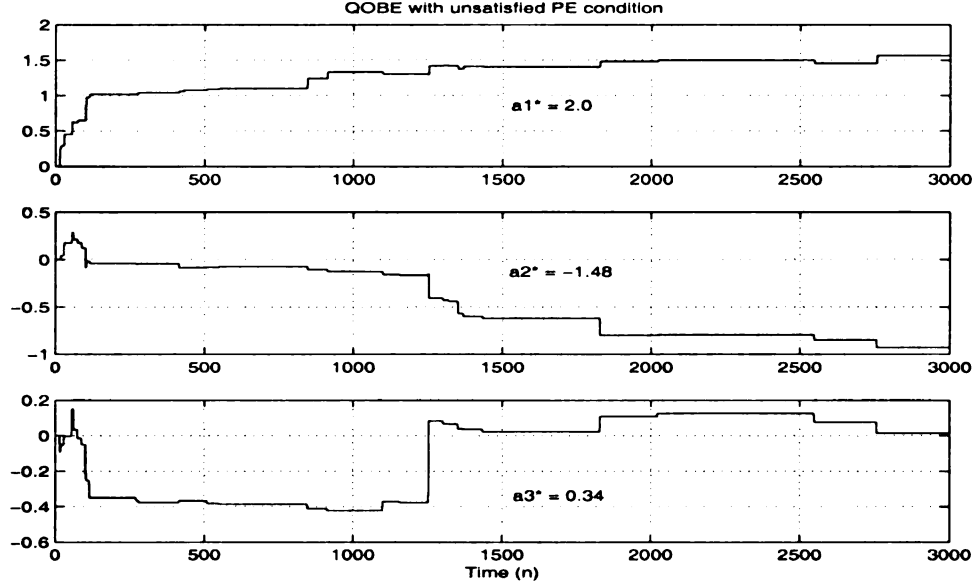


Figure 3.29: Estimators of $a_{1*} = 2.0$, $a_{2*} = -1.48$ and $a_{3*} = 0.34$ using QOBE for identification of model (3.1) when the PE condition is not satisfied.

3.3 Effect of Unsatisfied Conditions

In this section we want to show the effect of missing conditions or unsatisfied constraints from Theorems 2.1 - 2.3 on the performance of QOBE. This will enhance the illustration of the validity of the proof of these theorems. It should be noted that in all of the following simulations, the same results would be obtained even we increased the number of points higher than 3000 points indefinitely.

Figure 3.29 shows the effect of losing the PE condition on the performance of QOBE. It is seen that when the observation sequence is correlated then the algorithm will tend to converge to away from the true parameters.

Figure 3.30 shows the effect of missing the UCT condition on the performance of QOBE. In this case the noise or error sequence will not be uniformly conditionally tailed around the neighborhood of the bounds and this effect is seen the algorithm drifting away from the true parameters.

Figure 3.31 shows the performance of QOBE when both the PE and the UCT are not satisfied. We notice that the algorithm in this case will not converge properly and

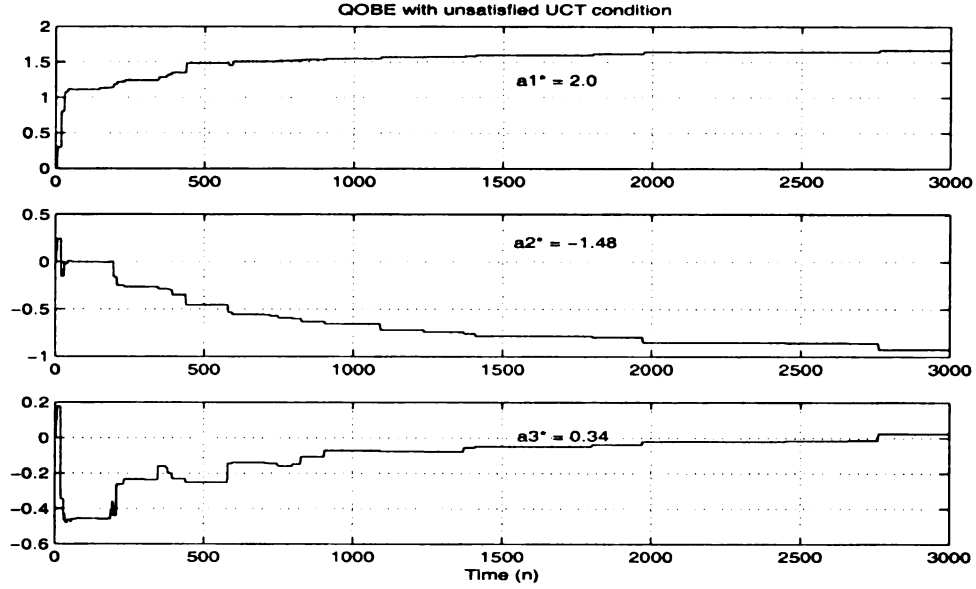


Figure 3.30: Estimators of $a_{1*} = 2.0$, $a_{2*} = -1.48$ and $a_{3*} = 0.34$ using QOBE for identification of model (3.1) when the UCT condition is not satisfied.

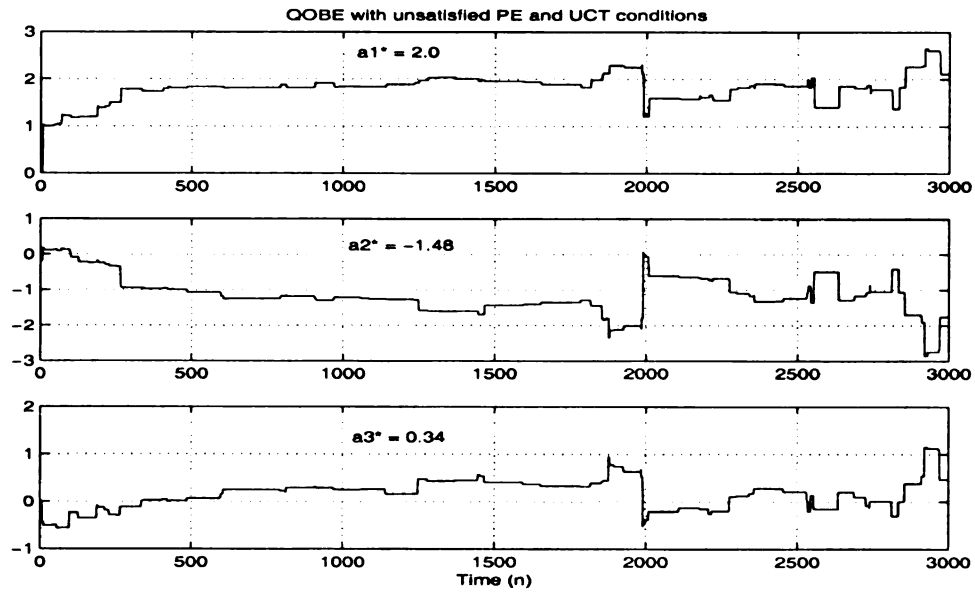


Figure 3.31: Estimators of $a_{1*} = 2.0$, $a_{2*} = -1.48$ and $a_{3*} = 0.34$ using QOBE for identification of model (3.1) when both the PE and the UCT conditions are not satisfied.

it will bounce in an undefined manner due to uncorrectly meeting the noise bounds and also having a correlation factor in the observation sequence.

Figures 3.29, 3.30 and 3.31 all show that meeting the constraints of Theorems 2.1 to 2.3 are required to ensure the convergence of the QOBE central estimator.

3.4 Tracking Performance of QOBE

In this section we will illustrate the performance of the QOBE algorithm by studying two time varying cases where the system exhibits time varying parameters. This will help show the efficiency of the tracking performance of the algorithm to time varying settings. We will consider an AR(2) model with time varying parameters that is constructed as follows:

$$y_n = a_{1*} y(n-1) + a_{2*} y(n-2) + \varepsilon_n^* \quad (3.8)$$

in which a_{1*} and a_{2*} will have the following different variations:

Case 9: AR(2) as in (3.8) where $\varepsilon_n^* \sim U(-1, 1)$ is an i.i.d. sequence with zero mean and is uniformly distributed on $[-1, 1]$, a_{1*} is a slow square wave that varies between 1.6 and -1.6 every 1000 points and $a_{2*} = 0.68$.

Case 10: AR(2) as in (3.8) where $\varepsilon_n^* \sim U(-1, 1)$ is an i.i.d. sequence with zero mean and is uniformly distributed on $[-1, 1]$, a_{1*} is a fast square wave that varies between 1.6 and -1.6 every 500 points and $a_{2*} = 0.68$.

Figures 3.32 - 3.33 illustrate the simulations for Case 9 and they show the good tracking capability of QOBE over OBE in slowly time varying systems where the time variation is abrupt. However, for Case 10, and from Figures 3.34 - 3.35, it is seen that OBE loses tracking capability while QOBE performs quite well in a system where the abrupt variation has doubled.

