Performance Studies on a "Quasi-OBE" Algorithm for Real-time Signal Processing

Shirish Nagaraj, Sridhar Gollamudi, John R. Deller, Yih-Fang Huang and Samir Kapoor
Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556, USA

***Abstract*—A novel algorithm is studied for the identification of linear-in-parameters models. This technique is dubbed "Quasi-OBE" (QOBE) because it is based on the principles of optimal bounding ellipsoid (OBE) identification, but has other geometric and classic least-squares interpretations that enhance its interpretability and application potential. Convergence behavior of both the central point estimate and measures of the hyperellipsoidal membership-set will be discussed in general terms, particularly in comparison with several conventional OBE algorithms. The QOBE algorithm uses highly-selective updating and exhibits excellent tracking ability in time-varying environments.**

I. Formulation and background

Optimal bounding ellipsoid (OBE) algorithms [1]–[3], [6], [7] are becoming increasingly popular in signal processing, system identification, control, and communication problems as alternatives to conventional least-square-error (LSE) identifiers. For detailed discussion on the advantages and properties of OBE algorithms, refer [1], [2], [4].

Recently, a novel algorithm was introduced in [5] that shares most of the motivating principles and algorithmic structure with the OBE algorithms, but which has novel interpretations and operational properties that make it uniquely different. We shall refer to this algorithm as *quasi-OBE* (QOBE) to connote both the similarities and differences. The purpose of the present work is to generally examine the convergence properties of QOBE and to provide some insights into its operation in practical applications.

The OBE and QOBE algorithms are used to identify a *linear-in-parameters* model

$$y_t = \boldsymbol{\theta}_*^T \boldsymbol{x}_t + \varepsilon_{t*} \tag{1}$$

in which $\boldsymbol{\theta}_* \in \Re^m$ is the unknown "true" parameter vector to be identified; $\{\boldsymbol{x}_t\}$ is a sequence of measurable vectors of dimension m; and $\{\varepsilon_{t*}\}$ is an error sequence. These algorithms are based on the premise that $\{\varepsilon_{t*}\}$ has a point-wise energy bound that is known *a priori* ,

$$\varepsilon_{t*}^2 \leq \gamma_t^2, \quad \text{for all}\, t. \tag{2}$$

At each t, these bounds imply two hyperstrips in parameter space, say $\mathcal{H}_t^+ = \{\boldsymbol{\theta} | y_t = \boldsymbol{\theta}^T \boldsymbol{x}_t + \gamma_t\}$ and $\mathcal{H}_t^- = \{\boldsymbol{\theta} | y_t = \boldsymbol{\theta}^T \boldsymbol{x}_t - \gamma_t\}$, between which $\boldsymbol{\theta}_*$ must lie. The intersection of these strips forms a polytope in $\Re^m$, which is a subset of a hyperellipsoidal set at time t given by

$$\Omega_t \stackrel{\text{def}}{=} \left\{ \boldsymbol{\theta} | \left(\boldsymbol{\theta} - \boldsymbol{\theta}_t\right)^T \boldsymbol{C}_t \left(\boldsymbol{\theta} - \boldsymbol{\theta}_t\right) < \kappa_t \right\}. \tag{3}$$

The ellipsoid center, $\boldsymbol{\theta}_t$, and matrix $\boldsymbol{P}_t \stackrel{\text{def}}{=} \boldsymbol{C}_t^{-1}$ are computed recursively using

$$\begin{aligned} G_t &= \boldsymbol{x}_t{}^T \boldsymbol{P}_{t-1} \boldsymbol{x}_t & (4)\\ \varepsilon_t &= y_t - \boldsymbol{\theta}_{t-1}^T \boldsymbol{x}_t & (5)\\ \boldsymbol{P}_t &= \frac{1}{\alpha_t}[\boldsymbol{P}_{t-1} - \frac{\beta_t \boldsymbol{P}_{t-1} \boldsymbol{x}_t \boldsymbol{x}_t{}^T \boldsymbol{P}_{t-1}}{\alpha_t + \beta_t G_t}] & (6)\\ \boldsymbol{\theta}_t &= \boldsymbol{\theta}_{t-1} + \beta_t \boldsymbol{P}_t \boldsymbol{x}_t \varepsilon_t & (7)\\ \kappa_t &= \alpha_t \kappa_{t-1} + \beta_t \gamma_t - \frac{\alpha_t \beta_t \varepsilon_t^2}{\alpha_t + \beta_t G_t}. & (8) \end{aligned}$$

