BEACON: An Adaptive Set-Membership Filtering Technique with Sparse Updates

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

Abstract—This paper deals with adaptive solutions to the so-called *set-membership filtering (SMF)* problem. The SMF methodology involves designing filters by imposing a deterministic constraint on the output error sequence. A set-membership decision feedback equalizer (SM-DFE) for equalization of a communications channel is derived, and connections with the minimum mean square error (MMSE) DFE are established. Further, an adaptive solution to the general SMF problem via a novel optimal bounding ellipsoid (OBE) algorithm called BEACON is presented. This algorithm features sparse updating, wherein it uses about 5–10% of the data to update the parameter estimates without any loss in mean-squared error performance, in comparison with the conventional recursive least-squares (RLS) algorithm. It is shown that the BEACON algorithm can also be derived as a solution to a certain constrained least-squares problem. Simulation results are presented for various adaptive signal processing examples, including estimation of a real communication channel. Further, it is shown that the algorithm can accurately track fast time variations in a nonstationary environment. This improvement is a result of incorporating an explicit test to check if an update is needed at every time instant as well as an optimal data-dependent assignment to the updating weights whenever an update is required.

I. INTRODUCTION

THE PROBLEM of designing a linear-in-parameter filter, given knowledge of the input and the corresponding desired output, is studied in this paper. Traditional methodologies include the minimum mean square error (MMSE) filters and their deterministic counterparts [the least-squares error (LSE) filters], which seek a filter by minimizing the 2-norm of the error sequence [1]. Performance of the MMSE filters relies on the accurate knowledge of the statistics of the input and output observations, whereas an off-line LSE procedure requires data processing in batches, which is not computationally attractive. On-line, or recursive, methods to iteratively achieve the same goals include the least-mean squares (LMS) and the recursive least-squares (RLS) algorithms, which have been studied extensively in the literature; see, e.g., [1] and [2].

In conventional system identification problems, we have some *a priori* knowledge about the system to be estimated. Incorporating this in the form of constraints in the estimation procedure leads to solutions that are consistent with that knowledge. In a more general framework of filter design, constraints can be imposed in the estimation procedure if we need assurance of *good* performance on a deterministic (point-wise) basis. As before, these requirements then have to be incorporated in the design of the estimator/filter to ensure such an acceptable performance. Moreover, by taking the structure and constraints of the problem into account, it is likely that computationally attractive recursive algorithms emerge as possible solutions. With these motivations in mind, the authors have recently introduced, in [3], a methodology for the design of filters that bound the worst-case error achieved by the filter. This method is termed *set-membership filtering (SMF)*, and the resulting filter is called an SM filter [3]. SMF owes its name to the so-called set-membership identification (SMI) technique [4]–[8], which is applicable only for identifying a linear-in-parameter plant with output corrupted by additive bounded noise. It is straightforward to show that SMI is a special case of the SMF problem.

The objective of an SM filter is to estimate a member of a so-called *feasibility set*. This set defines the SM filter's performance specification. SMF requires that this specification be met for every possible input-desired output pair of data that come from a certain *design space*. Any closed-form solution of an SMF problem requires accurate characterization of the design space over which the filter is required to meet the specification. Such a description of the design space would, in general, need knowledge of a functional relationship between the input and desired outputs. In problems where such knowledge is unknown or is not accurate enough, we require a tool to estimate a point in the feasibility set in a recursive fashion. It has been shown that one such method is given by the class of optimal bounding ellipsoid (OBE) algorithms. These algorithms were originally developed for SMI [9] and have gained much attention in the past decade [7], [8], [10] due to some of their attractive features. Among others, the OBE algorithms employ a *discerning update* rule, i.e., they use the data selectively in updating the parameter estimates.

In this paper, we present a novel OBE algorithm called the Bounding Ellipsoidal Adaptive CONstrained least-squares (BEACON) algorithm, which shares many of the desirable features exhibited by the various OBE algorithms [5] developed to date. In addition, however, the proposed algorithm features
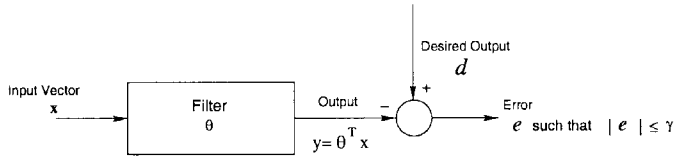
Fig. 1. Set-membership filtering.

some favorable characteristics that are unique and different from traditional OBE algorithms. The BEACON algorithm is found to update much less frequently in comparison with the existing OBE algorithms such as the one developed by Dasgupta and Huang called DH-OBE [4], whereas it exhibits better mean-squared error (MSE) performance than both DH-OBE and the weighted RLS. This paper examines some issues regarding the updating mechanism and provides a novel geometrical interpretation. Moreover, it investigates the behavior of this algorithm when the initial conditions are incorrectly chosen in such a way that the basic premise of bounding ellipsoids fails. Surprisingly, it turns out that a certain point estimate obtained via the BEACON recursions is insensitive to these *model violations*, and some convergence results also hold. It is also observed that BEACON exhibits very good tracking capabilities for time-varying systems.

The above features exhibited by the point estimate motivate the development of a least-squares-like estimation scheme that is constrained on the specification of bounded errors. This approach, although it is decoupled from the OBE method, is shown to lead to the same recursions for the parameter (point) estimate as before. Moreover, it also establishes a clear link between least-squares and OBE estimation paradigms.

The next section first overviews the concept of set-membership filtering and then addresses an application for a special but important problem in digital communications, namely, that of channel equalization. Specifically, conditions for the existence of the *feasibility set* for a *decision feedback equalizer (DFE)* will be derived. Relations between the so-called SM-DFE and the commonly used MMSE-DFE are studied. Section III introduces the BEACON algorithm as a recursive solution to SMF, gives an explicit formula to compute the weights for updating, and presents some analysis of its performance. Simulation results are also provided in this section. The alternative recursive solution method for SMF, via a constrained least-squares method, is developed in Section IV. Finally, Section V concludes the paper.

II. SET-MEMBERSHIP FILTERING

A. Problem Formulation of SMF

Set-membership filtering is a filtering method distinct from traditional least-squares filtering in terms of performance criterion [3], as depicted in Fig. 1. The objective is to design a filter whose output error is not greater than a specified value for all possible input-desired output pairs of data. In this paper, we restrict our attention to linear-in-parameter filters of the form (1). To formalize the above idea, assume that the input, which is denoted by $\boldsymbol{x}$, and the corresponding desired output d come from a certain design space $\mathcal{D} \subset \mathbb{C}^n \times \mathbb{C}$, where $\mathbb{C}^n$ is the

n-dimensional complex Euclidean space. The problem can be posed as follows: *Given $\mathcal{D}$ and a designer-specified positive real number γ, design a filter, $f_\theta\colon \mathbb{C}^n \times \mathbb{C} \to \mathbb{C}$ that (linearly) maps the pairs $(\boldsymbol{x}, d)$ to the output error*[1]

$$e \triangleq f_\theta(\boldsymbol{x}, d) = d - \theta^T \boldsymbol{x} \tag{1}$$

where $\theta \in \mathbb{C}^n$. The objective of SMF is to choose a parameter vector θ such that

$$|e| = |f_\theta(\boldsymbol{x}, d)| \leq \gamma \quad \forall (\boldsymbol{x}, d) \in \mathcal{D}. \tag{2}$$

In other words, find a parameter vector that meets the error specification for all input-desired output pairs in $\mathcal{D}$. Therefore, the SMF problem can be cast in the form of the following two parts:

- *Does there exist a θ such that*

$$\sup_{(\boldsymbol{x},d)\in\mathcal{D}} |f_\theta(\boldsymbol{x}, d)| \leq \gamma? \tag{3}$$

- *If so, find a θ^o such that (3) is satisfied with $\theta = \theta^o$.*

In contrast, traditional methodologies like LSE and MMSE minimize the squared error either in a deterministic or a stochastic setting. The question of the existence of an SM filter for a given specification is an important issue.

In general, the bounded error specification (2) can be met by a *set* of filters, which is referred to as the *feasibility set*, since any one member of this set is a valid SM filter. Therefore, the objective in SMF is to estimate this feasibility set or a member of it. This set is given by

$$\begin{aligned}
\boldsymbol{\Theta}(n, \gamma) &\triangleq \bigcap_{(\boldsymbol{x},d)\in\mathcal{D}} \{\boldsymbol{\theta} \in \mathbb{C}^n\colon |d - \theta^T\boldsymbol{x}|^2 \leq \gamma^2\} \\
&= \{\boldsymbol{\theta} \in \mathbb{C}^n\colon \sup_{(\boldsymbol{x},d)\in\mathcal{D}} |f_\theta(\boldsymbol{x}, d)| \leq \gamma\}
\end{aligned}$$

.

The feasibility set is an ensemble intersection over all the possible input-desired output pairs that come from the design space. A few properties of this set are listed below.

- $\boldsymbol{\Theta}(n, \gamma)$ is a convex set.
- $\boldsymbol{\Theta}(n, \gamma) \subset \boldsymbol{\Theta}(n + m, \gamma)$ for any $m > 0$.
- $\boldsymbol{\Theta}(n, \gamma_1) \subset \boldsymbol{\Theta}(n, \gamma_2)$, where $\gamma_1 < \gamma_2$.
- If $\mathcal{D}_1 \subset \mathcal{D}_2$, then $\boldsymbol{\Theta}_2(n, \gamma) \subset \boldsymbol{\Theta}_1(n, \gamma)$, where

$$\boldsymbol{\Theta}_i(n, \gamma) \triangleq \bigcap_{(\boldsymbol{x},d)\in\mathcal{D}_i} \{\boldsymbol{\theta} \in \mathbb{C}^n\colon |d - \theta^T\boldsymbol{x}|^2 \leq \gamma^2\}$$

$$i = 1, 2.$$

It may happen that there exists no parameter vector that can meet the performance criterion for a particular choice of γ and $\mathcal{D}$, as mentioned earlier. Therefore, once the design space is fixed, the chance of finding a nonempty $\boldsymbol{\Theta}(n, \gamma)$ increases with n and γ, which is intuitively well understood. In addition, a larger design space leads to a smaller feasibility set. That is, as we increase the number of data pairs for which the specification is to be met, there are fewer filters that can

---

[1] For a vector $\boldsymbol{x} \in \mathbb{C}^n$, $\|\boldsymbol{x}\|_1$ and $\|\boldsymbol{x}\|_2$, denote the standard 1- and 2-norms while $\boldsymbol{x}^T, \boldsymbol{x}^*, \boldsymbol{x}^H$ stand for the transpose, conjugate, and Hermitian, respectively.
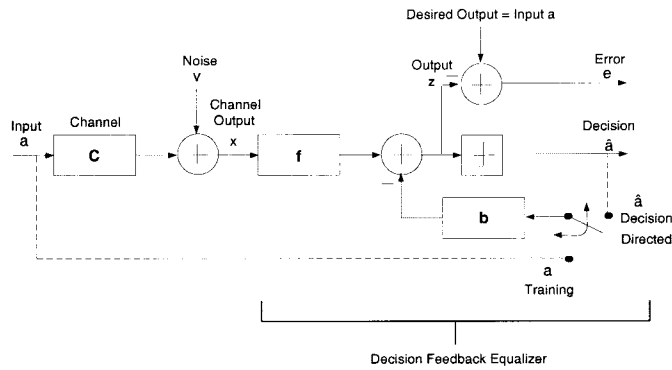
Fig. 2.   Discrete-time model of a communication system with a DFE.

meet the criterion. The following observation is utilized for designing an SM filter.