Case 11: AR(2) as in (3.8) where $\varepsilon_n^* \sim U(-1, 1)$ is an i.i.d. sequence with zero

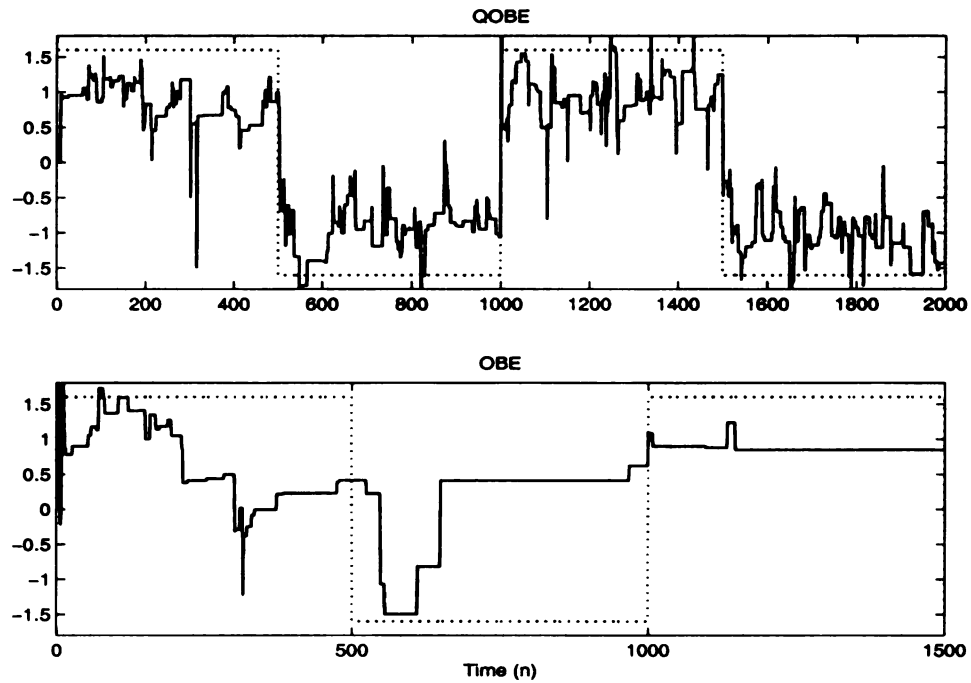


Figure 3.32: Case 9: Estimators of a_{1*} (time-varying parameter with slow abrupt change) convergence using QOBE and OBE algorithms for identification of model (3.8), $\gamma_n = 1.1$.

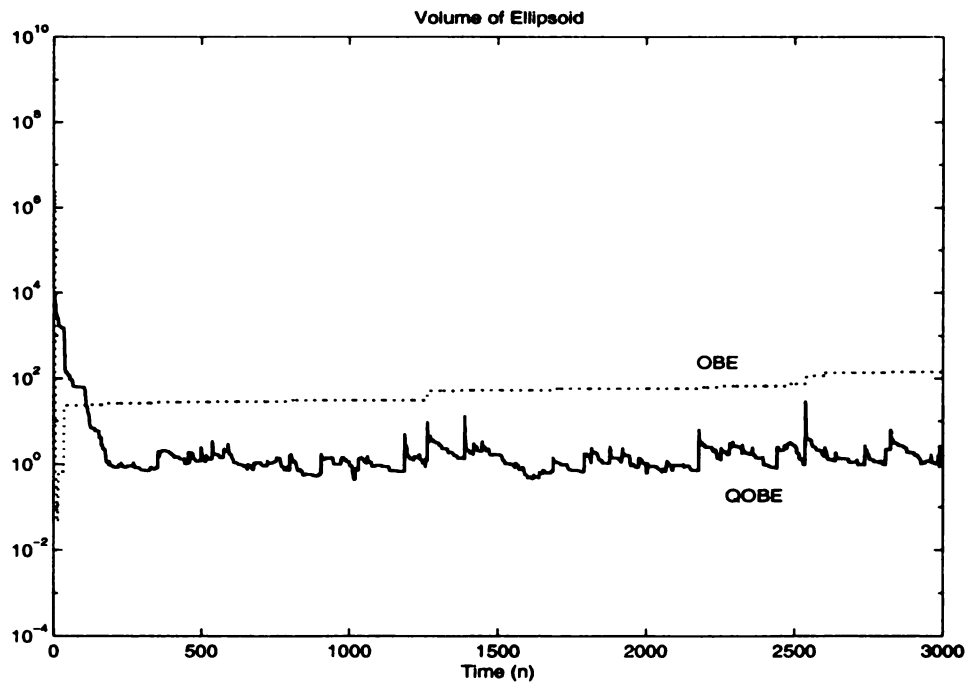


Figure 3.33: Case 9: Volume of the ellipsoids in QOBE and OBE algorithms for identification of model (3.8), $\gamma_n = 1.1$, and a_{1*} is time-varying parameter with slow abrupt change.

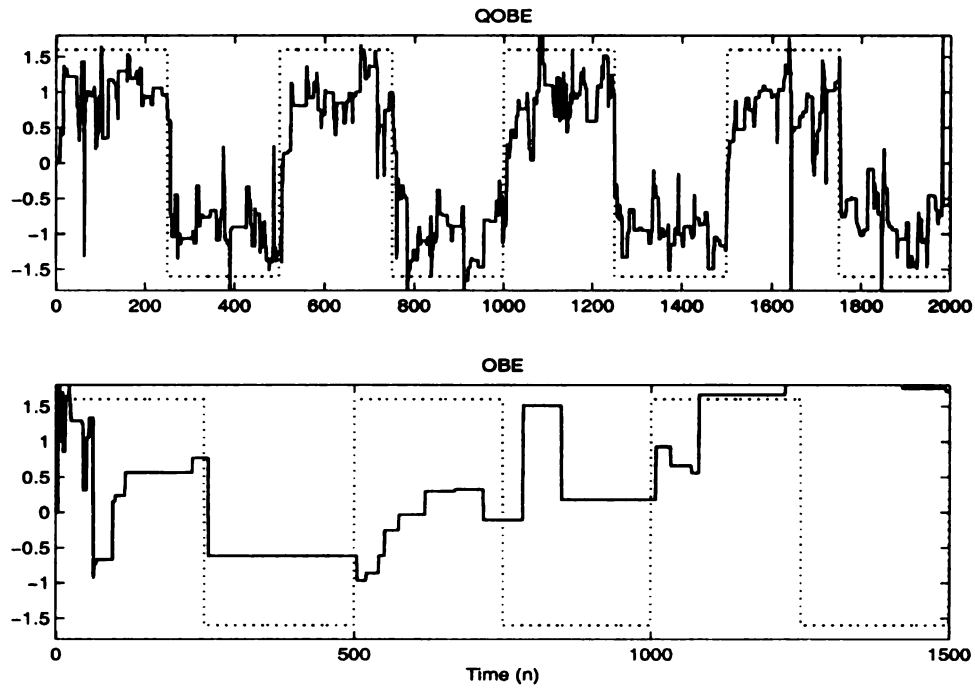


Figure 3.34: Case 10: Estimators of a_{1*} (time-varying parameter with fast abrupt change) convergence using QOBE and OBE algorithms for identification of model (3.8), $\gamma_n = 1.1$.

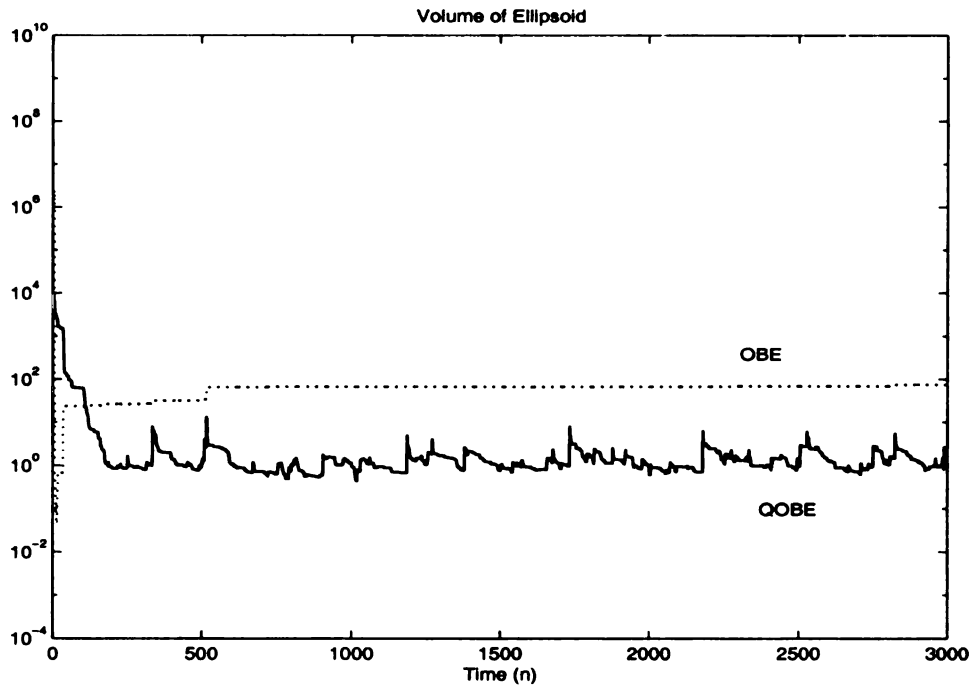


Figure 3.35: Case 10: Volume of the ellipsoids in QOBE and OBE algorithms for identification of model (3.8), $\gamma_n = 1.1$, and a_{1*} is time-varying parameter with fast abrupt change.

mean and is uniformly distributed on $[-1,1]$, a_{1*} is a slow sinusoidal wave that varies between 1.6 and -1.6, specifically, $a_{1*} = 1.6 \sin(2\pi n/1000)$, and $a_{2*} = 0.68$.