It is also useful for theoretical purposes to note that $\boldsymbol{C}_t$, the so-called *covariance matrix*, can be computed recursively as

$$\boldsymbol{C}_t = \alpha_t \boldsymbol{C}_{t-1} + \beta_t \boldsymbol{x}_t \boldsymbol{x}_t^T. \tag{9}$$

Recursions (4) – (8) comprise the basis for a general OBE or QOBE algorithm. The ellipsoid center $\boldsymbol{\theta}_t$ is used as an estimator of the parameters $\boldsymbol{\theta}_*$ at each t. The process is initialized with $\boldsymbol{\theta}_0 = \mathbf{0}$, $\kappa_0 > 0$ and $\mathbf{P_0} = \frac{1}{\mu}\boldsymbol{I}$, where μ is a small number, typically $\mu\kappa_0 = 10^{-4}$. $\{\alpha_t\}$ and $\{\beta_t\}$ are positive weighting sequences chosen according to the particular OBE algorithm employed. In almost every OBE algorithm, the weights are chosen to minimize the "size" (in some sense) of the set Ω_t at each iteration. When such optimal weights do not exist, the updating need not take place.

In the QOBE algorithm [5], the quantity κ_t is minimized at each iteration by letting $\beta_t = \lambda_t$, $\alpha_t = 1$, then seeking the optimal λ_t in light of the current measurements. QOBE is a "quasi" OBE algorithm in that it is different from the OBE algorithms, including the "SM-WRLS" OBE [2] and the "D/H-OBE" [1], in having an

This work was supported, in part, by the National Science Foundation under Grant MIP-9705173 and in part by the Center for Applied Mathematics, University of Notre Dame. J. R. Deller is on leave at the Univ. of Notre Dame from Michigan State University, East Lansing, MI.

interpretation that allows it to be divorced from the inherent ellipsoid; yet the ellipsoid is readily available in QOBE if it is useful. In fact, QOBE may be interpreted as a method which determines whether $\boldsymbol{\theta}_{t-1}$ lies between the hyperstrips $\mathcal{H}_t^+$ and $\mathcal{H}_t^-$ at time t. If so, the new data are ignored, and $\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1}$, $\boldsymbol{P}_t = \boldsymbol{P}_{t-1}$, etc. If not, $\boldsymbol{\theta}_{t-1}$ is moved to the nearest hyperstrip in the direction of $\boldsymbol{P}_t\boldsymbol{x}_t$, where it becomes $\boldsymbol{\theta}_t$. Further details are found in [5], where it has been shown that the central QOBE estimate can be derived, independent of OBE considerations, as a solution of a certain least-squares cost function constrained upon the need for the estimate to lie in between the hyperstrips. This kind of an interpretation is unique to QOBE and which is not shared by any other OBE algorithms to date.

One of the interesting features of the QOBE algorithm is its extraordinary ability to track time-varying dynamics while using very few data. Further, since QOBE does not explicitly depend upon the bounding ellipsoid, the need to assure that $\boldsymbol{\theta}_*$ remains inside the ellipsoid, which is critical in OBE processing, is not important here. Finally, the set (ellipsoid) associated with QOBE has interesting convergence behavior *in practice* which, while seemingly undesirable from a theoretical point of view, may also benefit tracking behavior. These issues are all discussed below in the assessment of QOBE performance.

II. Convergence behavior of QOBE

A. Introduction

In this section, we generally describe the convergence behavior of QOBE, particularly in comparison to its OBE cousins. We begin with some formal results. The following definitions are used throughout the reminder of the paper:

Definition 1. The sequence of regressor vectors $\{\boldsymbol{x}_t\}_{t=1}^{\infty}$ will be said to be *persistently exciting* (PE) if, for any nonzero $\boldsymbol{u} \in \Re^m$, there exist an infinite subsequence, say $\{\boldsymbol{x}_{t_k}\}_{k=1}^{\infty}$, and a number $\eta > 0$ (both dependent upon $\boldsymbol{u}$), such that $\boldsymbol{x}_{t_k}^T\boldsymbol{u} \geq \eta$ for all k.

