

Set-Membership Filtering and a Set-Membership Normalized LMS Algorithm with an Adaptive Step Size

Sridhar Gollamudi, *Student Member, IEEE*, Shirish Nagaraj, *Student Member, IEEE*,
Samir Kapoor, *Student Member, IEEE*, and Yih-Fang Huang, *Fellow, IEEE*

Abstract—Set-membership identification (SMI) theory is extended to the more general problem of linear-in-parameters filtering by defining a set-membership *specification*, as opposed to a bounded noise *assumption*. This sets the framework for several important filtering problems that are not modeled by a “true” unknown system with bounded noise, such as adaptive equalization, to exploit the unique advantages of SMI algorithms. A recursive solution for set membership filtering is derived that resembles a variable step size normalized least mean squares (NLMS) algorithm. Interesting properties of the algorithm, such as asymptotic cessation of updates and monotonically nonincreasing parameter error, are established. Simulations show significant performance improvement in varied environments with a greatly reduced number of updates.

I. INTRODUCTION

SET-MEMBERSHIP identification (SMI) and a popular class of recursive SMI algorithms known as *optimal bounding ellipsoids* (OBE) algorithms (see, e.g., [1], [2], [4]) are well-established paradigms in the area of system identification that exploit the assumption of a bounded noise process added to a linear-in-parameter model. The OBE algorithms compute ellipsoidal approximations to regions in the parameter space that are consistent with the observed data and model assumptions. Several investigations (see, e.g., [1]–[4], [6], [7]) have shown the attractive features offered by the OBE algorithms and their advantages over conventional adaptive techniques. Formulations for the OBE recursions look strikingly similar to those of the recursive least-squares (RLS) algorithm. In fact, the OBE algorithms essentially execute weighted least-squares recursions with an *optimized* adaptive weighting sequence (see Deller *et al.* [6]). Furthermore, the OBE algorithms offer significant advantages over the RLS algorithms in several respects. First, they provide region estimates in addition to point estimates. Second, they exhibit better convergence and tracking properties due to the optimization of the weighting sequence (see [3], [6], and [10]). Most OBE algorithms have proven convergence [1], [2], [5]. Third, the

Manuscript received April 25, 1997. 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. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. J. J. Shynk.

The authors are with the Laboratory for Image and Signal Analysis, Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556 USA (e-mail: huang.2@nd.edu).

Publisher Item Identifier S 1070-9908(98)04236-9.

OBE algorithms feature a unique selective update criterion that requires parameter updates to be computed for only a small fraction of the data (typically less than 20%). This feature can be exploited to obtain significant computational reduction by time-sharing schemes [8], [9].

However, the theory of SMI has so far been limited to system identification, which excludes several important filtering problems like inverse filtering, channel equalization, echo cancellation, interference suppression, and beamforming that are not modeled by a “true” linear-in-parameter system with additive bounded noise. The extension of set-membership theory to the general filtering problem is therefore of immense academic and practical interest. This paper proposes a framework for set-membership filtering (SMF) that encompasses, but is not limited to, system identification. The problem is formulated on the basis of a bounded-error *specification* as opposed to a bounded-noise *assumption*. The class of adaptive solutions to this problem, referred to here as *set-membership adaptive recursive techniques* (SMART), are seen to include the OBE algorithms. A novel algorithm in SMART is derived whose recursions are identical to those of the normalized least mean-squares (NLMS) algorithm with an “optimized” adaptive step size. This establishes a set-membership equivalent of the stochastic gradient algorithm in much the same way as the OBE algorithms are equivalents of the deterministic least-squares algorithms. This new algorithm, referred to here as SM-NLMS, is derived from two distinct approaches: the optimal bounding spheroids (OBS) approach and the projection-based approach. The new algorithm offers automatic selection of the step-size that is very useful in unknown or time-varying environments and a proof of convergence in terms of guaranteed cessation of updates. Some other interesting properties of the algorithm are also established.