Case 12: AR(2) as in (3.8) where $\varepsilon_n^* \sim U(-1, 1)$ is an i.i.d. sequence with zero mean and is uniformly distributed on $[-1,1]$, a_{1*} is a fast sinusoidal wave that varies between 1.6 and -1.6, specifically, $a_{1*} = 1.6 \sin(2\pi n/500)$, and $a_{2*} = 0.68$.

Figures 3.36 - 3.39 represent the performance of the QOBE algorithm with respect to the OBE algorithm in a system where the time variation is gradually changing. It is seen that in slow changing variation QOBE performs better than OBE that loses track of the parameter estimate. When the time variation of the parameter a_{1*} increases by the double, QOBE still can keep track of the parameter estimate while OBE can not. It is to be noted that in all of these simulations the QOBE algorithm has been very efficient in the use of number of updates which was much less than that of the OBE case.

Cases 9 through 12 show the feasibility of using the QOBE algorithm in a time varying system and indicates its superior performance over the conventional OBE algorithm. With its less number of computations, its minimization criteria that is independent of the volume size, and its robustness to time variations, QOBE proves to be a well behaved class of the set membership algorithms.

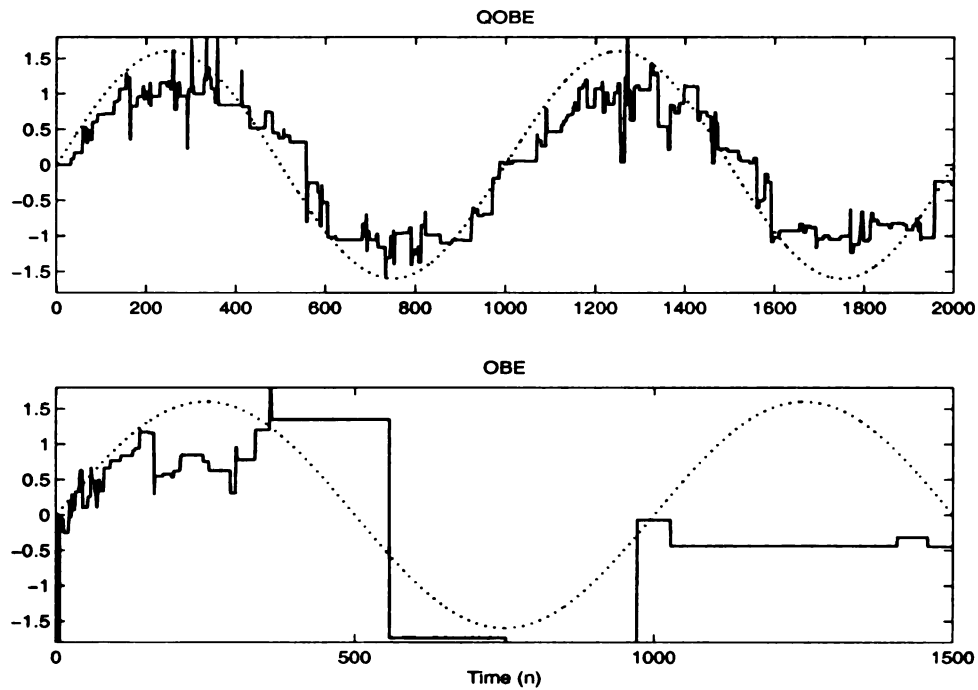


Figure 3.36: Case 11: Estimators of a_{1*} (time-varying parameter with slow gradual change) convergence using QOBE and OBE algorithms for identification of model (3.8), $\gamma_n = 1.1$.

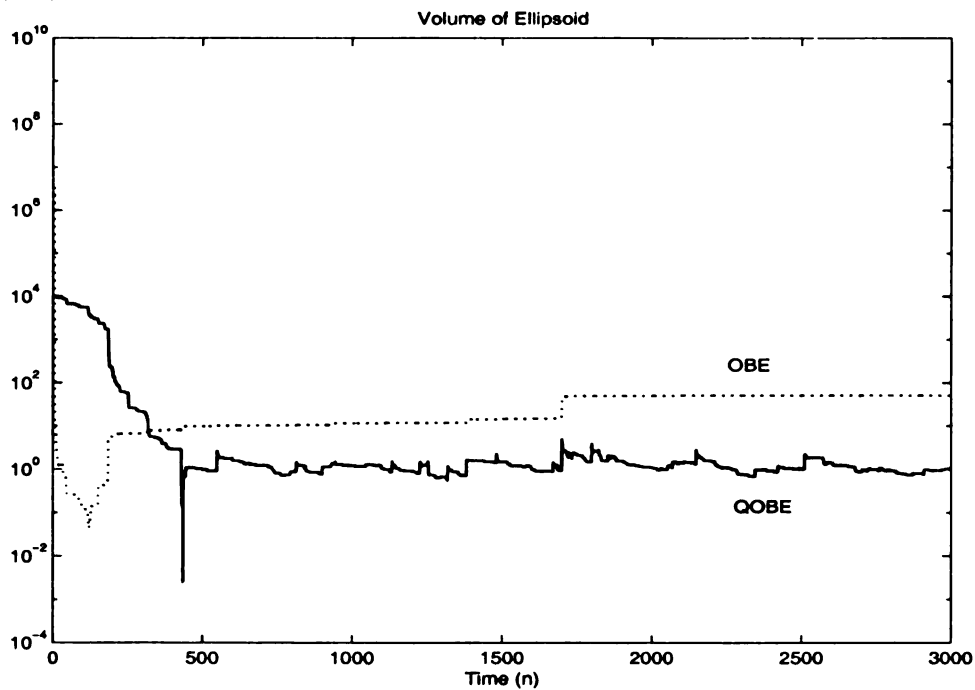


Figure 3.37: Case 11: Volume of the ellipsoids in QOBE and OBE algorithms for identification of model (3.8), $\gamma_n = 1.1$, and a_{1*} is time-varying parameter with slow gradual change.

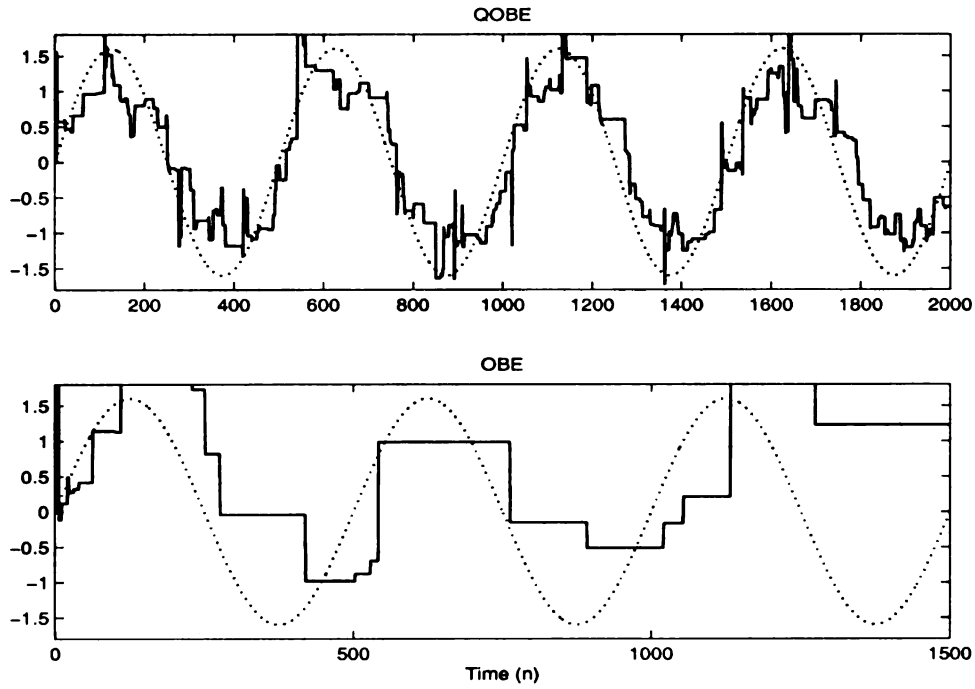


Figure 3.38: Case 12: Estimators of a_{1*} (time-varying parameter with fast gradual change) convergence using QOBE and OBE algorithms for identification of model (3.8), $\gamma_n = 1.1$.