In order to determine if the feasibility set $\Theta(n, \gamma)$ is nonempty or not, it is enough to check if $|d - \theta^{o T} x|^2 \leq \gamma^2$ for all $(x, d) \in \mathcal{D}$, where

$$\theta^o \triangleq \arg \min_{\theta \in \mathbb{C}^n} \sup_{(x, d) \in \mathcal{D}} |f_\theta(x, d)| \qquad (4)$$

where $f_\theta(x, d)$ is as defined in (1). That is, $\Phi(n, \gamma)$ is nonempty if and only if the above condition is met. In the subsequent sections, the applicability of the concept of SMF is demonstrated for the problem of channel equalization in communications using a decision feedback equalizer structure. For a linear equalizer, certain sufficient and necessary conditions for the existence of a nonempty feasibility set have been derived in [11]. The *set-membership decision feedback equalizer (SM-DFE)* is introduced in this paper using the DFE structure [12] with the resulting equalizer designed according to an SMF criterion. Conditions for the existence of a nonempty set of SM-DFE coefficients, and importantly, a result showing connections to the well-known MMSE-DFE are derived. It is also shown that the SM linear equalizer is a special case of the SM-DFE.

B. Set-Membership Decision Feedback Equalization

To set up the model for channel equalization, consider the equivalent complex baseband discrete-time channel model [12], as depicted in Fig. 2. The channel consists of the transmitting filter, modulator, the physical channel, the demodulator, and the receiving filter. Let $\{a_i\}_{i=-\infty}^{\infty}$ be the transmitted symbol sequence and $\{\nu_i\}_{i=-\infty}^{\infty}$ be the additive noise at the channel output. The discrete-time sampled output of the channel $\{x_i\}_{i=-\infty}^{\infty}$ is given by

$$x_i = \sum_{k=-D}^{D} c_k a_{i-k} + \nu_i \qquad (5)$$

where $\{c_k\}_{k=-D}^{D}$ is the (finite length and sampled) channel impulse response.

Consider a decision feedback equalizer whose output z_i at time i is given by

$$z_i = \sum_{k=-N_f}^{N_f} f_k x_{i-k} - \sum_{k=1}^{N_b} b_k \hat{a}_{i-k} = f^T x_i - b^T \hat{a}_i \qquad (6)$$

where $f = [f_{-N_f}, \cdots, f_{N_f}]^T$ is the vector of coefficients of the feedforward filter, and $b = [b_1, \cdots, b_{N_b}]^T$ is the (strictly causal) feedback filter tap weight vector. Depending on the sampling rate, this model can be shown to encompass fractionally or symbol-spaced equalizers. The vector x_i is the regressor input to the feedforward equalizer, i.e., the vector comprising of the channel outputs $x_i \triangleq [x_{i+N_f}, \cdots, x_{i-N_f}]^T$, which, from (5), can be written as

$$x_i = C^T a_i + v_i \qquad (7)$$

where

$$a_i = [a_{i+K}, \cdots, a_{i-K}]^T, \qquad K = D + N_f$$
$$v_i = [\nu_{i+N_f}, \cdots, \nu_{i-N_f}]^T.$$

The channel convolution matrix C is

$$C = \begin{pmatrix} c_{-D} & 0 & \cdots & & 0 \\ c_{-D+1} & c_{-D} & & & \\ \vdots & c_{-D+1} & \ddots & & \vdots \\ c_D & & \ddots & \ddots & 0 \\ 0 & \ddots & & \ddots & c_{-D} \\ \vdots & & \ddots & & c_{-D+1} \\ & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & c_D \end{pmatrix}_{(2K+1)\times(2N_f+1)} .$$

In addition, $\hat{a}_i$ is the regressor vector of the previously decoded bits

$$\hat{a}_i = [\hat{a}_{i-1}, \cdots, \hat{a}_{i-N_b}]^T.$$

Let the transmitted symbols come from a constellation $\mathcal{A}$. In the case of equalization, it is intuitively appealing to fix the upper bound on the output error at $d_{\min}/2$, where $d_{\min}$ is the minimum Euclidean distance of two distinct symbols from the constellation $\mathcal{A}$. In other words, by such a design, the SM-DFE will have no decoding errors whenever the data pairs come from the design space $\mathcal{D}$.

In order to guarantee that the output error be bounded by some $\gamma > 0$, we need to define the design space for which the specification can hold. The input to the DFE can be represented by a concatenated vector $\tilde{x}_i \triangleq (x_i^T \quad \hat{a}_i^T)^T$, whereas the desired output is a_i. From (7), it follows that x_i depends only on a_i and v_i if C is assumed known. The vector $\hat{a}_i$ is not only specified by a_i and v_i but also depends on a nonlinear function of the parameter vectors f and b. In order to remove this dependence and to make the analysis tractable, the so-called *DFE assumption* [12] is commonly invoked, namely, that the previous decoded bits are correct. This implies

$$\hat{a}_i = P^T a_i$$

where

$$P^T = (0_{N_b \times (K+1)} | I_{N_b \times N_b} | 0_{N_b \times (K-N_b)}).$$

Here, $0_{N_b \times (K+1)}$ refers to a zero matrix of size N_b by $K+1$ and $I_{N_b \times N_b}$ to an identity matrix of size N_b by N_b.

The concatenated input to the DFE ($\tilde{x}_i$) is now completely specified by the pair of vectors a_i and v_i. The desired output
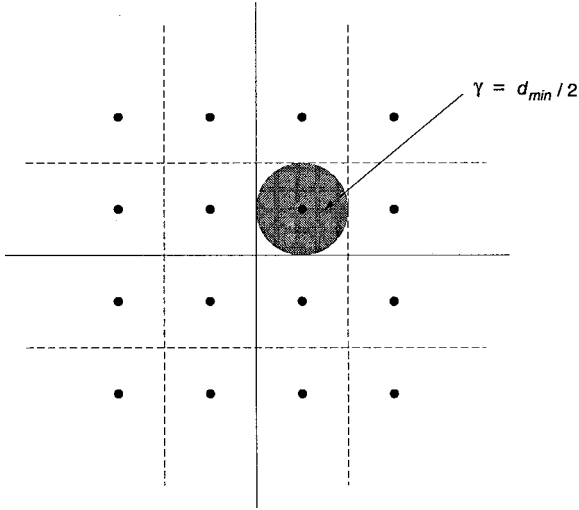
Fig. 3. Set-membership equalization criterion for a 16-QAM system. The equalized symbol is constrained to lie in a circle of radius γ centered at the transmitted symbol.

of the equalizer can be expressed as $a_i = e_0^T a_i$, where $e_0 = [0, \cdots, 1, \cdots, 0]^T$ is the unit vector of dimension $(2K+1)$ with a "1" in the $(K+1)$st position. Therefore, the pair $(\tilde{x}_i, a_i)$ is equivalent to

$$\tilde{x}_i = \begin{pmatrix} C^T a_i + v_i \\ P^T a_i \end{pmatrix}.$$
$$a_i = e_0^T a_i$$

As a result, the design space, which consists of the input-desired output pairs, can be equivalently described in terms of the Cartesian product space of the vectors a_i and v_i for all i, where a_i is taken from $\mathcal{A}^{2K+1}$ and $v_i \in \mathcal{V}^{2N_f+1}$. The set $\mathcal{V}$ consists of all noise components that are bounded

$$\mathcal{V} \triangleq \{v \in \mathbb{C}: |v|^2 \leq \gamma_\nu^2\} \tag{8}$$

for some noise bound $\gamma_\nu > 0$. Designing $\mathcal{D}$ requires fixing of the symbol constellation, the value of γ_ν, and the order of the equalizer. Rewriting (6) for the SM-DFE, we have [13]

$$z(f, b, a, v) = f^T x - b^T \hat{a},$$
$$\hat{a} = P^T a, \qquad (a, v) \in \mathcal{D}.$$

The dependence on the time variable has been dropped in this equation because the equation has to hold for all possible pairs from the design space and not just for a particular sequence.

The aim here is to ensure that the maximum Euclidean distance between the transmitted and equalized outputs (before decision) is upper bounded by a specified value $\gamma > 0$ for all the data sets that belong to the design space, i.e.,

$$|a - z(f, b, a, v)| \leq \gamma, \quad \forall (a, v) \in \mathcal{D}. \tag{9}$$

Condition (9) requires that the equalized output remain in a circle of radius γ centered at the transmitted symbol a, as shown in Fig. 3 for 16-QAM signaling with $\gamma = d_{\min}/2$.

We define the set of all feed-forward and feedback equalizer weight vectors of length $(2N_f+1)$ and N_b, respectively, that

satisfy (9) as the *feasible SM-DFE set* $\boldsymbol{\Theta}(n, \gamma)$

$$\boldsymbol{\Theta}(n, \gamma) \triangleq \bigcap_{(\boldsymbol{a}, \boldsymbol{v}) \in \mathcal{D}} \{(\boldsymbol{f} \in \mathbb{C}^{2N_f+1}, \boldsymbol{b} \in \mathbb{C}^{N_b}):$$
$$|a - z(\boldsymbol{f}, \boldsymbol{b}, \boldsymbol{a}, \boldsymbol{v})| \leq \gamma\}$$
$$= \bigcap_{(\boldsymbol{a}, \boldsymbol{v}) \in \mathcal{D}} \{(\boldsymbol{f}, \boldsymbol{b}):$$
$$|a - (\boldsymbol{f}^T \boldsymbol{x} - \boldsymbol{b}^T \hat{\boldsymbol{a}})| \leq \gamma, \hat{\boldsymbol{a}} = P^T \boldsymbol{a}\}$$

where $n \triangleq 2N_f + 1 + N_b$. Further, (9) can be rewritten as

$$\sup_{(\boldsymbol{a}, \boldsymbol{v}) \in \mathcal{D}} |(C\boldsymbol{f} - P\boldsymbol{b} - e_0)^T \boldsymbol{a} + \boldsymbol{f}^T \boldsymbol{v}| \leq \gamma. \tag{10}$$

It should be noted that the SM-DFE feasibility set is convex only when the DFE assumption is true, i.e., when all the past decisions are correct. The following proposition provides conditions for the existence of a nonempty parameter set $\boldsymbol{\Theta}(n, \gamma)$. For simplicity of discussion, assume that the input symbols and noise come from a zero-mean independent and identically distributed (i.i.d.) sequence (hence white) and that the noise and input symbols are uncorrelated.