The letter is organized as follows. Section II defines the problem formulation for set-membership filtering and suggests adaptive solutions by SMART. The set-membership NLMS algorithm, a linear complexity NLMS-like recursive algorithm in SMART is derived in Section III.

II. SET-MEMBERSHIP FILTERING

Consider a general linear-in-parameter filter in which $\mathbf{x} \in \mathbb{C}^N$ and complex scalar d are, respectively, the input and desired output. The output of the filter is $y(\theta) = \theta^T \mathbf{x}$,

where $\theta \in \mathbb{C}^N$ is the parameter vector. The filter error is defined as $e(\theta) = d - y(\theta)$. The class of filters modeled by the above includes finite impulse response (FIR) and infinite impulse response (IIR) filters that are commonly used in signal processing and communication systems as well as several nonlinear systems. Specifically, \mathbf{x} may contain nonlinear functions of past filter outputs, as is the case in decision feedback equalization.

Conventional filtering schemes estimate the parameter vector θ so as to minimize a cost function which is usually a direct function of the estimation error e . Examples of such cost functions are the time-average of e^2 (weighted least-squares), or the ensemble average of e^2 (MSE). In contrast, the objective of SMF is to achieve a specified *bound* on the magnitude of the estimation error e over a model space of interest. We start by defining a model space \mathcal{S} comprising input vector-desired output pairs over which we wish to impose the bounded error criterion. Any parameter estimate that results in the error being less than the specified bound γ for all data pairs from \mathcal{S} is an acceptable solution. Therefore, the solution to the SMF problem is a *set* in the parameter space rather than a point estimate. The SMF criterion is to find θ that satisfies

$$|e(\theta)|^2 \leq \gamma^2 \quad \text{for all } (\mathbf{x}, d) \in \mathcal{S}. \quad (1)$$

The desired region estimate, called the *feasibility set* Θ , is the set of parameter vectors that satisfy (1). In other words

$$\Theta \triangleq \bigcap_{(\mathbf{x}, d) \in \mathcal{S}} \{\theta \in \mathbb{C}^N: |d - \theta^T \mathbf{x}|^2 \leq \gamma^2\}. \quad (2)$$

If the bound on the error γ is properly chosen, then the feasibility set is nonempty and any point in it is a valid estimate. If the specification is too stringent, however, the feasibility set may be empty for a given model space of interest. It is assumed in the remainder of this paper that the feasibility set is nonempty.

It is important to emphasize that the inequality (1) is a *specification* on the adaptive filter, as opposed to an assumed *bound* on the additive disturbance, as is the case in set-membership identification (SMI). SMI can be seen to be a special case of the above problem when (\mathbf{x}, d) are the input, output pair of the plant to be identified. If it is known that the additive noise to this plant is bounded by γ_v , then Θ is clearly nonempty for $\gamma \geq \gamma_v$.

The feasibility set or a point in it can be estimated offline if the model space \mathcal{S} is known *a priori* (see, e.g., [9]). However, there is no closed-form solution for the feasibility set given an arbitrary model space, and solutions for particular cases may be too complex. Also, in practice, the model space is not completely known and/or is time-varying, as is always the case in SMI and in channel equalization. There is thus a necessity for developing *adaptive* algorithms that estimate the feasibility set or one of its members. We shall call such techniques *SMART* (defined above).

It is assumed that a sequence of data sets $(\mathbf{x}_n, d_n) \in \mathcal{S}$ is available for "training." At time n , the *constraint set* \mathcal{H}_n is defined as the set of all parameter vectors that are consistent with the specification (1) and the observation at time instant

n as follows:

$$\mathcal{H}_n = \{\theta \in \mathbb{C}^N: |d_n - \theta^T \mathbf{x}_n| \leq \gamma\}. \quad (3)$$

The constraint set is the region enclosed by the set of parallel hyperplanes described by $|d_n - \theta^T \mathbf{x}_n| = \gamma$.