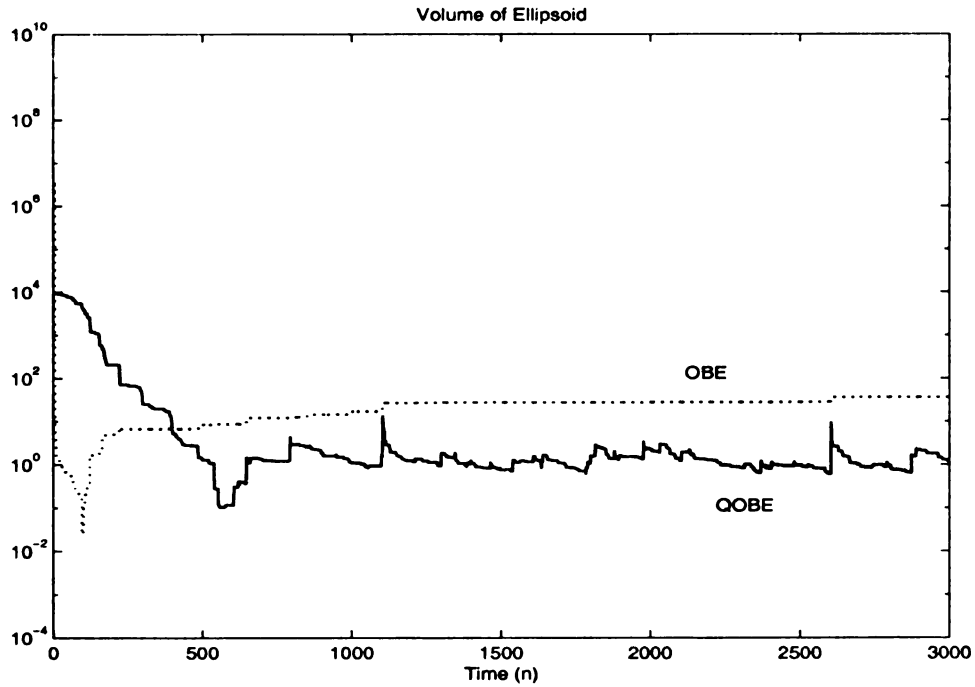


Figure 3.39: Case 12: Volume of the ellipsoids in QOBE and OBE algorithms for identification of model (3.8), $\gamma_n = 1.1$, and a_{1*} is time-varying parameter with fast gradual change.

Chapter 4

Application of QOBE in Classification

4.1 Introduction

Classification is a classic problem where the criterion is to associate each of a set of events with one of a finite number of classes [16]. An event is associated with the class to which it is “closest” according to a measure upon which each event-class pair is evaluated [54]. In this chapter we investigate the performance of the QOBE algorithm in classification problems to illustrate its application potential. This will also relate to our convergence analysis in Chapter 2 by implementing classification problems where parametric models, such as AR models, are used.

4.2 Classification Using QOBE Algorithms

Let us consider an n class problem in which class c corresponds to a model of form (1.1) with parameters $\theta_{c\star}$, so that class c is effectively represented by a point in $\mathbb{R}^m$. The classification problem will naturally employ a metric, say d , over $\mathbb{R}^m$, so that $\theta_{c\star}$ is a point in metric space $(\mathbb{R}^m, d)$. Suppose that the true, but

unknown, class is represented by vector θ_* [or some mapping, $F(\theta_*)$]. Suppose that a QOBE algorithm is used to classify a signal frame over some time frame ending at t . The analysis produces the hyperellipsoidal set Ω_t including, in particular, the central estimate θ_t . For reference, let us assume that we have the conventional unweighed least-square error (LSE) estimate of θ_* , say $\theta_{t,LSE}$. Regardless, of the estimator (or set of estimators), the assignment rule for the current signal frame is typically of the form

$$\text{Choose class } c_* \text{ , where } c_* = \arg \min_{c \in [1,n]} \delta \left(\hat{\theta}_t, \theta_{c*} \right) \quad (4.1)$$

and where $\hat{\theta}_t$ is the estimator derived for the frame ($\theta_{t,LSE}$, θ_t , or Ω_t) and δ is a measure of distance. To illustrate the use of these constructions, we develop the following example where we seek to classify the output of the two AR(2) systems of the form (1.1) with

$$\theta_{1*} = [-1.38, -0.64]^T \quad (4.2)$$

$$\theta_{2*} = [0.1, -0.56]^T \quad (4.3)$$

A sequence of length 3000 from each system is generated and the resulting signals are segmented into 10 frames of 300 points each. Each of the 20 frames is classified three times, once using $\theta_{t,LSE}$, θ_t , or Ω_t . The classification rule is of form (4.1) with δ (henceforth “ d ” in this example) indicating the Euclidean distance between the class parameters and relevant estimate. The criteria of $d(\Omega_t, \theta_{c*})$ is such that at each t , 50 quasi-random estimates are chosen from Ω_t by selecting random directions from the center of the ellipsoid (each angle being uniformly distributed over the interval $[0, 2\pi]$) and random magnitudes in the chosen direction (the length being uniformly distributed between 0 and the maximum length in that direction). If the 50 selected

points comprise the set $\Theta_t = \{\theta_{t,i} \mid i = 1, 2, \dots, 50\} \subset \Omega_t$, then

$$d(\Omega_t, \theta_{c*}) \stackrel{\text{def}}{=} \min_{i \in [1, 50]} d(\theta_{t,i}, \theta_{c*}). \quad (4.4)$$

In other words, $d(\Omega_t, \theta_{c*})$ implies a nearest-neighbor rule in the sense of Euclidean distance between θ_{c*} and Θ_t . The estimator so chosen is called simply the “other” estimate in the following. The three scatter plots, one for each estimate are shown in Figures 4.1 - 4.3. The system identity for a given point is indicated by the symbol used for plotting. Each point in each diagram is classified according to the relevant distance metric. We notice the lessened variance associated with the clusters resulting from the “other” estimator. Of course, this effect would be entirely expected since an exhausted search of Ω_t were performed for the minimum distance classifier. Although this could be an *ad hoc* approach, it shows evidence of significant improvement in classification.

An important feature of QOBE is that it does not explicitly depend upon the bounding ellipsoid, therefore the need to assure that θ_* remains inside the ellipsoid, which is critical in OBE processing, is not important in QOBE. Also, the ellipsoidal set associated with QOBE has interesting convergence behavior which are depicted in Figures 4.1 - 4.3 and can benefit QOBE tracking behavior.

One way to validate the usage of “other” estimates, can be by illustrating their effect on the magnitude spectrum of those estimates along the major and minor axes of the final ellipsoid with comparison to that of the true parameter. Figure 4.4 shows that those estimates along the minor axis corresponding to data points in the direction of error have a comparable energy to the true parameter than those along the major axis where the data points were not in the direction of error and therefore did not contribute much information to the algorithm update. Therefore, the potential of using these “other” estimates is promising for application in many signal processing

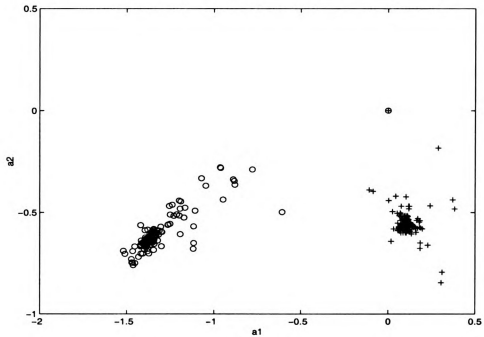


Figure 4.1: QOBE central estimator θ_t of θ_{1*} and θ_{2*} of two AR(2) systems represented by '+' and 'o', respectively.