Proposition 1—Existence of a Feasible Set of SM-DFE Filter Weights: Consider a linear channel model (5) and a DFE equalizer as described above. The following are true:

1) *Sufficiency:* $\boldsymbol{\Theta}(n, \gamma)$ is nonempty if the channel satisfies

$$\gamma_{\overline{a}} \|(C\boldsymbol{f}^{(s)} - P\boldsymbol{b}^{(s)} - e_0)\|_1 + \gamma_\nu \|\boldsymbol{f}^{(s)}\|_1 \leq \gamma \tag{11}$$

where $\gamma_{\overline{a}}$ is the maximum amplitude in the constellation $\mathcal{A}$, $\boldsymbol{f}^{(s)} = (C^H(C - PH) + (\gamma_\nu^2/\gamma_{\overline{a}}^2)I)^{-1}C^H e_0$, and $\boldsymbol{b}^{(s)} = H\boldsymbol{f}^{(s)}$, where H consists of the $(K + 2)$ to $(K + N_b + 1)$ rows of C. If (11) is satisfied, then $(\boldsymbol{f}^{(s)}, \boldsymbol{b}^{(s)}) \in \boldsymbol{\Theta}(n, \gamma)$.

2) *Necessity:* If $\boldsymbol{\Theta}(n, \gamma)$ is nonempty, then

$$\sigma_a^2 \|(C\boldsymbol{f}^{(n)} - P\boldsymbol{b}^{(n)} - e_0)\|_2^2 + \gamma_\nu^2 \|\boldsymbol{f}^{(n)}\|_2^2 \leq \gamma^2 \tag{12}$$

where σ_a is the variance of the symbols in the constellation $\mathcal{A}$, $\boldsymbol{f}^{(n)} = (C^H(C - PH) + (\gamma_\nu^2/\sigma_a^2)I)^{-1}C^H e_0$, and $\boldsymbol{b}^{(n)} = H\boldsymbol{f}^{(n)}$, where H is as before.

Remark 1: Similar conditions have been derived for a linear equalizer in [11]. It can be shown that the conditions derived here apply to a linear equalizer as a special case since the linear equalizer is obtained by setting the feedback weights of the DFE to zero. The equivalent conditions for a linear equalizer can be obtained by setting $P = 0$. Moreover, the necessary condition (12) is an improvement over a similar result in [11] when linear equalization is considered.

Remark 2: If the sufficient condition (11) is met, then $(\boldsymbol{f}^{(s)}, \boldsymbol{b}^{(s)})$ is a member of the feasibility set, and hence, it is a valid SM-DFE filter. The structure of this filter is very similar to that of the MMSE-DFE [12] in that the feedback filter cancels part of the combined *post-cursor intersymbol interference (ISI)* effects [12] of the channel and the feedforward filter. The best cancellation is achieved by setting $N_b = K$. For the remaining part of the discussion, we shall assume $N_b = K$.

It turns out that the MMSE-DFE is a member of the feasible SM-DFE set of filters under some conditions. If the sufficient

condition (11) holds with $\gamma_\nu = \gamma_{\mathrm{mmse}}$, where

$$\gamma_{\mathrm{mmse}} \triangleq \frac{\gamma_{\overline{a}}\sigma_\nu}{\sigma_a}$$

with σ_ν^2 being the variance of the additive (Gaussian) noise, then $(\boldsymbol{f}^{(s)}, \boldsymbol{b}^{(s)})$ is the MMSE-DFE [14] and belongs to the feasibility set.

Remark 3: The above result implies that if the sufficient condition (11) holds, then the MMSE-DFE assures that the equalized outputs lie within a γ radius of the true transmitted symbol whenever the squared magnitude of each component of the channel noise vector is less than γ_{mmse}^2. Further, if (11) holds with $\gamma = d_{\min}/2$, then the MMSE-DFE makes no errors in decoding whenever the data comes from the design space.

C. SM-DFE Design Methodology and an Example

Designing an SM-DFE filter that results in error-free equalization for a subset of the data sets would require testing for condition (11) to hold with $\gamma = d_{\min}/2$. The LHS of this equation depends only on the parameters γ_ν and N_f and is denoted by $g(\gamma_\nu, N_f)$. There are a set of values of these parameters that satisfy the condition $g(\gamma_\nu, N_f) = \gamma$. Within this set, we would like to choose a value for the noise bound (γ_ν) and order of the filter (proportional to N_f), which results in the *largest* design space $\mathcal{D}$. It is shown that this requirement can be cast as the optimization of a single objective function. This goal is to maximize the probability that the data comes from the design space. We have

$$\begin{aligned}
\mathrm{Prob}&((\boldsymbol{a}, \boldsymbol{v}) \in \mathcal{D}) \\
&= \mathrm{Prob}(\boldsymbol{a} \in \mathcal{A}^{2K+1}) \cdot \mathrm{Prob}(\boldsymbol{v} \in \mathcal{V}^{2N_f+1}) \\
&= [\mathrm{Prob}(|\nu|^2 \leq \gamma_\nu^2)]^{(2N_f+1)}.
\end{aligned} \tag{13}$$

The random variable $|\nu|^2 = \nu_R^2 + \nu_I^2$ is Rayleigh distributed since the real and imaginary parts of the noise component, given by ν_R and ν_I, respectively, are independent Gaussian random variables [15]. The constrained optimization problem is

$$\begin{aligned}
\text{Maximize} \quad &\mathrm{Prob}((\boldsymbol{a}, \boldsymbol{v}) \in \mathcal{D}) \\
&= (1 - e^{-(\gamma_\nu^2/\sigma_\nu^2)})^{(2N_f+1)} \\
&\qquad \text{with respect to } (\gamma_\nu, N_f) \\
\text{Subject to:} \quad &g(\gamma_\nu, N_f) \\
&= \gamma_{\overline{a}}\|(C\boldsymbol{f}^{(s)} - P\boldsymbol{b}^{(s)} - \boldsymbol{e}_0)\|_1 \\
&\quad + \gamma_\nu\|\boldsymbol{f}^{(s)}\|_1 = \gamma.
\end{aligned}$$

In general, it is not easy to obtain an analytical solution to this problem. In our case, since the maximum order of the filter is limited by the computational constraint on the designer, a simple exhaustive search procedure can be used. Practically, the maximum value of N_f can be assumed to be of the order of half the channel impulse response length (equal to 9 in the example used below). Moreover, it can be shown that the function $g(\gamma_\nu, N_f)$ is approximately linear with respect to γ_ν for values in the range of interest. For each value of N_f, there is only a single γ_ν that meets the constraint. This can
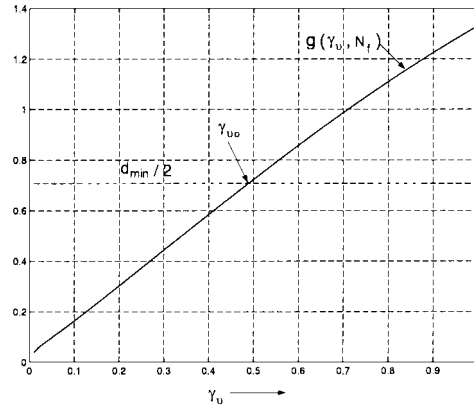

Fig. 4. Plot of sufficiency condition versus γ_ν and $N_f = 2$ for design of SM-DFE with $\gamma = d_{min}/2$ for a 20-tap microwave radio channel with QPSK signaling.

be seen easily from Fig. 4, where $N_f = 2$. A practical design methodology using these ideas is as follows.

1) Given a certain constellation $\mathcal{A}$, set $\gamma = d_{\min}/2$.
2) Check if $g(0, 1) \leq \gamma$. If not, increase the value of γ until the condition is met. Fix this γ to be the specified error bound.
3) For each $N_f \in \{1, 2, \cdots, N_{\max}\}$, find $\gamma_{\nu o}$ such that $g(\gamma_{\nu o}, N_f) = \gamma$ with associated cost $(1 - e^{-(\gamma_{\nu o}^2/\sigma_\nu^2)})^{(2N_f+1)}$.
4) Find the value of N_f for which this cost is maximum.

Intuitively, we might expect γ_ν to be as large and N_f to be as small as possible. Smaller order filters are also highly desirable from a computational point of view.

To illustrate this method, consider a microwave radio channel obtained from actual field measurements [16]. A symbol-spaced equalizer is used and also the channel impulse response is truncated to 20 significant coefficients (i.e., $D = 9$). A QPSK signaling is assumed, resulting in $\gamma_{\overline{a}} = 1$ and $d_{\min} = \sqrt{2}$. The value of γ was set equal to $d_{\min}/2$. The signal-to-noise ratio (SNR) at the receiver input was taken to be 20 dB. The design procedure described above is followed with $N_{\max} = D = 9$. The optimal values were found to be

$$\gamma_{\nu o} = 0.49 \quad N_f = 2 \quad N_b = 11.$$

For this SNR of 20 dB, the probability that the data comes from the design space was 0.999 999 982, leading to an upper bound on the probability of error of 1.8×10^{-8}, which is consistent with the results shown in Fig. 5. The probability of error curves for different values of SNR (using the appropriate optimal values of $\gamma_{\nu o}$ and N_f) are plotted for both the SM and MMSE-DFE's in Fig. 5. It was observed that the performance of the SM-DFE was very similar, and almost identical, to that of the MMSE-DFE in the range of 0–20 dB of SNR.

In many practical situations, however, the design space $(\mathcal{D})$ is not known *a priori* or might be time varying. For example, in most wireless communication systems, the channel characteristics cannot be assumed to be known *a priori*, especially due to time variations. Adaptive equalization is then required to mitigate the effects of ISI. Two approaches to formulating a recursive (adaptive) algorithm for the general
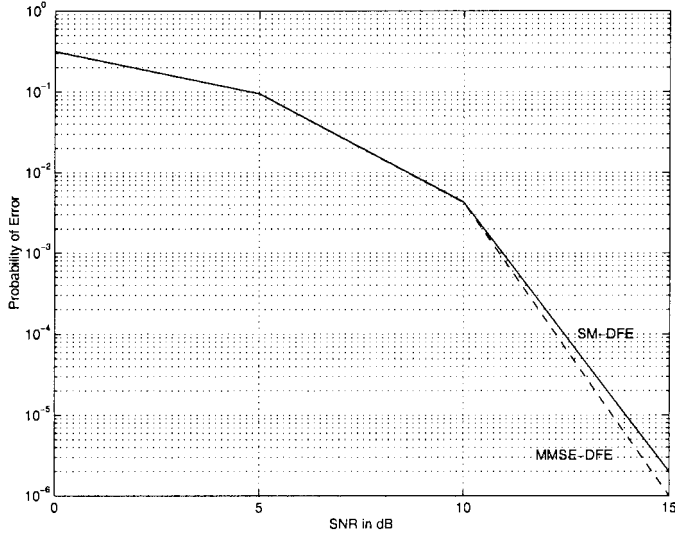
Fig. 5. Plot of probability of error for SM-DFE versus signal-to-noise ratio (SNR).