The minimal set estimate for Θ at time n is the *membership set* $\psi_n = \bigcap_{i=1}^n \mathcal{H}_i$. Clearly, the membership set is a superset of the feasibility set. However, exact membership sets are convex polytopes in the parameter space that are analytically and computationally intractable. Adaptive solutions for the SMF problem can be derived from two separate approaches: i) by forming a sequence of parameter estimates that (hopefully) converges into the feasibility set (point-wise approach), or ii) by forming a sequence of sets that *tightly* outer-bound the membership sets (set bounding approach). The OBE algorithms can be easily verified to belong to the latter.

III. SET-MEMBERSHIP NLMS ALGORITHM

The SM-NLMS algorithm, as a SMART, can be derived from both approaches mentioned above. The point-wise approach is to take projections of the previous estimate on the constraint set at each time.

Formulation 1—Point-Wise Approach: Find $\hat{\theta}_n$ so as to minimize the Euclidean norm of the change in the estimate given by $\|\hat{\theta}_n - \hat{\theta}_{n-1}\|$, subject to the constraint that $\hat{\theta}_n \in \mathcal{H}_n$.

In other words, the goal is to move into the constraint set at time n by traversing the minimum possible distance in the parameter space, in accordance with *the principle of minimum disturbance*, [11]. When $\hat{\theta}_{n-1} \notin \mathcal{H}_n$, this corresponds to choosing the foot of the perpendicular from $\hat{\theta}_{n-1}$ to the nearest bounding hyperplane of the constraint set, as shown in Fig. 1(a). If $\hat{\theta}_{n-1} \in \mathcal{H}_n$, however, Formulation 1 implies $\hat{\theta}_n = \hat{\theta}_{n-1}$, i.e., no update of the parameter estimate is necessary.

The set-bounding approach is to compute outer approximations of the membership set by minimal volume spheroids. A spheroid, S_n , in the N -dimensional complex parameter space is described by

$$S_n = \{\theta \in \mathbb{C}^N: \|\theta - \hat{\theta}_n\|^2 \leq \sigma_n^2\} \quad (4)$$

where $\hat{\theta}$ is the center of the spheroid and σ_n is its radius. Then, the criterion for update is as follows.

Formulation 2—Optimal Bounding Spheroids Approach: Compute a smallest spheroid that contains the intersection of the spheroidal set estimate S_{n-1} and the constraint set \mathcal{H}_n .

Theorem 1: The solutions to both Formulations 1 and 2 are unique and identical. They are given by the following recursions:

$$\hat{\theta}_n = \hat{\theta}_{n-1} + \alpha_n \frac{\delta_n \mathbf{x}_n^*}{\mathbf{x}_n^H \mathbf{x}_n} \quad (5)$$

$$\sigma_n^2 = \sigma_{n-1}^2 - \alpha_n^2 \frac{|\delta_n|^2}{\mathbf{x}_n^H \mathbf{x}_n} \quad (6)$$

where the prediction error δ_n and the gain α_n are given by

$$\delta_n = d_n - \hat{\theta}_{n-1}^T \mathbf{x}_n$$

$$\alpha_n = \begin{cases} 1 - \frac{\gamma}{|\delta_n|}, & \text{if } |\delta_n| > \gamma \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