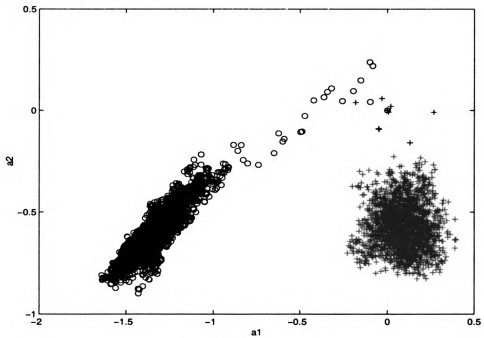


Figure 4.2: $\theta_{t,LSE}$ estimator of θ_{1*} and θ_{2*} of two AR(2) systems represented by '+' and 'o', respectively.

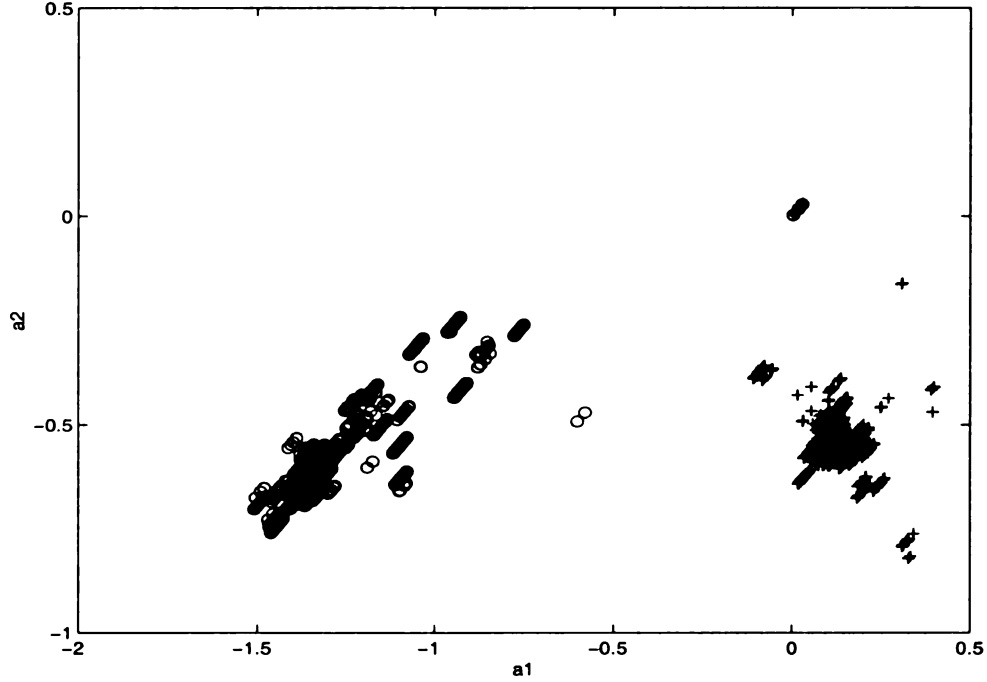


Figure 4.3: QOBE other estimators Ω_t of θ_{1*} and θ_{2*} of two AR(2) systems represented by '+' and 'o', respectively.

realm.

To show the potential of QOBE in classification problem with higher model order, we develop the following example where we seek to classify the output of the two AR(14) systems of the form (1.1) with

$$\begin{aligned} \theta_{1*} = & [1.0574, -0.9442, 1.4281, -1.5123, 0.5309, -0.8630, 0.6631, \\ & 0.1423, -0.2438, -0.0809, -0.2310, 0.1333, 0.0272, 0.0214]^T \end{aligned} \quad (4.5)$$

$$\begin{aligned} \theta_{2*} = & [0.1578, -1.3503, 1.4395, -0.2691, 1.8091, -0.0284, -0.0307, \\ & -0.3629, -1.0732, 0.1254, -0.5244, 0.4307, -0.1156, 0.1906]^T \end{aligned} \quad (4.6)$$

In this context we will apply also the nearest-neighbor rule in which $d(\Omega_t, \theta_{c*})$ implies a minimum Euclidean distance between θ_{c*} and Θ_t . The estimator so chosen will be again called the “other” estimate in the following. Figures 4.5 - 4.7 show the three scatter plots for the first two parameters of the AR(14) model, one for each

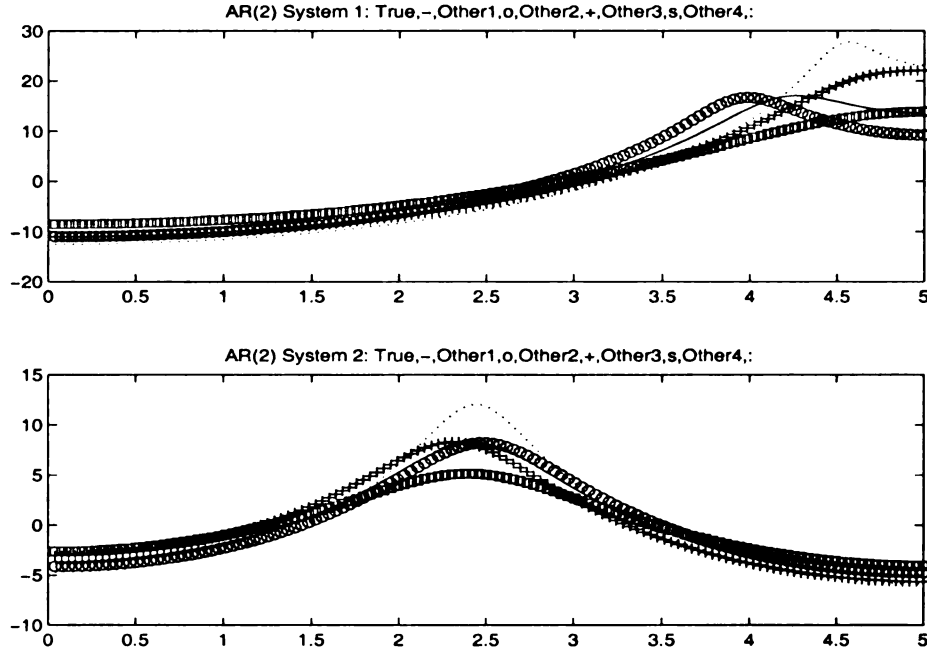


Figure 4.4: The direct 256 points FFT spectrum of the impulse response for AR(2). The upper graph for System1. The lower graph for System2. In each graph: the true parameters (-); estimators along the minor axis Other1 (o) and Other2 (+); and estimators along the major axis Other3 ($\square$) and Other4 (.).

estimation method: the QOBE central estimator, the LS estimator and the “other” estimator, respectively. Although this may be a crude way of plotting, however, it can give a sense of how the estimators are behaving in a higher order model. The system identity for a given point is indicated by the symbol used for plotting. Each point in each diagram is classified according to the relevant distance metric. We notice again the lessened variance associated with the clusters resulting from the “other” estimator.

Suppose a class C is to be classified with either members of System 1 or System 2, where class C , for example, is on the form

$$\begin{aligned} \text{class } C = & [0.6076, -1.1473, 1.4338, -0.8907, 1.17, -0.4457, 0.3162, \\ & -0.1103, -0.6585, 0.0223, -0.3777, 0.282, 0.0442, 0.1060]^T \end{aligned}$$

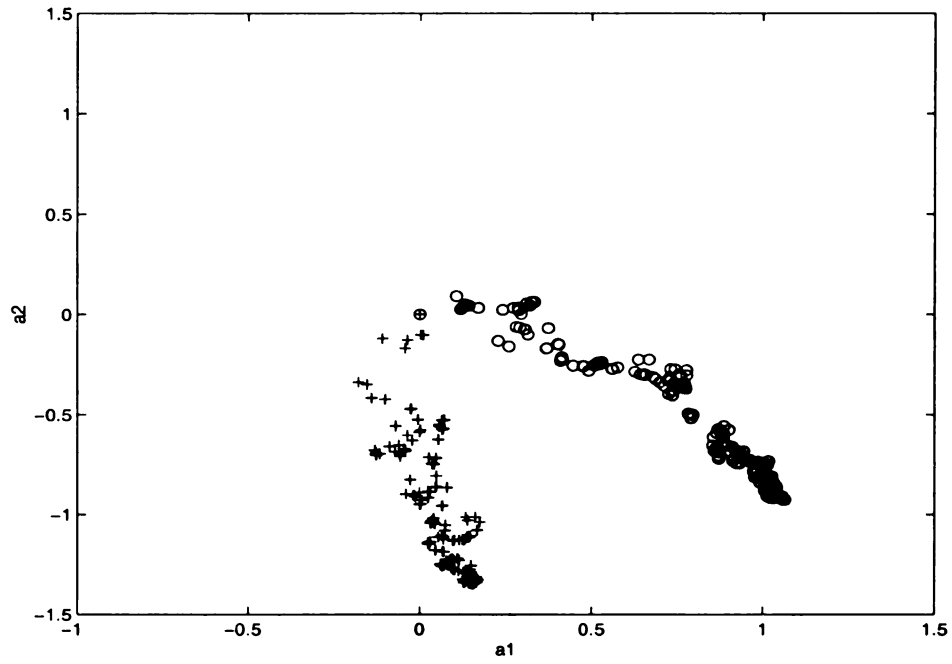


Figure 4.5: QOBE central estimator θ_t of θ_{1*} and θ_{2*} of two AR(14) systems represented by '+' and 'o', respectively.