Fig. 6. OBE procedure.

SMF problem that estimate a member of the feasibility set are derived Sections III and IV. It is shown that both approaches yield the same estimates on a pointwise basis. The first approach is based on the concept of optimal bounding ellipsoid (or OBE) algorithms [9], which gives a set, as well as a point estimate. The second approach employs a least-squares-like strategy, which is shown to give the same point estimates as the former approach.

III. BEACON: A RECURSIVE ELLIPSOIDAL BOUNDING APPROACH

An OBE algorithm called BEACON is presented first. Explicit rules for updating as well as assignments to the weighting sequence are derived. A number of results on the point estimates obtained by BEACON are established that motivate and lead to the least-squares formulation discussed in Section IV.

A. Recursions and Updating Mechanism

One of the important considerations in developing real-time estimation methods is that of keeping computational expenses low without compromising significantly on performance (measured in terms of, e.g., convergence speed and asymptotic estimation error). The family of OBE algorithms, developed for SMI, use outer bounding ellipsoids to capture the *true parameter*. In the case of SMF, however, we do not necessarily need to assume the existence of a true system model (that is, linear-in-parameters with bounded output noise). However, it is straightforward to show that OBE algorithms can be used to approximate the feasibility set in the context of SMF.

For developing the recursions, the performance specification is imposed on the real-time input-desired output pairs. Thus, the filter model takes the form

$$d_i = \theta^T x_i + e_i \quad \forall i \tag{14}$$

where $d_i \in \mathbb{C}$ is the desired filter output, $e_i \in \mathbb{C}$ is the error, $x_i \in \mathbb{C}^n$ is the known input sequence, $(x_i, d_i) \in \mathcal{D}$, and

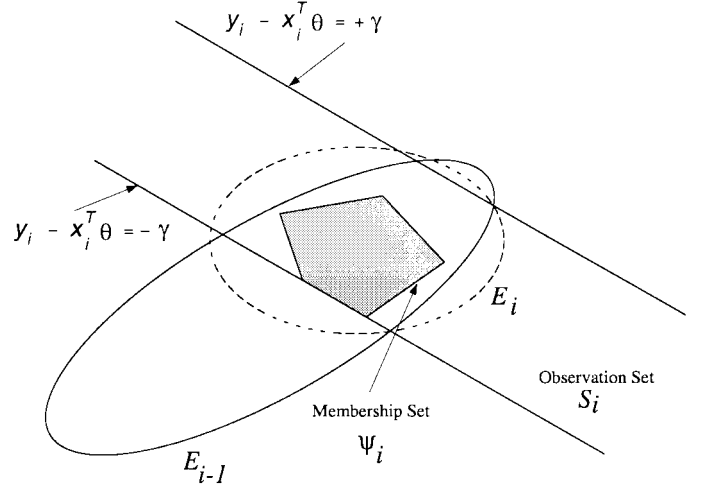$\theta \in \mathbb{C}^n$ is an estimate of a point in $\Theta(n, \gamma)$. SMF requires that θ be such that

$$|e_i|^2 \leq \gamma^2 \quad \forall i = 1, 2, \cdots$$

for the given error bound $\gamma > 0$.

The above equation combined with (14) results in a set of all the possible parameters at each instant i that are consistent with the data observed at that time. Obviously, $\Theta(n, \gamma)$ lies in this so-called *observation-induced set* (which is denoted by $\mathcal{S}_i$) and given by

$$\mathcal{S}_i = \{\theta \in \mathbb{C}^n \colon |d_i - \theta^T x_i|^2 \leq \gamma^2\}.$$

The above equation describes a degenerate ellipsoid in the parameter space. Given observations $(x_k, d_k)_{k=1}^i$, we define the *membership set* Ψ_i as the intersection over time, up to and including the present instant, of all the observation-induced sets. That is, $\Psi_i = \cap_{k=1}^i \mathcal{S}_k$. Since $\Theta(n, \gamma) \subset \mathcal{S}_k$ for all $k = 1, 2, \cdots, i$, it follows that $\Theta(n, \gamma) \subset \Psi_i$ for all i.

The basic idea of OBE algorithms is to outer bound the membership set at each instant by a mathematically tractable ellipsoid $\mathcal{E}_i$. Specifically, given an ellipsoid $\mathcal{E}_{i-1}$ that outer bounds Ψ_{i-1}, we obtain an ellipsoid $\mathcal{E}_i$ such that

$$\mathcal{E}_i \supset (\mathcal{E}_{i-1} \cap \mathcal{S}_i) \supset \Psi_i \quad \forall i.$$

This process is depicted in Fig. 6. Given an initial ellipsoid $\mathcal{E}_o = \{\theta \in \mathbb{C}^n \colon (\theta - \hat{\theta}_o)^H P_o^{-1}(\theta - \hat{\theta}_o) \leq \sigma_o\}$ with some properly chosen initial parameter estimate $\hat{\theta}_o$ and $P_o = \mu I (\mu > 0), \sigma_o > 0$ such that $\Theta(n, \gamma) \subset \mathcal{E}_o$, the algorithm establishes a recursive procedure for computing the sequence of ellipsoids $\{\mathcal{E}_i\}$. The feasibility set $\Theta(n, \gamma)$ lies in $\mathcal{E}_i$, where $\mathcal{E}_i$ is obtained as an outer bounding ellipsoid of the intersection of $\mathcal{S}_i$ with $\mathcal{E}_{i-1}$. The recursions for the BEACON algorithm are presented in the result below [17].

Proposition 2—The Update Equations: Given the observation-induced set $\mathcal{S}_i$ and the ellipsoid $\mathcal{E}_{i-1}$, the feasibility set

$\Theta(n, \gamma)$ lies in $\mathcal{E}_i$, where

$$
\begin{aligned}
\mathcal{E}_i &\triangleq \{\theta \colon (\theta - \hat{\theta}_{i-1})^H P_{i-1}^{-1} (\theta - \hat{\theta}_{i-1}) + \lambda_i | d_i \\
&\quad - \theta^T \boldsymbol{x}_i |^2 \leq \sigma_{i-1} + \lambda_i \gamma^2 \} \\
&= \{\theta \colon (\theta - \hat{\theta}_i)^H P_i^{-1} (\theta - \hat{\theta}_i) \leq \sigma_i \}
\end{aligned}
$$

$$
P_i^{-1} = P_{i-1}^{-1} + \lambda_i \boldsymbol{x}_i^* \boldsymbol{x}_i^T \tag{15}
$$

$$
\hat{\theta}_i = \hat{\theta}_{i-1} + \lambda_i P_i \boldsymbol{x}_i^* \delta_i \tag{16}
$$

$$
\sigma_i = \sigma_{i-1} - \frac{\lambda_i \delta_i^2}{1 + \lambda_i G_i} + \lambda_i \gamma^2 \tag{17}
$$

for any $\lambda_i \geq 0$, where the *prediction error* is $\delta_i = d_i - \hat{\theta}_{i-1}^T \boldsymbol{x}_i$, and $G_i = \boldsymbol{x}_i^T P_{i-1} \boldsymbol{x}_i^*$.

The proof is on similar lines as in [4]. For a properly chosen γ, P_o, and $\sigma_o > 0$, we have $\sigma_i > 0$ for all i. The centroid of the ellipsoid at each instant $\hat{\theta}_i$ can be considered to be a point estimate at that instant if need be. In the actual implementation of the algorithm, the matrix P_i, rather than its inverse, is updated. That is

$$
P_i = P_{i-1} - \frac{\lambda_i P_{i-1} \boldsymbol{x}_i^* \boldsymbol{x}_i^T P_{i-1}}{1 + \lambda_i G_i}.
$$

This update can be derived easily from the matrix inversion lemma [4]. Each value of λ_i yields a different bounding ellipsoid. To compute λ_i, we adopt a measure of optimality similar to the one used in DH-OBE [4]. The difference between DH-OBE and this approach in terms of formulation is the manner in which the previous ellipsoid is combined with the data constraint set at the current instant. This difference results in a simpler update checking rule and assignment to the time-varying weights in our algorithm. The objective, therefore, is to minimize σ_i with respect to λ_i under the constraint that $\lambda_i \geq 0$. When the minimum of σ_i occurs at $\lambda_i = 0$, it results in no update [17]. The *discerning update* capability of the proposed algorithm follows as a result.

The optimal value of λ_i, denoted by λ_i^o, is obtained by maximizing the following function:

$$
\lambda_i \left\{ \frac{\delta_i^2}{\gamma^2} \left(\frac{1}{1 + \lambda_i G_i} \right) - 1 \right\}.
$$

This yields

$$
\lambda_i^o = \begin{cases} 0, & \text{if } |\delta_i| \leq \gamma \\ \frac{1}{G_i} \left(\frac{|\delta_i|}{\gamma} - 1 \right), & \text{if } |\delta_i| > \gamma. \end{cases} \tag{18}
$$

It has been noted in [5] that σ_i, as a measure of optimality in OBE algorithms, is not a physically interpretable measure of the size of the ellipsoid. The minimum volume and minimum trace criteria proposed by Fogel and Huang [9], which was also employed in other OBE algorithms [7], [10], are easily interpretable optimality measures since they have direct relation to some notion of the geometrical size of the ellipsoid. The volume of the ellipsoid $\mathcal{E}_i$ is proportional to $\det\{\sigma_i P_i\}$, and the sum of squares of the semi-axes is given by $\mathrm{trace}\{\sigma_i P_i\}$. The justification for σ_i minimization has been that it is a natural bound on the Lyapunov function in the convergence analysis in [4] and that it bears relations to the volume and trace criteria asymptotically [18]. Although the

OBE algorithms derived using the Fogel–Huang criterion also update selectively, this does not result in any computational savings. This is so because the test for checking whether an update is needed or not is itself an $\mathcal{O}(n^2)$ operation, which is the same as the computation required in carrying out the actual update. The σ_i minimization strategy leads to an $\mathcal{O}(n)$ check for updating. With BEACON's sparse updating property, this implies significant computational savings. It is further shown in this paper that by minimizing σ_i, BEACON yields sequences of volume and trace measures of the ellipsoids that are monotone nonincreasing. This feature is not shared by DH-OBE, which is another σ_i minimizing OBE algorithm [4]. Therefore, the size of the ellipsoids generated by BEACON decreases with time, even though the decrease may not be optimal at every time step. For the next result and much of the subsequent analysis, the following definitions are needed.

Definition 1: The *bounding hyperplanes* to the observation-induced set $\mathcal{S}_i$ are denoted by $\mathcal{S}_i^\phi$, where $\phi \in [0, 2\pi)$, and given by

$$
\mathcal{S}_i^\phi = \{\theta \in \mathbb{C}^n \colon d_i - \theta^T \boldsymbol{x}_i = \gamma e^{j\phi} \} \quad \forall \phi \in [0, 2\pi). \tag{19}
$$

In particular, if $\theta \in \mathbb{R}^n$, then there are only two bounding hyperplanes given by $\mathcal{S}_i^{(\phi=0)}$ and $\mathcal{S}_i^{(\phi=\pi)}$. Moreover, the *interior* of $\mathcal{S}_i$ is

$$
\mathcal{S}_i^{in} = \{\theta \in \mathbb{C}^n \colon |d_i - \theta^T \boldsymbol{x}_i| < \gamma \}.
$$

From the above construction, we have $\mathcal{S}_i = \cup_\phi \mathcal{S}_i^\phi \cup \mathcal{S}_i^{in}$.