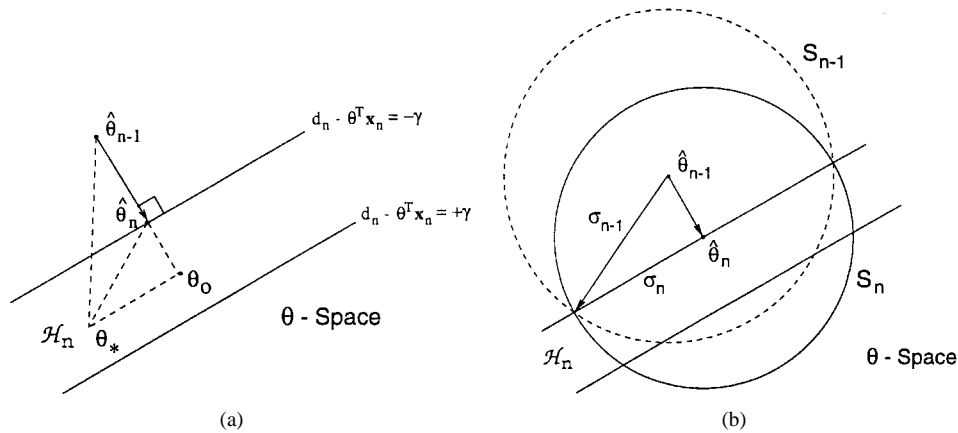


Fig. 1. Derivation of the SM-NLMS algorithm. (a) Point-wise approach. (b) Optimal bounding spheroids approach.

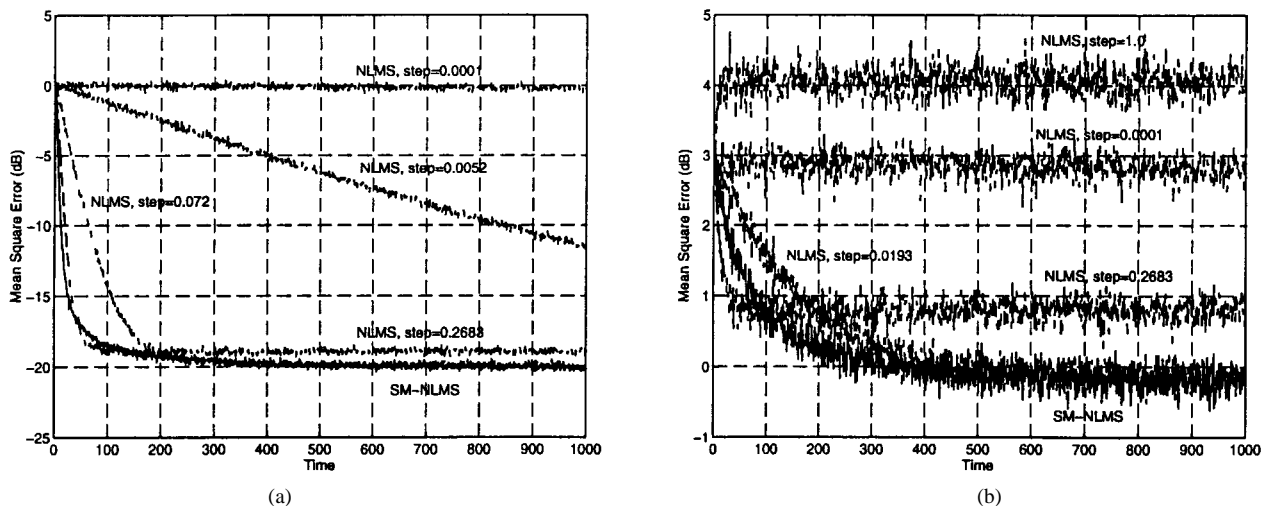


Fig. 2. Mean square error performance of the SM-NLMS algorithm compared with the NLMS algorithm. (a) Eigenvalue spread of the input vectors = 100 and SNR = 20 dB. (b) Eigenvalue spread = 1 and SNR = 0 dB.