Definition 2. The term *infinite visitation* (IV) refers to the condition in which the true disturbance sequence, $\{\varepsilon_{t*}\}$, visits arbitrarily small neighborhoods of its bounds infinitely often (i.o.). That is, IV occurs if for any $\epsilon > 0$, $|\varepsilon_{t*}| > \gamma_t - \epsilon$ i.o.

Definition 3. The term *infinite updating* (IU) refers to the condition in which the algorithm does not cease updating in finite time.

Further, while much of the the discussion of convergence below is not rigorous, the following (or similar) assumptions must be made in any formal pursuit of these issues. Generally speaking, these conditions will be taken for granted in the remainder of the paper:

Assumption 1. The regressor sequence $\{\boldsymbol{x}_t\}$ is PE as defined in Def. 1.

Assumption 2. $\{\boldsymbol{x}_t\}$ has a bounded norm: $0 < ||\boldsymbol{x}_t|| \leq \bar{\boldsymbol{x}}$ for every t.

Assumption 3. The true disturbance magnitude is bounded for all t: $\varepsilon_{t*}^2 \leq \gamma_t^2$ where $0 < \underline{\gamma} \leq \gamma_t \leq \bar{\gamma} < \infty$ for all t for some constants $\underline{\gamma}$ and $\bar{\gamma}$.

B. Fundamental results

The following results are fundamental to the understanding and ultimate proofs of QOBE convergence properties. The proofs of the following propositions are omitted due to lack of space. The first describes the elements of the optimization process.

Proposition 1 *[5]* An optimal QOBE (in the sense of minimizing κ_t) weight, $\lambda_t > 0$, exists at time t if and only if

$$|\varepsilon_t| > \gamma_t. \tag{10}$$

When condition (10) is satisfied, the optimal weight is given by

$$\lambda_t = \frac{1}{G_t}\left(\frac{|\varepsilon_t|}{\gamma_t} - 1\right), \tag{11}$$

and the reduced value of κ_t can be computed as

$$\kappa_t = \kappa_{t-1} - \frac{1}{G_t}(|\varepsilon_t| - \gamma_t)^2. \tag{12}$$

Remark: Condition (10) also indicates the existence of an optimal weight *in the sense of minimizing ellipsoid volume* at time t [2, p. 57]. It means that whenever κ_t is minimized, volume is (or could be) minimized at that time as well (see Section C.). •

Proposition 2 *[5]* The sequence of weights, $\{\lambda_t\}_{t=1}^{\infty}$, diminishes to zero. Further, the innovation sequence magnitude is asymptotically "upper bounded by γ_t" in the following sense: $\limsup_{t\to\infty}(|\varepsilon_t| - \gamma_t) \leq 0$ with equality when IU occurs.

The final result will be important in attempts to prove set convergence for QOBE. In particular, it can be used to assert that ellipsoid volume does not shrink to zero as a result of κ_t converging to zero.

Proposition 3 $\kappa_t > 0$ for any finite t, and $\lim_{t\to\infty}\kappa_t > 0$ unless (i) $|\varepsilon_{t*}| = \gamma_t$ for every t selected by QOBE, and (ii) $\boldsymbol{\theta}_*$ is on the boundary of the initial ellipsoid.

C. Convergence in QOBE and volume-minimizing OBE algorithms

In practice, the convergence behavior of QOBE is observed to be quite different from that of the volume-minimizing OBE algorithms to which it is closely related. QOBE tends to select very few points and usually cease updating rather quickly. Also of great interest is the fact that the tracking ability of QOBE appears to be superior to volume OBE in many cases.