Definition 2: The *nearest bounding hyperplane* to a vector, say $\zeta \notin \mathcal{S}_i$ at time instant i denoted by $\mathcal{NB}_i(\zeta)$, is

$$
\mathcal{NB}_i(\zeta) \triangleq \mathcal{S}_i^{\phi_o}, \quad \phi_o = \angle(d_i - \zeta^T \boldsymbol{x}_i). \tag{20}
$$

By the assignment to λ_i, the following result for the *a posteriori* error $\epsilon_i \triangleq d_i - \hat{\theta}_i^T \boldsymbol{x}_i$ can be shown, from which we obtain a geometrical interpretation to the update rule.

Proposition 3: The *a posteriori* error given by ϵ_i is always less than or equal to the noise bound γ in magnitude. That is

$$
|\epsilon_i| = \begin{cases} \gamma & (< |\delta_i|), & \text{if } i \text{ is an updating instant} \\ |\delta_i| & (\leq \gamma), & \text{if } i \text{ is not an updating instant.} \end{cases}
$$

Geometrically, the update rule can be interpreted as—*No update is needed if $\hat{\theta}_{i-1} \in \mathcal{S}_i$, else, update $\hat{\theta}_{i-1}$ in the direction of the vector $P_{i-1} \boldsymbol{x}_i^*$ such that $\hat{\theta}_i \in \mathcal{NB}_i(\hat{\theta}_{i-1})$.*

B. Effect of Initial Conditions and Convergence Issues

The choice of the initial conditions, specifically that of σ_o and μ (where $P_o = \mu I$), is treated in this section. The motivation for the following analysis is two fold.

1) It has been observed that various choices of σ_o lead to significantly different performance in DH-OBE
2) A problem that may occur in tracking time-varying parameters is that σ_i may become negative for some i.

By a certain result pertaining to robustness of the estimates with respect to choice of initial conditions, we can effectively explain the good tracking properties of the BEACON algorithm.

In the absence of any *a priori* knowledge regarding the parameter, the origin $(\hat{\theta}_o = 0)$ forms the initial estimate of $\Theta(n, \gamma)$. Theoretically, the condition to be satisfied while choosing σ_o and μ is

$$\sup_{\theta \in \Theta(n,\gamma)} \|\theta\|_2^2 \leq \mu \sigma_o. \tag{21}$$

Since this cannot be implemented in practice, a *large enough* σ_o and μ are picked to start the algorithm. First, the effect of any arbitrary scaling on σ_o, keeping the product $\mu\sigma_o$ constant such that (21) is satisfied, is investigated.

Proposition 4: Consider two realizations of the algorithm with different initial conditions: 1) an initial condition $\sigma_o^{(1)}, \mu^{(1)}$ and 2) $\sigma_o^{(2)} = K\sigma_o^{(1)}, \mu^{(2)} = \mu^{(1)}/K$, for any $K > 0$. Then, for all i, the following are true.

- The sequence of updating instants are the same.
- $\lambda_i^{o(2)} = K\lambda_i^{o(1)}, P_i^{(2)} = P_i^{(1)}/K$, and $\sigma_i^{(2)} = K\sigma_i^{(1)}$.
- Consequently, $\hat{\theta}_i^{(1)} = \hat{\theta}_i^{(2)}$. In addition, $\mathcal{E}_i^{(1)}$ and $\mathcal{E}_i^{(2)}$ are identical.

Remark 4: For any volume or trace minimizing the OBE algorithm, these results are trivially true. In σ_i minimizing OBE algorithms, however, it is not essential that the performance be invariant to such scaling. In fact, it has been observed that none of the above results are true for DH-OBE.

It is possible to generalize the above result to the potentially catastrophic case of a wrong choice of σ_o and μ such that (21) is not satisfied. In such a scenario, the basic assumption that $\mathcal{E}_o$ is a bounding ellipsoid fails, leading to what we term as a *model violation*. Extending the above result shows that the BEACON algorithm, given by (15)–(18), is robust with respect to such model violations [17]. It is important to note that other types of deviations in the assumed system model, in the form of, e.g., decision errors in a DFE, are not addressed in this work. We address the case when the initial ellipsoid fails to outer bound the feasibility set, leading to a violation in the assumed model that the OBE algorithms are based on. Specifically, we hve the following.

For a fixed μ, the sequence of updating instants, weights (λ_i^o), and parameter estimates $(\{\hat{\theta}_i\} \forall i)$ are the same for all $\sigma_o \in \mathbb{R}$.

Therefore, the point estimates are not affected, even if $\sigma_o < 0$. Under model violations, the algorithm no longer operates according to bounding ellipsoid principles. The result suggests that the point estimate given by (15)–(18) might have a more general interpretation. An alternative solution method to the SMF problem, resulting from an explicit least-squares-like cost function optimization, is addressed in Section IV.

Remark 5: It is pertinent to note that although the central (point) estimate $\hat{\theta}_i$ is independent of the the value of σ_i, the ellipsoidal set estimate $(\mathcal{E}_i)$ will change if σ_o is varied with μ held constant.

Convergence issues in OBE algorithms for set-membership identification were addressed by Dasgupta and Huang in [4] and Nayeri *et al.* in [19], which established convergence

results for most known OBE algorithms along with conditions to guarantee point convergence to the true parameter in a stochastic setting. In this paper, we analyze some convergence properties for the BEACON algorithm in the SMF context and prove asymptotic convergence of the prediction error even under model violations. First, it is shown that if the estimate converges, then in the limit as $i \to \infty$, the estimate is a point on the *boundary* of the feasibility set that is denoted by $\Pi(n, \gamma)$. That is, $\theta \in \Pi(n, \gamma)$ if $\theta \in \Theta(n, \gamma)$, and $|d^o - \theta^T x^o| = \gamma$ for some $(x^o, d^o) \in \mathcal{D}$.

Proposition 5: Consider the sequence of estimates $\hat{\theta}_i$.

- If there exists some $N > 0$ such that $\hat{\theta}_N \in \Pi(n, \gamma)$, then $\lambda_{N+l}^o = 0$, i.e., $\hat{\theta}_{N+l} = \hat{\theta}_N \forall l \geq 0$.
- Assume that $\hat{\theta}_o \notin \Theta(n, \gamma)$ and $\lim_{i \to \infty} \hat{\theta}_i = \hat{\theta}_\infty \in \Theta(n, \gamma)$. Then, $\hat{\theta}_\infty \in \Pi(n, \gamma)$.
- Let x_i be bounded, that is, $0 \leq \|x_i\|_2^2 \leq \overline{X}$ for all i. Then, for any σ_o, $\limsup_{i \to \infty} |\delta_i| \leq \gamma$.

In the special case of system identification with bounded output noise (i.e., SMI), it is shown that using an overestimated value for the noise bound in the BEACON recursions leads to a biased estimate, assuming the estimates converge. As a degree of confidence, convergence of the estimate to a region around the true parameter is also shown. The SMI model assumes that the output noise in the system to be identified is bounded by some $\gamma_{\nu,\text{true}}$. That is, there exists θ_{true} and $\gamma_{\nu,\text{true}}$ such that

$$d_i = \theta_{\text{true}}^T x_i + \nu_i, \quad |\nu_i| \leq \gamma_{\nu,\text{true}} \quad \forall i. \tag{22}$$

Proposition 6: Choose some $\gamma = \gamma_{\nu,\text{true}} + \varepsilon, \varepsilon > 0$ in the OBE recursions. Further, assume the sequence of estimates converge to some $\hat{\theta}_\infty$. Then, the following are true.

- $\hat{\theta}_\infty \neq \theta_{\text{true}}$, where θ_{true} is the true parameter.
- $\limsup_{i \to \infty} \|\hat{\theta}_{i-1} - \theta_{\text{true}}\|_2^2 \leq (\gamma^2 - \sigma_\nu^2)/\sigma_x^2$, where σ_ν^2 is the noise variance, and σ_x^2 is the minimum positive singular value of R_{xx}, which is the autocorrelation matrix of the (stationary) input.

C. Simulation Studies

Some simulation studies using the BEACON algorithm are described, and a comparison with the traditional least-squares algorithm like RLS is made. Superior tracking capabilities of the BEACON algorithm for time-varying systems is demonstrated via an example. Simulation results for the following cases are presented:

- identification of an eigth-order FIR filter under uniformly distributed noise;
- estimation of an actual *microwave radio communications channel* with truncated Gaussian noise;
- tracking of a fast time-varying second-order FIR filter.

Minute details of the simulation are omitted for the sake of brevity. The results are shown in Figs. 7–10. For all cases considered, the SNR was 15 dB, and the input was taken from a binary alphabet.

The performance of the BEACON algorithm is similar to RLS for Case 1. Importantly, the number of updates was around 30 out of the total of 1000 samples. It is pertinent
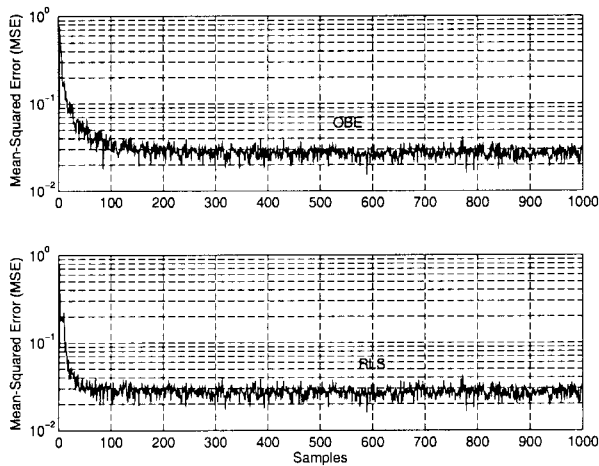
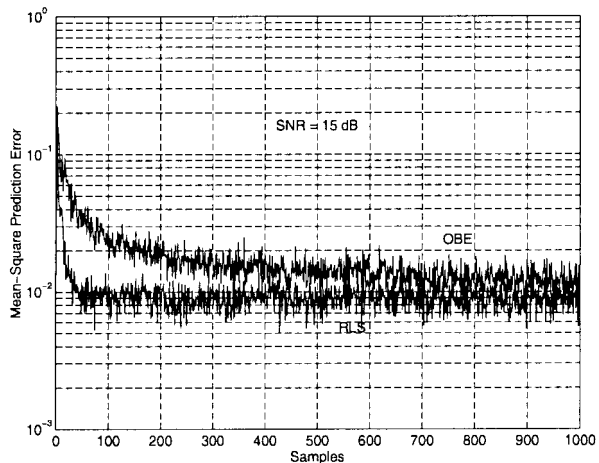Fig. 7. Plot of mean squared prediction error for estimation of an eight-tap FIR filter with uniform noise.