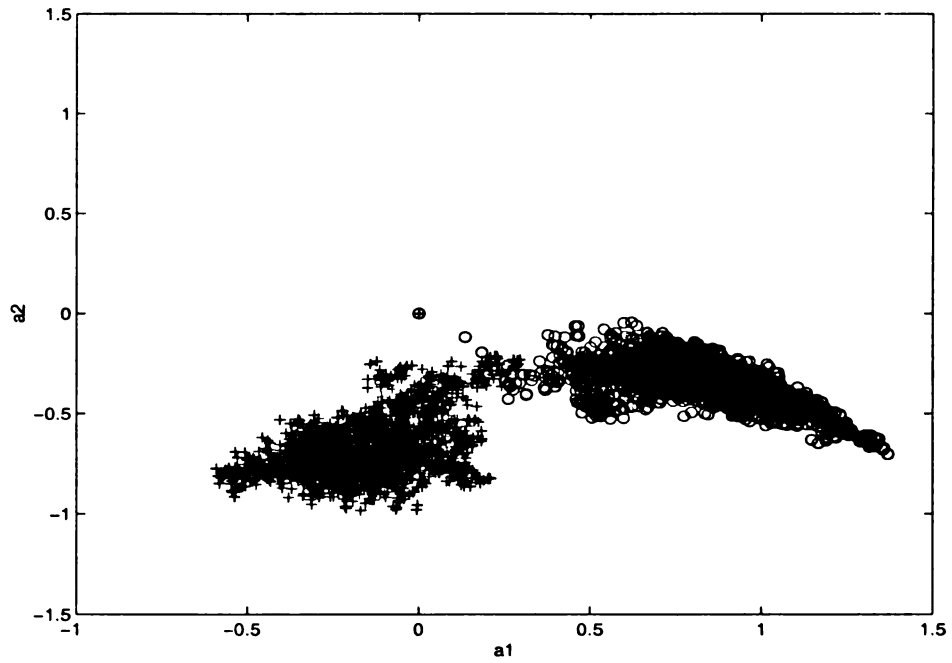


Figure 4.6: $\theta_{t,LSE}$ estimator of θ_{1*} and θ_{2*} of two AR(14) systems represented by '+' and 'o', respectively.

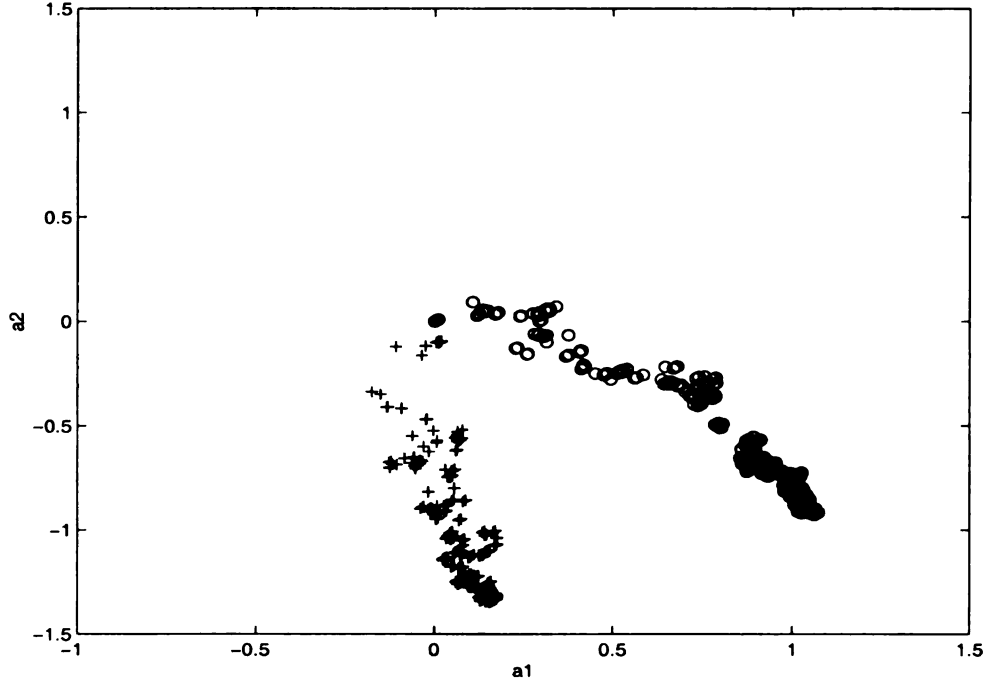


Figure 4.7: QOBE other estimators Ω_t of θ_{1*} and θ_{2*} of two AR(14) systems represented by '+' and 'o', respectively.

For experimental analysis, the class was chosen such that it is more emphasized towards System 2. Figure 4.8 shows a sketch of how the classification works. We trimmed the dimensions of the ellipsoid by throwing away those eigen vectors of the covariance matrix that do not contribute useful information to the updates and calculate the minimum distance between the class and the particular estimator through the Euclidean distance measure explained previously. The minimum distance between the class and the central estimator for both systems, “other” estimates on the minor axes (representing the maximum eigen value of the ellipsoid matrix) for both systems, and “other” estimates on the major axes (representing the minimum eigen value of the ellipsoid matrix) for both systems are calculated and tabulated in Tables 4.1 - 4.4.

Table 4.1 shows the minimum distances when using the central estimators and the “other” estimates on the minor axis for both systems. It can be seen that both the central estimator and the “other” estimator for System 2 are having small distances, with the “other” estimator of System 2 giving the minimum distance of all. Class

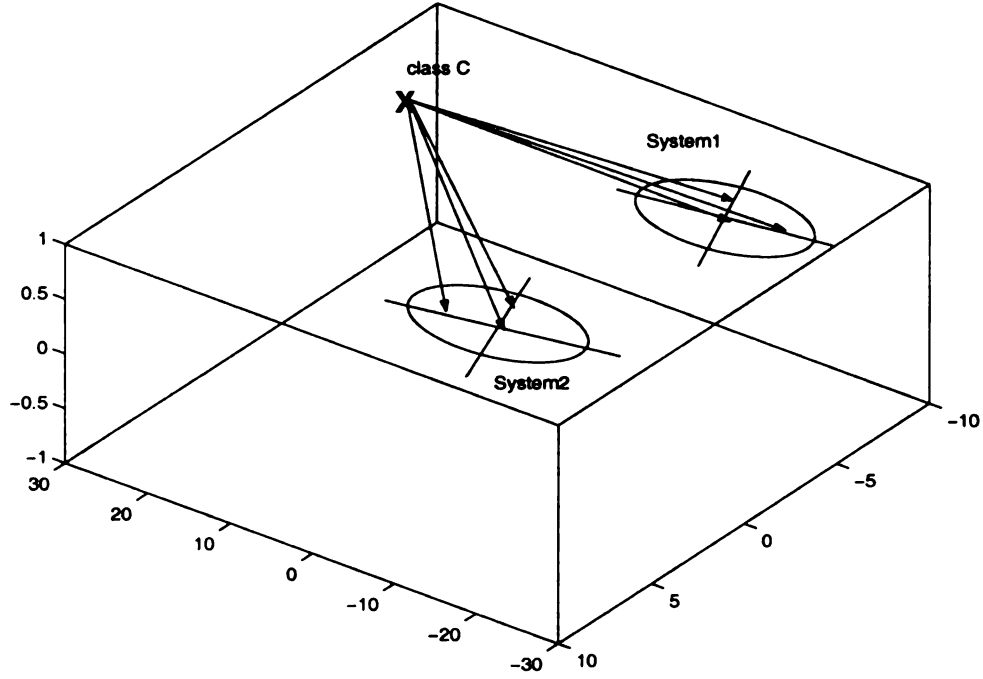


Figure 4.8: Final ellipsoids of System 1 and System 2 used for associating class C to its membership set

$d_{\min}(\theta_{center, System1})$	0.0057
$d_{\min}(\theta_{other, System1})$	0.0052
$d_{\min}(\theta_{center, System2})$	7.6904×10^{-4}
$d_{\min}(\theta_{other, System2})$	7.0183×10^{-4}

Table 4.1: The minimum distance between class C and estimates of System 1 and System 2 along the minor axis

$d_{\min}(\theta_{center, System1})$	0.0019
$d_{\min}(\theta_{other, System1})$	3.5695×10^{-5}
$d_{\min}(\theta_{center, System2})$	3.607×10^{-4}
$d_{\min}(\theta_{other, System2})$	6.085×10^{-5}

Table 4.2: The minimum distance between class C and estimates of System 1 and System 2 along the major axis

$d_{\min}(\theta_{center, System1})$	4.49×10^{-5}
$d_{\min}(\theta_{other, System1})$	2.9×10^{-5}
$d_{\min}(\theta_{center, System2})$	1.21×10^{-5}
$d_{\min}(\theta_{other, System2})$	1.4199×10^{-7}

Table 4.3: The minimum distance between class C and estimates of System 1 and System 2 after rotating the major axis of System 2 by 90 degrees

$d_{\min}(\theta_{center, System1})$	0.0025
$d_{\min}(\theta_{other, System1})$	5.32×10^{-5}
$d_{\min}(\theta_{center, System2})$	0.0016
$d_{\min}(\theta_{other, System2})$	0.0012

Table 4.4: The minimum distance between class C and estimates of System 1 and System 2 after rotating the major axis of System 1 by 90 degrees

C will be correctly classified to System 2 and this can be attributed to the “other” estimate since it is on the minor axis where the data is mostly contributing information to update the system parameters.