The reason for these behaviors is concerned with the QOBE selection process which tends to discard data as soon as the central estimator becomes small. In turn, this often happens very quickly with respect to the shrinkage of the ellipsoid. The primary criteria for the selection of a point at time t by QOBE are the ability of the point to reduce estimation error (reduce $||\tilde{\boldsymbol{\theta}}_t||$), and the proximity of $|\varepsilon_{t*}|$ to the bound γ_t. This can be observed from the fact that $\varepsilon_t = \varepsilon_{t*} + \tilde{\boldsymbol{\theta}}_{t-1}^T \boldsymbol{x}_t$, hence

$$|\varepsilon_t| \leq |\varepsilon_{t*}| + |\tilde{\boldsymbol{\theta}}_{t-1}^T \boldsymbol{x}_t|, \tag{13}$$

so that when $|\tilde{\boldsymbol{\theta}}_{t-1}^T \boldsymbol{x}_t| = \epsilon$ and $|\varepsilon_{t*}| < \gamma_t - \epsilon$, no update occurs regardless of the size of the ellipsoid. Whereas the reduction of $||\tilde{\boldsymbol{\theta}}_t||$ will tend to reduce the size of the ellipsoid [since error tends to be in the direction of the "weakest" eigenvalues of $\boldsymbol{C}_t$ (largest axes of $\boldsymbol{\Phi}_t$)], the explicit concern for set shrinkage is not as strong in QOBE as in the case of the volume-minimizing OBE algorithms. In volume OBE algorithms, the information checking criterion (ICC) used to determine the potential of data to improve the estimate is given by (e.g., [2])

$$\varepsilon_t^2 > \gamma_t^2 - \frac{\kappa_{t-1} G_t}{m}. \tag{14}$$

With respect to the ICC for QOBE, $\varepsilon_t^2 > \gamma_t^2$, the "volume" ICC contains an extra term on the right side of the inequality that "lowers the threshold" for updating at t in consideration of the potential of the current data to reduce the size of the ellipsoid, independent of its central estimator. The quantity

$$\kappa_{t-1} G_t = \kappa_{t-1} \boldsymbol{x}_t^T \boldsymbol{C}_{t-1}^{-1} \boldsymbol{x}_t = \boldsymbol{x}_t^T \boldsymbol{\Phi}_{t-1}^{-1} \boldsymbol{x}_t, \tag{15}$$

where $\boldsymbol{\Phi}_t$ is the "ellipsoid matrix" $\boldsymbol{C}_t/\kappa_t$, represents a normalized instantaneous signal energy, weighted in proportion to the sizes of the axes of the existing ellipsoid. This weighting tends to emphasize signal components in the directions of "greatest need," that is, along axes where the ellipsoid is still the largest. Conversely, signal energy in directions with small axes are downplayed.

From the analysis above, it is apparent that, given identical conditions at time $t-1$, a volume OBE algorithm would update at time t if QOBE does (see remark below Proposition 1). Further, it is interesting to note that the ellipsoid volume associated with QOBE *is* decreased each time the algorithm updates. This does not imply, however, that the volume of the QOBE approaches zero, even with IU. Empirical evidence suggests that the volume, in fact, does not vanish. Proof of rigorous conditions for point-set convergence, however, remains an open problem for QOBE.

On the other hand, the QOBE central estimator $\boldsymbol{\theta}_t$ regularly approaches $\boldsymbol{\theta}_*$ in simulation studies. Here, too, the precise conditions remain unknown. It is reasonable to conjecture that IU is theoretically necessary and sufficient for estimator convergence, but this remains to be proven. In turn, an IV condition on $\{\varepsilon_{t*}\}$ is almost certainly requisite for estimator convergence as implied by the discussion below (13). Sufficiency of IV is also suggested by Proposition 2 since ε_{t*} and ε_t can only "match" (both quantities in arbitrarily small neighborhoods of $\pm\gamma_t$) infinitely often if $||\tilde{\boldsymbol{\theta}}_t|| \to 0$.