Fig. 8. Estimation of a microwave radio channel.

Fig. 9. Tracking performance comparison.

Fig. 10. Two-tap FIR filter—evolution of the ellipsoids and the central estimate through time, k (the x- and y-axes represent the two components of the parameter vector).

to note that such performance is obtained even though the algorithm uses very few inputs for updating.

For Case 2, it can be observed that the rate of convergence of BEACON is slightly slower than RLS, although the final MSE is comparable in spite of sparse updates. The selective updating feature is one of the main reasons for the somewhat slower convergence. The value of γ can be varied to trade off between achievable performance and computational complexity. A higher value of γ would result in fewer updates and, consequently, a higher MSE (since the algorithm essentially stops taking inputs for updating once the prediction error goes below γ in magnitude). On the other hand, a lower γ, leading to a more stringent error specification, would imply a lower MSE at the expense of an increased number of updates. When multiple filters need to be estimated simultaneously, the selective update feature can be used to share a small number of updating processors among many filters, leading to a saving in hardware requirements [11]. The BEACON algorithm is an ideal candidate for such sharing as its frequency of updating is one of the least among the known OBE algorithms.

In Case 3, for tracking time-varying parameters, identification of a second-order FIR filter was simulated similar to
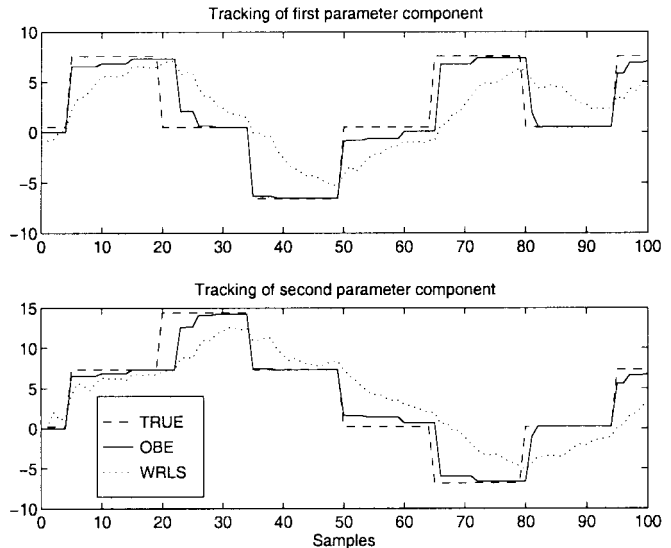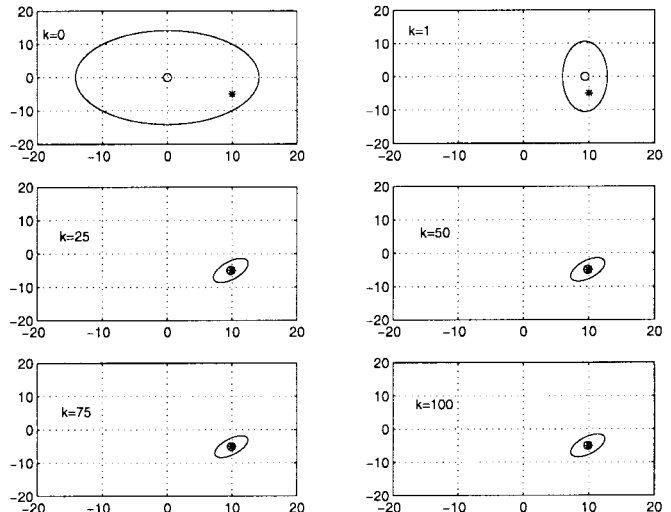
Case 1, but the coefficients varied randomly, thus making it a time-varying system. The time variations in the parameters are introduced by having random, but bounded, jumps at every 15 samples. It can be seen from Fig. 9 that the BEACON algorithm outperforms weighted RLS (with an exponential forgetting factor of 0.9) in terms of tracking while using only around 25 data points out of 100 for updating. This superior tracking capability is an outcome of the data-dependent choice of weights as well as the selective updating mechanism. In comparison with RLS, the sparse updating by BEACON results in a slight loss in convergence speed in a stationary environment. On the other hand, in the presence of time variations, it is this selective updating capability along with the dependence of the weights on the prediction error that result in the excellent tracking characteristics of BEACON.

Finally, Fig. 10 is a plot of the evolution of the ellipsoids through time for the case of identifying a second-order FIR

filter under uniform noise. The "$*$" denotes θ_{true}, whereas "o" denotes $\hat{\theta}_i$. It can be seen that in just 25 samples, the central estimate seems to converge to the true parameter and that the associated ellipsoid shrunk to a much smaller size, albeit not yet a point. This can be explained from the fact that the centroid of the BEACON algorithm can also be looked upon as the solution of a certain least-squares like problem under the constraint of bounded errors. The robustness to model violations of the centroid also indicates the existence of an alternative estimation approach, which leads to the same recursions as (15)–(17) that is not hinged upon the choice of σ_o. The next section is devoted to the formulation of such a methodology.

IV. LEAST-SQUARES FILTERING UNDER BOUNDED ERRORS

The idea here is to formulate a least-squares cost function, given the knowledge that the errors are bounded but otherwise unknown. The problem of least-squares estimation with a dead zone has been treated in [20]. We derive the least-squares method as a constrained optimization problem, and the resulting weights are shown to be optimal in the sense of satisfying the constraint at every instant. Similar algorithms have been derived in [21] for the problem of model validation, where the motivation is to develop least-squares estimates that (asymptotically or otherwise) lie in the membership set. Optimality issues in identification for systems with bounded constraints on the error sequences are addressed in [22].

A. Problem Formulation

Assume that at time $i-1$, a certain quadratic cost function is constructed, given information up to and including time instant $i-1$, that is, $(\boldsymbol{x}_k, d_k)_{k=1}^{i-1}$. Denote this function by $V_{i-1}(\theta)$, which is defined as

$$
\begin{aligned}
V_{i-1}(\theta) = & \{(\theta - \hat{\theta}_o)^H P_o^{-1}(\theta - \hat{\theta}_o) - \sigma_o\} \\
& + \sum_{k=1}^{i-1} \alpha_k \{|y_k - \theta^T \boldsymbol{x}_k|^2 - \gamma^2\} \\
= & (\theta - \hat{\theta}_{i-1})^H P_{i-1}^{-1}(\theta - \hat{\theta}_{i-1}) - \sigma_{i-1} \quad (23)
\end{aligned}
$$

for some appropriately chosen weighting sequence $\alpha_k \geq 0, k = 1, 2, \cdots, i-1$. The choice of the time-varying and (possibly) data-dependent weights will be explained shortly. Note that the additive constants appearing in the cost function do not affect the optimization procedure. They only facilitate connections to the OBE derivation. The first term in (23) represents the confidence on the initial guess.

Consider the data set at time i given by the pair $(\boldsymbol{x}_i, d_i) \in \mathcal{D}$. The information contained by the new data, due to the requirement of the bounded errors, is characterized by $|d_i - \theta^T \boldsymbol{x}_i|^2 \leq \gamma^2 \Leftrightarrow \theta \in \mathcal{S}_i$. The basic idea of deriving this least-squares estimator is the following. *Given the cost function,* $V_{i-1}(\theta)$ *and the observation-induced set* $\mathcal{S}_i$, *find the new estimate, say,* $\hat{\theta}_i$, *that minimizes* $V_{i-1}(\theta)$ *subject to* $\hat{\theta}_i \in \mathcal{S}_i$. In other words, the strategy is to move from $\hat{\theta}_{i-1}$ to a new estimate to incur the least increase in the cost function $V_{i-1}(\theta)$ under the constraint that the new estimate has to lie in the observation-induced set $\mathcal{S}_i$.
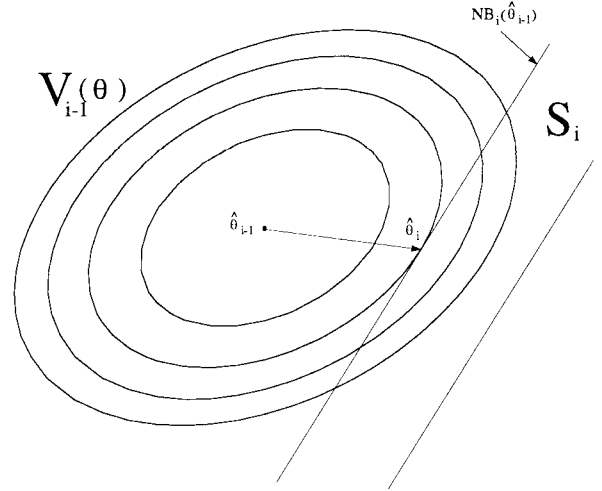

Fig. 11. Contour plot of $V_{i-1}(\theta)$ along with geometric representation of updating process.

Remark 3: When $\hat{\theta}_{i-1} \in \mathcal{S}_i$, there is no need to move to a new estimate as $\hat{\theta}_{i-1}$ itself is the global minimum of $V_{i-1}(\theta)$. This results in *no update* of the parameter estimate. Moreover, the condition for no update is exactly the same as for BEACON, given by Proposition 3. Therefore, the discerning update rule for this least-squares-type problem is established and shown to coincide with the condition in BEACON.

This estimation strategy is cast as a constrained optimization problem. At each instant i

$$
\text{Minimize} \quad V_{i-1}(\theta) = (\theta - \hat{\theta}_{i-1})^H P_{i-1}^{-1}(\theta - \hat{\theta}_{i-1}) - \sigma_{i-1} \quad (24)
$$

$$
\text{Subject to:} \quad |d_i - \theta^T \boldsymbol{x}_i|^2 \leq \gamma^2. \quad (25)
$$

The parameter vector minimizing this problem is the new estimate, which is denoted by $\hat{\theta}_i$.

Henceforth, assume that $\hat{\theta}_{i-1} \notin \mathcal{S}_i$ because if it did, we have shown that there is no update and that $\hat{\theta}_{i-1}$ is indeed the solution. Since the cost function is convex and the constraint set is also a convex set, it is easy to establish that the solution to the problem (24) and (25), which is denoted by $\hat{\theta}_i$, belongs to $\mathcal{NB}_i(\hat{\theta}_{i-1})$. This, in turn, is given by $\mathcal{S}_i^{\phi_o}$, where ϕ_o is the phase of the prediction error δ_i (see Definition 2). Therefore, the phase of the *a posteriori* error is the same as the phase of the prediction error. Recall that the above is identical to the result obtained in Proposition 3, which was derived using OBE methods. With this, the problem constraint becomes an equality, rather than an inequality, as in (25).