Sketch of Proof: The solution to Formulation 1 can be derived geometrically using Fig. 1(a). In the solution for Formulation 2, S_n is the smallest spheroid containing $\mathcal{H}_n \cap S_{n-1}$. Therefore the “diameter” of S_n is the nearest bounding hyperplane of \mathcal{H}_n bounded by S_{n-1} . The center $\hat{\theta}_n$ and radius σ_n can be derived from Fig. 1(b). Extension to multidimensional complex vectors is straightforward. \square

It may be observed that the SM-NLMS recursion, although derived from a completely different philosophy, is identical to the recursion in the conventional NLMS algorithm, except for the assignment to the gain term α_n . The gain α_n is a constant in conventional NLMS, whereas it is adaptive in SM-NLMS, reflecting the information provided by the current data set. The variable step size is *optimal* in the sense of Formulations 1 or 2, as shown above. Furthermore, the SM-NLMS algorithm guarantees nonincreasing parameter error, as described by the following theorem.

Theorem 2: The magnitude of the parameter error $\|\theta_* - \hat{\theta}_n\|$, where θ_* is any point in the feasibility set Θ and $\hat{\theta}_n$ is given by (5), is a monotonically nonincreasing sequence.

Sketch of Proof: In Fig. 1(a), let the extension of the line connecting $\hat{\theta}_{n-1}$ and $\hat{\theta}_n$ intersect the hyperplane ($\theta -$

$\theta_*)^T \mathbf{x}_n = 0$ at θ_o . Then, using $(\hat{\theta}_{n-1} - \theta_o) \perp (\theta_o - \theta_*)$, we have

$$\begin{aligned} \|\theta_* - \hat{\theta}_n\|^2 &= \|\theta_* - \theta_o\|^2 + \|\theta_o - \hat{\theta}_n\|^2 \\ &\leq \|\theta_* - \theta_o\|^2 + \|\theta_o - \hat{\theta}_{n-1}\|^2 = \|\theta_* - \hat{\theta}_{n-1}\|^2. \end{aligned} \quad (8)$$

At every instant of time, the above theorem implies that *each* point in the feasibility set is closer to the new estimate than the previous estimate. If the feasibility set is time-varying, the point estimate always moves closer to each point in the *present* feasibility set, indicating good tracking capability in a time-varying environment. In terms of set estimates, Theorem 1 implies that the SM-NLMS algorithm provides a sequence of *set error* $\{S_n \setminus \Theta\}$ with monotonically nonincreasing volumes.

Convergence of the algorithm in terms of the learning-rate α_n converging to zero is established by the following theorem. Note that this requires no conditions on the data except that $(\mathbf{x}_n, d_n) \in \mathcal{S}$ for all n . In particular, no persistence of excitation conditions are necessary.

Theorem 3: The magnitude of parameter updates and the sequence of step-sizes $\{\alpha_n\}$ of SM-NLMS converge to zero,

i.e.,

$$\lim_{n \rightarrow \infty} \|\hat{\theta}_n - \hat{\theta}_{n-1}\| = \lim_{n \rightarrow \infty} \alpha_n = 0 \quad (9)$$

and the magnitude of prediction error is asymptotically smaller than γ , i.e.,

$$\limsup_{n \rightarrow \infty} |\delta_n| \leq \gamma. \quad (10)$$

Proof: Let the time index n denote just the updating instants, since $\|\hat{\theta}_n - \hat{\theta}_{n-1}\| = \alpha_n = 0$, otherwise. For all n , $\sigma_n^2 > 0$, since the spheroid S_n is nonempty (it contains the feasibility set). And since σ_n^2 is monotone decreasing, the sequence $\{\sigma_n^2\}$ is convergent. From (6), this implies $\alpha_n^2 |\delta_n|^2 / \|\mathbf{x}_n\|^2 \rightarrow 0$. It follows from (5) that $\|\hat{\theta}_n - \hat{\theta}_{n-1}\| = \alpha_n |\delta_n| / \|\mathbf{x}_n\| \rightarrow 0$. Assuming $\|\mathbf{x}_n\|$ is bounded and using $|\delta_n| > \gamma > 0$ in updating instants, it also follows that $\alpha_n \rightarrow 0$. Using (7), this implies that $|\delta_n| \rightarrow \gamma$ in the updating instants and $|\delta_n| < \gamma$ otherwise, resulting in the required result (10). \square