Finally we note that the fast-estimator-convergence — reluctant-set-convergence behavior that is observed for QOBE in practice may offer an explanation for its robust and excellent tracking performance. The ability to quickly seek the true parameters is obviously the essence of quick tracking performance, but the tendency for the ellipsoidal set to remain large may be an asset with respect to other OBE methods. The theory of OBE algorithms does not support the identification of time-varying dynamics. However, the maintenance of a larger ellipsoidal set in the QOBE case may fortuitously benefit tracking in this regard.

In the next section, we illustrate this theoretical discussion with simulation studies.

D. *Simulation studies*

The convergence properties of QOBE are demonstrated via simulation studies, first for the identification of a 4^{th} order FIR filter with the output corrupted by uniformly distributed noise. The chosen filter is given by the following:

$$\begin{aligned} y_t &= u_t - 0.5\ u_{t-1} + 0.25\ u_{t-2} + 0.125\ u_{t-3} + \varepsilon_{t*} \\ y_t &= \boldsymbol{\theta}_*^T \boldsymbol{x}_t + \varepsilon_{t*} \end{aligned}$$

for all t, where

$$\begin{aligned} \boldsymbol{\theta}_* &= [1 \quad -0.5 \quad 0.25 \quad 0.125]^T \\ \boldsymbol{x}_t &= [u_t \quad u_{t-1} \quad u_{t-2} \quad u_{t-3}]^T \end{aligned}$$

It is assumed that the "true" noise bound is known. In this case, the value of $\gamma_t = 0.1122$ for all t. The Signal to Noise Ratio (SNR) is fixed at 25 dB for all the simulation examples. The pertinent points to note are that the convergence rate of QOBE is similar to that of RLS, although QOBE uses only 40 updates in 1000 samples, while RLS processes all the samples. Moreover, the mean squared parameter error (averaged over 200 runs) resulting from QOBE is almost a order of magnitude smaller than that by RLS, as can be seen in Figure (1).

The prediction error can be seen to lie at a value smaller than the (time-invariant) noise bound in Figure (2), as is predicted theoretically. The convergence of the update weights to zero (cessation of updating) was also observed in the simulations conducted.

Figure (3) is a plot of the evolution of the ellipsoids through time for the case of identifying a 2^{nd} order FIR filter under uniform noise. The '$*$' denotes $\boldsymbol{\theta}_*$ while 'o' denotes $\boldsymbol{\theta}_t$. It can be seen that in just 25 samples, the central estimate seems to converge to the true parameter and the associated ellipsoid shrank to a much smaller size, albeit not yet a point. This can be explained from the fact that the centroid of the QOBE algorithm can also be looked upon as the solution of a certain non-linear least-squares problem under the constraint of bounded noise [5].

For tracking time-varying parameters, we simulated a 2^{nd} order FIR filter similar to the earlier case, but with randomly time varying coefficients. The time variations in the parameter are introduced by having random, but bounded jumps at every 15 samples. It can be seen from Figure (4), that the QOBE algorithm outperforms weighted RLS (with an exponential forgetting factor of 0.9) in terms of tracking, while using only around 25 data points for updating, out of 100. This superior tracking capability is an outcome of the "intelligent" choice of weights as also the selective updating mechanism.

III. Conclusions

This paper presented studies on a novel OBE algorithm, termed QOBE, for real-time signal processing and identification. Excellent convergence properties exhibited by the algorithm in stationary and non-stationary environments were discussed. However, a rigorous convergence proof of this algorithm still remains an open issue although the analysis presented in this paper is a significant step in this direction. Simulation studies conducted demonstrated the superior performance offered by QOBE over standard techniques like RLS.

References

[1] 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 Transactions on Information Theory*, Vol. 33, No. 3, pp. 383-392, May 1987.

[2] J.R. Deller, M. Nayeri and M.S. Liu, "Unifying the landmark developments in optimal bounding ellipsoid identification," *Int. J. Automatic Control and Signal Processing*, vol. 8, pp. 48-63, Jan.-Feb. 1994.

[3] E. Fogel and Y. F. Huang, "On the value of information in system identification– Bounded noise case," *Automatica*, vol. 18, pp. 229-238, 1982.

[4] D. Joachim, J. R. Deller and M. Nayeri, "Practical considerations in the use of a new OBE algorithm that blindly estimates error bounds," to be presented at the 40^{th} *Midwest Symp. Circ. and Sys.*, Sacramento, CA, August 1997.