B. The Algebraic Solution

Construct an auxiliary function

$$
J_i(\theta, \kappa_i) = V_{i-1}(\theta) + \kappa_i(|d_i - \theta^T \boldsymbol{x}_i|^2 - \gamma^2)
$$

where κ_i is the Lagrangian multiplier. The minimizer is the parameter at which the normal vectors to the surfaces of $V_{i-1}(\theta)$, and the constraint function are colinear. This is depicted in Fig. 11.

Proposition 7: Given data set $(\boldsymbol{x}_i, d_i)$ and the cost function $V_{i-1}(\theta)$ at time i, the solution to the optimization problem posed in (24) and (25) is given by

$$\hat{\theta}_i = \hat{\theta}_{i-1} + \frac{\kappa_i^o P_{i-1} \boldsymbol{x}_i^* \delta_i}{1 + \kappa_i^o G_i}$$
$$\kappa_i^o = \frac{1}{G_i} \left(\frac{|\delta_i|}{\gamma} - 1 \right) \qquad (26)$$

where, as before, $G_i = \boldsymbol{x}_i^T P_{i-1} \boldsymbol{x}_i^*$.

Remark 7: By the above formulation, the recursive solution for the parameter estimate, the update mechanism, and the assignment to the time-varying weights are seen to be the same as for the centroid in the BEACON algorithm. In addition, the sequence of weights in the BEACON recursions given by $\{\lambda_i^o\}$ are the the same as the Lagrangian multipliers in the constrained optimization problem ($\{\kappa_i^o\}$). It is this weighting strategy that leads to the satisfaction of the constraint at each instant. The direction of movement from $\hat{\theta}_{i-1}$ given by $P_{i-1} \boldsymbol{x}_i^*$ is essentially a compromise between reaching the observation-induced set at each instant and incurring the least increase on the error surface $V_{i-1}(\theta)$.

A geometrical solution to the same problem is presented next without resorting to the standard Lagrange multiplier technique.

C. The Geometric Solution

Transforming the parameter space by the (invertible) matrix $P_{i-1}^{-1/2}$ results in a transformation of the ellipsoid to a n-dimensional spheroid. That is, a transformation of the kind $\xi = P_{i-1}^{-1/2}\theta$, where ξ is a vector from the transformed domain. If the singular value decomposition of P_{i-1} is given by $P_{i-1} = U\Sigma U^H$ for some unitary matrix U and a diagonal matrix Σ of positive elements, then the transformation by $P_{i-1}^{-1/2}$ can be looked upon as consisting of two operations—transformation by U^H and then scaling by $\Sigma^{-1/2}$. The first results in a rotation (under norm invariance) of the axes, aligning all the semi-axes of the ellipsoid along the (new) coordinates. The next operation amounts to normalizing the lengths of the semi-axes to unity, resulting in a spheroid in the new coordinates.

The input $\boldsymbol{x}_i$ is also transformed to, say, $\boldsymbol{y}_i$, where $\boldsymbol{y}_i = P_{i-1}^{*1/2}\boldsymbol{x}_i$ such that the product $\theta^T \boldsymbol{x}_i = \xi^T \boldsymbol{y}_i$. By this, the observation-induced set $\mathcal{S}_i$ remains unchanged. In addition, $G_i = \boldsymbol{x}_i^T P_{i-1} \boldsymbol{x}_i^* = \|\boldsymbol{y}_i\|_2^2$. Now, the optimization problem is as follows: *Minimize the cost function given by*

$$V_i^o(\xi) = \|\xi - \hat{\xi}_{i-1}\|_2^2$$

under the constraint that the solution belongs to $\mathcal{S}_i$ given by $\{\xi : |d_i - \xi^T \boldsymbol{y}_i|^2 \le \gamma^2\}$. This problem statement is equivalent to (24) and (25).

We know that the optimum solution lies on the nearest bounding hyperplane to $\hat{\xi}_{i-1}$ and that the solution occurs at the point where the normal direction to $\mathcal{NB}_i(\hat{\xi}_{i-1})$ coincides with that of $V_i^o(\xi)$. Since the contours of $V_i^o(\xi)$ are spheroids, this amounts to moving along the direction of $-\boldsymbol{y}_i$ to reach the nearest bounding hyperplane.

Proposition 8: The solution to the above problem is

$$\hat{\xi}_i = \hat{\xi}_{i-1} + \frac{\boldsymbol{y}_i}{\|\boldsymbol{y}_i\|_2^2} \left(|\delta_i| - \gamma \right)$$

which results in the same recursion (after transforming back to the original coordinates) as in Proposition 7. This result can be shown as follows: Let $\hat{\xi}_i - \hat{\xi}_{i-1} = -c_i \boldsymbol{y}_i / \|\boldsymbol{y}_i\|_2$, where c_i is a constant such that $d_i - \hat{\xi}_i^T \boldsymbol{y}_i = (\delta_i / |\delta_i|)\gamma$. This implies $c_i = (1/\|\boldsymbol{y}_i\|)(1 - (\gamma/|\delta_i|))\delta_i$. Some rearrangement and transforming back to the θ space yields the above result.

We also derive an explicit recursion for the cost function at time i, namely $V_i(\theta)$, given $V_{i-1}(\theta)$ and the data pair $(\boldsymbol{x}_i, d_i)$. The property that should be possessed by this function is that it is quadratic and of the same form as $V_{i-1}(\theta)$. In addition, $\hat{\theta}_i$ should be the global minimizer of this function.

Effectively, the question is with regard to the assignment to the weight $\alpha_i \ge 0$, as in (23). Therefore

$$V_i(\theta) = V_{i-1}(\theta) + \alpha_i \{|d_i - \theta^T \boldsymbol{x}_i|^2 - \gamma^2\}.$$

It turns out that the only assignment to α_i, which is consistent with the above formulation, is given by κ_i^o from (26), and this defines the updates for the matrix P_i and the scalar σ_i. Since we know that $\kappa_i^o = \lambda_i^o$, these recursions are the same as (15) and (17).

Thus, we have formulated a least-squares type cost function at each instant and obtained a solution constrained on the need for the estimate to lie in the observation-induced set at that instant. The solution was found by the Lagrange multiplier technique and by a geometrical argument. Additional constraints imposed on the model, in the form of a bound on the error process, allowed for a discerning method to pick the weighting sequence in a recursive least-squares-like algorithm.

D. Discussions on the Optimality Measure and Tracking Properties of BEACON

In a manner independent of one another, the OBE derivation was shown to coincide with the constrained least-squares derivation. Note that the set of all the parameters that result in $V_{i-1}(\theta) \le 0$ describe an n-dimensional ellipsoid in the parameter space. Moreover, the non-negative quantity σ_i is the depth of the point of minimum of $V_{i-1}(\theta)$ (which occurs at $\hat{\theta}_{i-1}$). It can also be be proved that minimizing the cost function at each instant (24) reduces to the problem of finding an extremum of σ_i. To show this, assume now that the estimator has been partially determined by setting the gradient with respect to θ to zero. Then, recognizing the dependence on κ_i, we denote the estimate by $\theta(\kappa_i)$

$$\theta(\kappa_i) = \hat{\theta}_{i-1} + \frac{\kappa_i P_{i-1} \boldsymbol{x}_i^*}{1 + \kappa_i G_i} \delta_i.$$

From the above equation, by using the equation for $\theta(\kappa_i)$ and some tedious algebra, we obtain

$$J_i(\theta(\kappa_i), \kappa_i) = - \left(\sigma_{i-1} + \kappa_i \gamma^2 - \frac{\kappa_i \delta_i^2}{(1 + \kappa_i G_i)} \right) = -\sigma_i.$$

Therefore, the process of obtaining an extremum of $J_i(\cdot)$ with respect to κ_i, in order to satisfy the constraint (25) and find

the Lagrange multiplier, is the same as finding the extremum of σ_i with respect to κ_i. Moreover, with κ_i^o denoting the value of κ_i, which achieves this extremum, it turns out that $(d^2\sigma_i/d\kappa_i^2) > 0$ for $\kappa_i = \kappa_i^o$. This implies that κ_i^o is a minimizer of σ_i. Hence, the problem of finding the extremum of the auxiliary cost function with respect to the Lagrange multiplier is the same as finding the minimum of σ_i over all $\kappa_i \geq 0$. As pointed out earlier, the geometrical significance of σ_i is that it defines the depth of the minimum point of $V_i(\theta)$ below the θ-space.

As far as tracking is concerned, one of the singular problems encountered in OBE algorithms is the case when the intersection of $\mathcal{S}_i$ with $\mathcal{E}_{i-1}$ becomes void at some time instant i. At such instants, σ_i becomes negative, and the OBE algorithms have to resort to certain rescue procedures so that tracking is restored [6]. However, as has been shown in the earlier section, σ_i becoming negative does not affect the updating recursions (15)–(17), although the OBE interpretation breaks down. In the absence of rescue procedures, other OBE estimators tend to lose track of the parameter for abrupt and significant changes in the parameter, while such is not the case with the BEACON algorithm, which explains the good tracking characteristics observed. From the least-squares perspective, this fact is even more apparent. The only objective in that formulation is to optimally move to the observation-induced set at the present time instant.

However, if we wish the estimator to operate according to OBE principles, a scheme using ideas from set-membership state estimation can be implemented for tracking, thus giving a concrete method to circumvent the need for rescue mechanisms. For more details on the state estimation via set-membership principles, see [23] and [24].

V. CONCLUSION

This paper dealt with a filtering problem known as set-membership filtering (SMF). The formulation imposed convex set constraints on the filter by requiring the resulting error sequences to be bounded in magnitude by a designer-specified value. Results for a special case of filtering, namely, decision feedback equalization of communication channels, was studied in this framework. Its relation to the existing MMSE-DFE was explored wherein it was shown that, in certain circumstances, the MMSE-DFE belongs to the class of SM-DFE filters. Next, a recursive algorithm called BEACON was derived according to an optimal bounding ellipsoid criterion as an adaptive solution to the SMF problem. This algorithm was shown to feature a highly selective update mechanism, wherein a large percentage of the data (greater than 90%) was not used to update the parameter estimate. Moreover, a least-squares-like criterion was used to derive the same algorithm. The analysis presented allowed for connections between OBE and least-squares estimation methods. The alternative formulation also shed light on the observed insensitivity to model violations and robust performance under time-varying conditions. The ability of the BEACON algorithm to track fast time variations was demonstrated via simulations that indicate much promise in emerging wireless communication systems where high

performance and low complexity adaptive signal processing methods are imperative.