Table 4.2 shows the minimum distances when using the central estimators and the “other” estimates on the major axis for both systems. It can be seen that the central estimators of both systems having small distances, with the central estimator of System 1 giving the minimum distance of all. This renders faulty classification of class C to System 1 and this is because the “other” estimates are on the major axis where the data does not give either enough or any information useful to update the system parameters.

Table 4.3 shows the minimum distances when using the central estimators and the “other” estimates after rotating the major axis of System 2 by 90 degrees. It can be seen that the “other” estimator of System 2 gives the minimum distance of all. This renders correct classification of class C to System 2 and indicates that the “other” estimates in this situation was much more helpful than the central estimator of the same System 2 in correctly classifying the class C.

Table 4.4 shows the minimum distances when using the central estimators and the “other” estimates after rotating the major axis of System 1 by 90 degrees. Here the “other” estimator of System 1 gives the minimum distance of all. Again this is attributed that the “other” estimator of System 2 was on the major axis and therefore did not contribute correct information. Therefore classification of class C will be to System 1, giving incorrect result. The above experiment illustrates that the “other”

estimates contribute a lot of useful information to the final classification result as long as their selection or location inside the ellipsoid is chosen carefully.

Chapter 5

Conclusion and Future Work

5.1 Conclusions

The research in this dissertation has been concerned with a specific class of SM algorithms known as the *quasi optimal bounding ellipsoid* (QOBE) algorithm. QOBE algorithm is based on a different minimization criteria than other OBE algorithms. QOBE algorithm's convergence properties have been analyzed and proven in a deterministic framework in the literature. In this work we have analyzed the convergence properties of the QOBE algorithm in a stochastic framework. This is an important issue since it will provide an understanding of the algorithm's performance and its implementation in many of practical real world applications. We have provided the conditions and constraints required to implement the algorithm in any stochastic settings through proven theorems and their supporting corollaries and lemmas. Through many simulations in Chapter 3 and 4 we have showed the performance of the algorithm in different noise conditions ranging from i.i.d. uniformly distributed noise to asymmetrical and colored noise sequences as well. The QOBE central estimator converges in all cases under the right conditions given in the supportive theorems in Chapter 2. However, we have shown that the membership set of the QOBE al-

gorithm can not converge to a point estimate and the results although expected are now proven. Finally we have introduced the potential of using the QOBE algorithm in classification problems. The “other” non-central estimators were shown to have promising use in such classification applications.

5.2 Contributions and Future Work

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

1. Introduced the importance of researching the stochastic analysis of the QOBE algorithm.
2. Proved the a.s. convergence and p. convergence of the QOBE algorithm central estimator in i.i.d., mixing, ergodic, and non-stationary noise cases by applying necessary analytical techniques and methods from probability and measure theory. Theorems in Chapter 2 provide the necessary theoretical foundation for the analysis and application of the QOBE algorithm in stochastic settings.
3. Proved that the membership set of the QOBE ellipsoid can not converge to a point estimate.
4. Provided extensive simulations that show the behavior and characteristics of the QOBE algorithms in the different noise cases and show its performance in this conditions.
5. Provided several simulations to show the feasibility of using other non-central estimators within the QOBE ellipsoid for classification problems.

Many interesting points of research can be followed in the footsteps of this work such as:

1. Exploring the effect of unsatisfied UCT (2.6) and UT (2.7) conditions on the performance of the QOBE algorithm.
2. Investigating the application of the QOBE algorithm in digital communication problems such as inter-symbol interference where output noise will be added to the system.
3. Exploring the potential use of the other non-central estimators in other digital signal processing applications with more emphasis on the conditions for selecting those estimates .

APPENDIX

Appendix

Lemma A.1 [13] *If $\{x_k : k \in \mathcal{K}\}$ is an ergodic stationary sequence, and $g : \mathbb{R}^k \rightarrow \mathbb{R}$ is measurable then $\{y_k : g(x_k, x_{k-1}, \dots) : k \in \mathcal{K}\}$ is also stationary and ergodic.*

Lemma A.2 [13] *If both $\{x_k : k \in \mathcal{K}\}$ and $\{y_k : k \in \mathcal{K}\}$ are ergodic, stationary sequences and $\{x_k : k \in \mathcal{K}\}$ is independent of $\{y_k : k \in \mathcal{K}\}$, then $\{z_k = x_k + y_k : k \in \mathcal{K}\}$ is also stationary and ergodic.*

Theorem A.1 [13] *Let $T : \Omega \rightarrow \Omega$ be a measure preserving transformation (m.p.t) and $A \in \mathfrak{S}$. Define $\tau_A = \inf \{n \geq 1 : T^n(w) \in A\}$ where $w = \{\dots, w_{-1}, w_0, w_1, \dots\} \in \Omega$. Then*

1. $\tau_A < \infty$ a.s. on A , i.e., $P(w \in A : \tau_A(w) = \infty) = 0$
2. $A \subset \{w : T^n(w) \in A \text{ i.o.}\}$
3. If T is ergodic and $P(A) > 0$ then $P(w : T^n(w) \in A \text{ i.o.}) = 1$

BIBLIOGRAPHY

Bibliography

- [1] M. H. Hayes. *Statistical Digital Signal Processing and Modeling*. John Wiley and Sons, Inc., New York, first edition, 1996.
- [2] T. Soderstrom and P. Stoica. *System Identification*. Prentice-Hall, Englewood-Cliffs, New Jersey, 1989.
- [3] M. Milanese and G. Belforte. Estimation theory and uncertainty intervals evaluation in presence of unknown but bounded errors: Linear families of models and estimators. *IEEE Trans. on Automatic Control*, vol. AC-27:408–414, April 1982.
- [4] S. Dasgupta and Y. F. Huang. Asymptotically convergent modified recursive least-squares with data-dependent updating and forgetting factor for systems with bounded noise. *IEEE Trans. on Information Theory*, vol. IT-33:383–392, May 1987.
- [5] G. Belforte, B. Bona, and S. Frediani. Optimal sampling schedule for parameter estimation of linear models with unknown but bounded measurement errors. *IEEE Trans. on Automatic Control*, vol. AC-32:179–182, February 1987.
- [6] J. P. Norton. Identification and application of bounded-parameter models. *Automatica*, vol. 23:497–507, 1987.
- [7] J. R. Deller, Jr. Set membership identification in digital signal processing. *IEEE ASSP Magazine*, vol.6:4–20, Oct. 1989.
- [8] A. K. Rao, Y. F. Huang, and S. Dasgupta. ARMA parameter estimation using a novel recursive estimation algorithm with selective updating. *IEEE Trans. Acoust., Speech, and Signal Process.*, vol. 38:447–457, March 1990.
- [9] J. R. Deller, Jr., M. Nayeri, and S. F. Odeh. Least-square identification with error bounds for real-time signal processing and control. *Proceedings of the IEEE*, vol. 81:815–849, June 1993.
- [10] J. R. Deller, Jr., M. Nayeri, and M. S. Liu. Unifying the landmark developments in optimal bounding ellipsoid identification. *Int. J. Adaptive Control and Signal Process.*, vol. 8:43–60, 1994.
- [11] M. Nayeri, M. S. Liu, and J. R. Deller, Jr. Do interpretable optimal bounding ellipsoid algorithms converge? Part I - The long-awaited set convergence proof.