The performance gain of the proposed algorithm over the NLMS algorithm is shown via simulations in Fig. 2. The figure shows MSE curves for the identification of a fifth-order plant with additive white truncated Gaussian noise, truncated to $\pm\sqrt{5\sigma_v^2}$, where σ_v^2 is the variance of the Gaussian distribution. The error bound γ is also chosen to be $\sqrt{5\sigma_v^2}$. Note the drastic reduction in the number of updates needed. The proposed algorithm combines the convergence speed of a high step-size NLMS and the low steady-state MSE of a low step-size NLMS algorithm. It also obviates *a priori* knowledge of input statistics that are necessary in the case of NLMS to heuristically set the step size.

IV. CONCLUSION

This letter introduced the notion of set-membership filtering, a novel formulation of the linear-in-parameter filtering problem that involves estimation of a feasible set of filter parameters to meet a specified bound on estimation error. A toolbox of adaptive algorithms for set-membership filtering, called SMART, was presented. Optimal bounding ellipsoids algorithms, originally developed for system identification, were shown to be viable algorithms in SMART. An algorithm for

SMART was derived in the set-membership filtering framework that has the same recursions as the normalized LMS algorithm, except that the new algorithm provides optimal data-dependent gain term that obviates the need for heuristics for its choice. The SM-NLMS algorithm guarantees nonincreasing magnitude of parameter error and also offers set estimates of nonincreasing size in the parameter space. A proof of cessation of updates is established. Further convergence analysis of the algorithms in SMART in general and the SM-NLMS algorithm in particular will be presented in a forthcoming work.

REFERENCES

- [1] E. Fogel and Y. F. Huang, "On the value of information in system identification-bounded noise case," *Automatica*, vol. 18, pp. 229–238, Mar. 1982.
- [2] 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. Inform. Theory*, vol. IT-33, pp. 383–392, 1987.
- [3] A. K. Rao and Y. F. Huang, "Tracking characteristics of an OBE parameter estimation algorithm," *IEEE Trans. Signal Processing*, vol. 41, pp. 1140–1148, Mar. 1993.
- [4] J. R. Deller, Jr., M. Nayeri, and M. S. Liu, "Unifying the landmark developments in optimal bounding ellipsoid identification," *Int. J. Adapt. Contr. Signal Process.*, vol. 8, pp. 43–60, Jan./Feb. 1994.
- [5] M. Nayeri, M. S. Liu, and J. R. Deller, "Do interpretable optimal bounding ellipsoid algorithms converge?—Part I and II," in *Proc. IFAC Symp. System Identification*, Copenhagen, Denmark, 1994.
- [6] J. R. Deller, Jr., M. Nayeri, and S. F. Odeh, "Least-squares identification with error bounds for real-time signal processing and control," *Proc. IEEE*, vol. 81, pp. 815–849, June 1993.
- [7] E. Walter and H. Piet-Lahanier, "Estimation of parameter bounds from bounded-error data: A survey," *Math. Comput. Simul.*, vol. 32, pp. 449–465, Dec. 1990.
- [8] S. Gollamudi and Y. F. Huang, "Updater-shared adaptive parallel equalization (U-SHAPE) using set-membership identification," in *Proc. IEEE Int. Symp. Circuits and Systems*, Atlanta, GA, May 13–15, 1996, pp. 13–16.
- [9] S. Gollamudi, S. Kapoor, S. Nagaraj, and Y. F. Huang, "Set-membership adaptive equalization and an updater-shared implementation for multiple channel communications systems," *IEEE Trans. Signal Processing*, to be published.
- [10] 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*, The Johns Hopkins Univ., Baltimore, MD, Mar. 1997.
- [11] B. Widrow and M. Lehr, "30 years of adaptive neural networks: Perceptron madaline and backpropagation," *Proc. IEEE*, vol. 78, Sept. 1990.