[5] S. Nagaraj, S. Gollamudi, S.Kapoor and Y. F. Huang, "Bounded error estimation: Set-theoretic and least-squares formulations," *Proc. Conf. on Info. Sci. Sys. (CISS)*, The Johns Hopkins University, Baltimore, MD, March 1997.

[6] J. P. Norton (editor), *Int. Journal of Automatic Control and Signal Proc.* (Special issues on bounded-error methods in system identification), Vol. 8, Jan-Feb, 1993-4.

[7] E. Walter and H. Piet-Lahanier, "Estimation of parameter bounds from bounded-error data: A survey," *Mathematics and Computers in Simulation*, vol. 32, pp. 449-468, 1990.

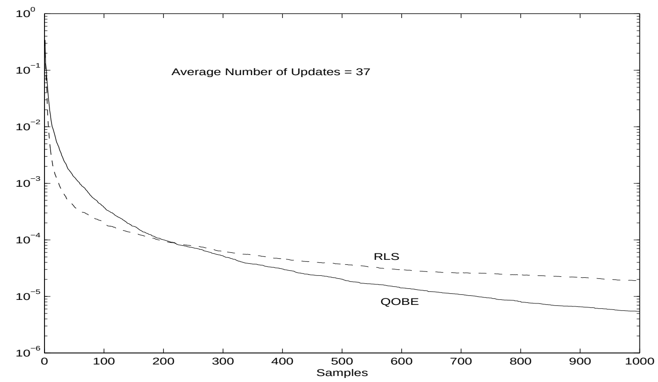

Fig. 1. Estimation of an 4 Tap FIR filter with Uniform Noise - Mean Square Parameter Error

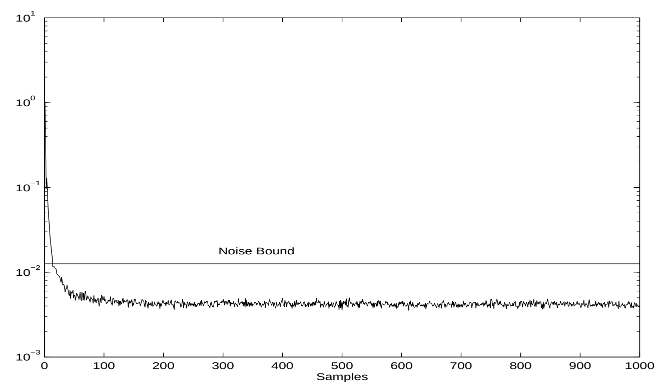

Fig. 2. Plot of Mean Squared Prediction error

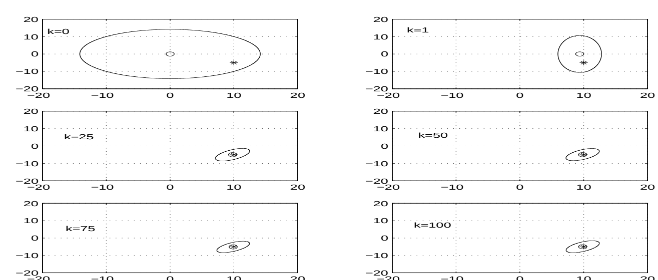

Fig. 3. 2-Tap FIR filter - evolution of the ellipsoids and the central estimate through time (k)

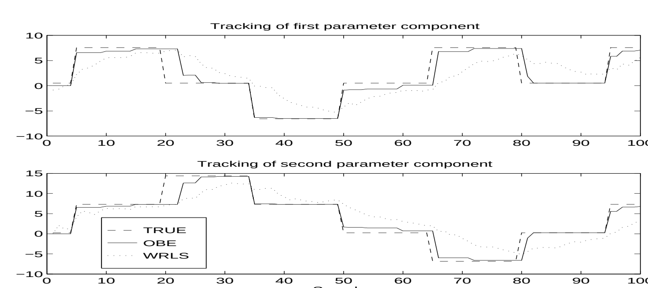

Fig. 4. Tracking performance comparison