- In *Proc. 10th IFAC Symp. On System Identification*, volume vol. 3, pages 389–394, Copenhagen, Denmark, July 1994.
- [12] M. S. Liu, M. Nayeri, and J. R. Deller, Jr. Do interpretable optimal bounding ellipsoid algorithms converge? Part II - OBE vs. RLS: Clearing the smoke. In *Proc. 10th IFAC Symp. On Systems Identification*, pages 395–400, Copenhagen, Denmark, July 1994.
 - [13] T. M. Lin. *Optimal bounding ellipsoid algorithms with automatic bound estimation*. PhD thesis, Michigan State University, East Lansing, 1996.
 - [14] M. Nayeri, J. R. Deller, Jr., and M. S. Liu. Stochastic convergence of optimal bounding ellipsoid algorithms. *J. Circuits, Systems and Computers*, vol. 7:607–639, 1997.
 - [15] T. M. Lin, M. Nayeri, and J. R. Deller, Jr. A consistently convergent OBE algorithm with automatic estimation of error bounds. *Int. J. Adapt. Control and Signal Process.*, vol. 12:305–324, 1998.
 - [16] D. Joachim, J. R. Deller, Jr., and G. I. Mandour. OBE set estimates in classification problems. In *Proc. 41th Midwest Symp. Circuits and Sys.*, pages 259–262, U. Notre Dame, August 1998.
 - [17] P. Stoica. Prediction of autoregressive lognormal processes. *IEEE Trans. on Automatic Control*, vol. AC-25:292–293, April 1980.
 - [18] L. Ljung and T. Soderstrom. *Theory and Practice of Recursive Identification*. MIT Press, Cambridge, Mass., 1983.
 - [19] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The John Hopkins University Press, Baltimore, Maryland, 1983.
 - [20] D. Graupe. *Time Series Analysis, Identification, and Adaptive Systems*. Krieger, Malabar, Florida, 1989.
 - [21] S. Haykin. *Adaptive Filter Theory*. Prentice-Hall, Englewood-Cliffs, New Jersey, 1995.
 - [22] D. G. Childers, J. C. Principe, and Y. T. Ting. Adaptive WRLS-VFF for speech analysis. *IEEE Trans. on Speech and Audio Processing*, vol. 3:209–213, May 1995.
 - [23] F. C. Schweppe. Recursive state estimation: unknown but bounded errors and system inputs. *IEEE Trans. Automatic Control*, AC-13:22–28, Feb. 1968.
 - [24] D. P. Bertsekas and I. B. Rhodes. Recursive state estimation for a set-membership description of uncertainty. *IEEE Trans. on Automatic Control*, vol. AC-16:117–128, April 1971.

- [25] E. Fogel. System identification via membership set constraints with energy constrained noise. *IEEE Trans. on Automatic Control*, vol. AC-24:752–758, October 1979.
- [26] R. R. Bitmead. Persistence of excitation conditions and the convergence of adaptive schemes. *IEEE Trans. on Information Theory*, vol. IT-30:183–191, March 1984.
- [27] G. Belforte, B. Bona, and V. Cerone. Parameter estimation algorithms for a set-membership description of uncertainty. *Automatica*, vol. 26:887–898, 1990.
- [28] M. Milanese and A. Vicino. Estimation theory for dynamic systems with unknown but bounded uncertainty: An overview. *Automatica*, vol. 27:997–1009, 1991.
- [29] S. M. Veres and J. P. Norton. Structure selection for bounded-parameter models: Consistency conditions and selection criterion. *IEEE Trans. on Automatic Control*, vol. 36:474–481, April 1991.
- [30] Y. F. Huang and J. R. Deller, Jr. On the tracking capabilities of optimal bounding ellipsoid algorithms. In *The 30th Annual Allerton Conference on Communications, Control and Computing*, Urbana, IL, September 30 - October 2 1992.
- [31] A. K. Rao and Y. F. Huang. Tracking characteristics of an obe parameter estimation algorithm. *IEEE Trans. on Signal Processing*, vol. 41, March 1993.
- [32] D. Joachim, J. R. Deller, Jr., and M. Nayeri. Multiweight optimization in OBE algorithms for improved tracking and adaptive identification. In *Proc. Int. Conf. On Acoustics, Speech, and Signal Processing*, pages 2201–2204, Seattle, WA, May 1998.
- [33] S. J. Pushparajah and J. A. Chambers. Frequency domain set membership fast normalized lms adaptive algorithm. *Electronics Letters*, vol. 35(No. 11):865–867, May 1999.
- [34] A. Garulli, A. Vicino, and G. Zappa. Conditional central algorithms for worst case set-membership identification and filtering. *IEEE Trans. on Automatic Control*, vol. 45(No. 1):14–23, January 2000.
- [35] D. Joachim and J. R. Deller, Jr. Adaptive optimal bounded-ellipsoid identification with an error under-bounding safeguard: Application in state estimation and speech processing. In *Proc. Int. Conf. On Acoustics, Speech, and Signal Processing*, pages 372–375, Istanbul, Turkey, June 2000.
- [36] D. Joachim and J. R. Deller, Jr. Some signal processing applications of set solutions. In *Proc. 12th IFAC/IFORS Sym. System Identification*, Santa Barbara, 2000.

- [37] C. Novara and M. Milanese. Set membership identification of nonlinear systems. In *Proc. 39th IEEE Conf. On Decision and Control*, Sydney, Australia, December 2000.
- [38] S. Werner and P. S. R. Diniz. Set-membership affine projection algorithm. *IEEE Signal Processing Letters*, vol. 8(No. 8):231–235, August 2001.
- [39] S. Gollamudi, S. Nagaraj, S. Kapoor, and Y. F. Huang. Set-membership state estimation with optimal bounding ellipsoids. In *Int. Symp. On Information Theory and its Applications*, Victoria, B. C. Canada, September 1996.
- [40] S. Nagaraj, S. Gollamudi, S. Kapoor, and Y. F. Huang. Bounded error estimation: Set-theoretic and least-squares formulations. In *Proc. Conf. Information Sciences and Systems*, John Hopkins University, Maryland, March 1997.
- [41] S. Nagaraj, S. Gollamudi, J. R. Deller, Jr., Y. F. Huang, and S. Kapoor. Performance studies of a 'quasi-OBE' algorithm for real-time signal processing. In *Proc. 40th Midwest Symp. Circuits and Systems*, pages 770–773, Sacramento, CA, August 1997.
- [42] J. R. Deller, Jr. and S. F. Odeh. Adaptive set-membership identification in $O(m)$ time for linear-in-parameters models. *IEEE Trans. on Signal Processing*, vol. 41:1906–1924, May 1993.
- [43] H. S. Witsenhausen. Sets of possible state of linear systems given perturbed observation. *IEEE Trans. on Automatic Control*, AC-13:556–568, October 1968.
- [44] E. Fogel and Y. F. Huang. On the value of information in system identification: Bounded noise case. *Automatica*, vol. 18:229–238, 1982.
- [45] S. Nagaraj, S. Gollamudi, S. Kapoor, and Y. F. Huang. An OBE estimation algorithm with enhanced tracking capabilities. *ISCAS*, 1997.
- [46] S. Gollamudi, S. Nagaraj, and Y. F. Huang. SMART: A toolbox for set-membership filtering. In *European Conference on Circuit Theory and Design*, Budapest, Hungary, August 31- September 3 1997.
- [47] S. Nagaraj and Y. F. Huang. A statistical approach to set-membership identification. In *ISIT 2001*, Washington, D. C., June 24-29 2001.
- [48] A. J. Krener. Kalman-bucy and minimax filtering. *IEEE Trans. on Automatic Control*, vol. AC-25:291–292, April 1980.
- [49] S. Nagaraj, S. Gollamudi, S. Kapoor, Y. F. Huang, and J. R. Deller, Jr. Multiuser detection based on a deterministic error specifications: Theory and low-complexity adaptive solutions. In *Proc. 31st Asilomar Conf. Signals, Systems, and Computers*, Monterey, 1997.

- [50] X.F. Sun and Y. Z. Fan. Comments on identification for systems with bounded noise. *IEEE Trans. on Automatic Control*, vol. 46(No. 5):808–809, May 2001.
- [51] Dale Joachim. *Multiweight optimization in optimal bounding ellipsoid algorithms*. PhD thesis, Michigan State University, East Lansing, 1998.
- [52] J. B. Moore. Persistence of excitation in extended least squares. *IEEE Trans. on Automatic Control*, vol. AC-28(No. 2):60–68, January 1983.
- [53] J. R. Deller, Jr., S. Gollamudi, S. Nagaraj, Y. F. Huang, and S. Kapoor. Convergence analysis of a new “quasi-OBE” algorithm for real-time signal processing. *Int. J. Adaptive Control and Signal Processing*.
- [54] D. Joachim and J. R. Deller, Jr. Short signal classification using set-membership identification with application to speech labeling. In *Proc. 43rd IEEE Midwest Symp. On Circuits and Systems*, Lansing, MI, August 8-11 2000.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 02470 0357