APPENDIX
PROOFS OF RESULTS

Proposition 1—Proof: "$\Rightarrow$:" The supremum of the LHS in (10) is equal to

$$\sup_{\boldsymbol{a}\in\mathcal{A}^{2K+1}} |(C\boldsymbol{f} - P\boldsymbol{b} - \boldsymbol{e}_0)^T\boldsymbol{a}| + \gamma_\nu\|\boldsymbol{f}\|_1$$

and is achieved when $|\nu_k| = \gamma_\nu$, $\angle(\nu_k) = \phi - \angle(f_k)$, where ϕ is such that $\angle(C\boldsymbol{f} - P\boldsymbol{b} - \boldsymbol{e}_0)^T\boldsymbol{a}) = \angle(\boldsymbol{f}^T\boldsymbol{v})$. An upper bound on the first term in the above expression is

$$\sup_{\boldsymbol{a}\in\mathcal{A}^{2K+1}} |(C\boldsymbol{f} - P\boldsymbol{b} - \boldsymbol{e}_0)^T\boldsymbol{a}| \leq \gamma_{\overline{a}}\|(C\boldsymbol{f} - P\boldsymbol{b} - \boldsymbol{e}_0)\|_1$$

which leads to a condition of the form (11). Since

$$\gamma_{\overline{a}}\|(C\boldsymbol{f} - P\boldsymbol{b} - \boldsymbol{e}_0)\|_1 + \gamma_\nu\|\boldsymbol{f}\|_1)^2$$
$$\leq \alpha^2[\gamma_{\overline{a}}^2\|(C\boldsymbol{f} - P\boldsymbol{b} - \boldsymbol{e}_0)\|_2^2 + \gamma_\nu^2\|\boldsymbol{f}\|_2^2]$$

because there exists a constant α such that $\|\boldsymbol{x}\|_1 \leq \|\boldsymbol{x}\|_2$. By minimizing this upper bound over all $(\boldsymbol{f}, \boldsymbol{b})$, we obtain the expression for $\boldsymbol{f}^{(s)}$ and $\boldsymbol{b}^{(s)}$.

"$\Leftarrow$:" The necessary condition is also obtained in a similar manner by noting that

$$\sup_{\boldsymbol{a}\in\mathcal{A}^{2K+1}} |(C\boldsymbol{f} - P\boldsymbol{b} - \boldsymbol{e}_0)^T\boldsymbol{a}| \geq \sigma_a\|(C\boldsymbol{f} - P\boldsymbol{b} - \boldsymbol{e}_0)\|_2$$

since $E[\boldsymbol{a}\boldsymbol{a}^H] = \sigma_a^2 I$. Further, since $\|\boldsymbol{f}\|_1^2 \geq \|\boldsymbol{f}\|_2^2$, we obtain a lower bound on (10) as

$$\sigma_a^2\|(C\boldsymbol{f} - P\boldsymbol{b} - \boldsymbol{e}_0)\|_2^2 + \gamma_\nu^2\|\boldsymbol{f}\|_2^2$$

and $(\boldsymbol{f}^{(n)}, \boldsymbol{b}^{(n)})$ are the filters that minimize this function, leading to the necessary condition (12). ∎

Proposition 3—Proof: First, if $i \notin$ updating instants, then $\hat{\theta}_i = \hat{\theta}_{i-1} \Rightarrow |\epsilon_i| = |\delta_i| \leq \gamma$. Else, from the update equation for $\hat{\theta}_i$, we have that the *a posteriori* error ϵ_i is

$$\epsilon_i = \delta_i - \frac{\lambda_i^o\boldsymbol{x}_i^T P_{i-1}\boldsymbol{x}_i^*}{1 + \lambda_i^o G_i}\delta_i = \frac{\delta_i}{|\delta_i|}\gamma$$

where the second equality comes from the optimal assignment of λ_i^o. Thus, the phase of ϵ_i is the same as that of δ_i, and $|\epsilon_i| = \gamma$, which implies that $\hat{\theta}_i \in \mathcal{NB}_i(\hat{\theta}_{i-1})$. ∎

Proposition 4—Proof: The result is proved by induction. Assume that $\hat{\theta}_o$ is the same in both the realizations. Then, all the equalities above are true at time $m = 0$. Further, let the above results be true at a time instant $m - 1, (m > 1)$. Since $\hat{\theta}_{m-1}$ is independent of the scale factor K, δ_m also does not depend on K. From (18), this implies that $\lambda_m^{o(2)} = K\lambda_m^{o(1)}$ since $P_{m-1}^{(2)} = P_{m-1}^{(1)}/K$ by assumption. This immediately leads to the fact that $P_m^{(2)} = P_m^{(1)}/K$ from (15) and from (16), and it is easy to conclude that $\hat{\theta}_m^{(2)} = \hat{\theta}_m^{(1)}$. The scaling of σ_m can also be shown similarly, which also implies that $\mathcal{E}_m^{(1)} = \mathcal{E}_m^{(2)}$. ∎

Proposition 5—Proof:

- If $\exists N$ such that $\hat{\theta}_N \in \Pi(n, \gamma)$, then $|\delta_{N+1}| = |d_{N+1} - \hat{\theta}_N^T \boldsymbol{x}_{N+1}| = \gamma$ since $(\boldsymbol{x}_{N+1}, d_{N+1}) \in \mathcal{D}$. This implies there is no update at time $N + 1$, and $\hat{\theta}_{N+1} = \hat{\theta}_N$. From this, it follows that $\hat{\theta}_{N+l} = \hat{\theta}_N$ for all $l \geq 0$.

- The result does not hold in the trivial case when $\hat{\theta}_o \in \Theta(n, \gamma)$ (since there would be no updates at all, and $\hat{\theta}_\infty = \hat{\theta}_o$). Otherwise, assume the contrary: Let $\hat{\theta}_\infty \in \Phi(n, \gamma) \setminus \Pi(n, \gamma)$. This is possible only if $\exists N > 0$ such that $\hat{\theta}_N \in \mathbb{C}^n \setminus \Theta(n, \gamma)$, and $\hat{\theta}_{N+1} \in \Theta(n, \gamma) \setminus \Pi(n, \gamma)$ [note that if $\hat{\theta}_{N+1} \in \Pi(n, \gamma)$, then by the previous result, there would be no more updates]. This implies that $\exists (\boldsymbol{x}_{N+1}, d_{N+1}) \in \mathcal{D}$ such that $|d_{N+1} - \hat{\theta}_{N+1}^T \boldsymbol{x}_{N+1}| = \gamma$ since the *aposteriori* error magnitude is γ at updating instants. Note that $\hat{\theta}_{N+1}$ lies strictly inside $\Theta(n, \gamma)$. That is, it belongs to $\Theta(n, \gamma) \setminus \Pi(n, \gamma)$. Therefore, there are some parameter vectors $\theta \in \Theta(n, \gamma) \setminus \mathcal{S}_{N+1}$ (which also belong to the feasibility set). Since these θ do not belong to $\mathcal{S}_{N+1}$, $|d_{N+1} - \theta^T \boldsymbol{x}_{N+1}| > \gamma$, which is not possible because θ also belongs to $\Phi(n, \gamma)$ and $(\boldsymbol{x}_{N+1}, d_{N+1}) \in \mathcal{D}$. Hence, $\hat{\theta}_\infty \in \Pi(n, \gamma)$.

- Assume that there is no model violation and that $\limsup_{i \to \infty} |\delta_i| \leq \gamma$ is not true. Then, there exists an $\epsilon > 0$ such that

$$\limsup_{i \to \infty} |\delta_i| = \gamma + \epsilon. \tag{27}$$

In other words, $(|\delta_i| - \gamma) > \epsilon/2$ for infinitely many i. At such instants i, by substituting the value of λ_i^o in the expression for σ_i (17), we have

$$\sigma_i = \sigma_{i-1} - \frac{(|\delta_i| - \gamma)^2}{G_i} < \sigma_{i-1} - \frac{\epsilon^2}{4\overline{G}}$$

where $G_i < \overline{G}$ for all i, and $\overline{G} = \mu \overline{X} > 0$. This, coupled with (27), implies that for some N large enough, we have $\sigma_N < 0$, which is not possible. However, by changing the value of σ_o by any factor does not change the sequence of estimates. Therefore, this result is extendible to arbitrary σ_o. ∎

Proposition 6—Proof:

- When the actual noise bound is overestimated, then there is a feasible set of plants that result in $|e| \leq \gamma$. This set is nonempty since θ_{true} belongs to it. This is so because

$$|d_i - \theta_{\text{true}}^T \boldsymbol{x}_i| \leq \gamma_{\nu, \text{true}} < \gamma$$

for all $(\boldsymbol{x}_i, d_i)$ generated by (22). Since $\varepsilon > 0$, we know that $\theta_{\text{true}} \notin \Pi(n, \gamma)$. However, even if the estimates converge to a point in $\Theta(n, \gamma)$, we have shown that they converge to the boundary of the feasibility set (from Proposition 5). This means that $\hat{\theta}_\infty \neq \theta_{\text{true}}$.

- Using the fact that $\limsup_{i \to \infty} |\delta_i|^2 \leq \gamma^2$ and lower bounding the LHS by the expected value of the squared prediction error, the result follows after substituting the expression for d_i from (22). ∎

Proposition 7—Proof: The gradient of the auxiliary function with respect to θ gives $\theta - \hat{\theta}_{i-1} = \kappa_i^o P_{i-1} \boldsymbol{x}_i^*(d_i - \theta^T \boldsymbol{x}_i)$. The satisfaction of the constraint implies $d_i - \theta^T \boldsymbol{x}_i = (\delta_i/|\delta_i|)\gamma$. From this, we obtain

$$\theta = \hat{\theta}_{i-1} + \frac{\kappa_i P_{i-1} \boldsymbol{x}_i \delta_i}{|\delta_i|/\gamma}$$

From the above results, $\gamma(\delta_i/|\delta_i|) = d_i - \theta^T \boldsymbol{x}_i = \delta_i - (\kappa_i^o G_i \delta_i/|\delta_i|/\gamma)$, which gives the desired result. It is straightforward to verify that this vector is a point of minimum. ∎

REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
[2] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.
[3] S. Gollamudi, S. Nagaraj, S. Kapoor, and Y. F. Huang, "Set-membership filtering and a set-membership normalized LMS algorithm with an adaptive step size," *IEEE Signal Processing Lett.*, vol. 5, May 1998.
[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. Inform. Theory*, vol. IT-33, pp. 383–392, May 1987.
[5] J. R. Deller, M. Nayeri, and M. S. Liu, "Unifying the landmark developments in OBE processing," *Int. J. Automat. Contr. Signal Process.*, vol. 8, pp. 43–60, Jan.–Feb. 1994.
[6] 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.
[7] J. P. Norton, Ed., *Int. J. Automat. Contr. Signal Process.* (Special issues on Bounded-Error Methods in System Identification), vol. 8, Jan.–Feb. 1993–1994.
[8] E. Walter and H. Piet-Lahanier, "Estimation of parameter bounds from bounded-error data: A survey," *Math. Comput. Simul.*, vol. 32, pp. 449–468, 1990.
[9] E. Fogel and Y. F. Huang, "On the value of information in system identification—bounded noise case," *Automatica*, vol. 18, no. 2, pp. 229–238, Mar. 1982.
[10] J. R. Deller, M. Nayeri, and S. F. Odeh, "Least-square identification with error bounds for real-time signal processing and control," *Proc. IEEE*, vol. 81, pp. 813–849, June 1993.
[11] S. Gollamudi, S. Kapoor, S. Nagaraj, and Y. F. Huang, "Set-membership adaptive equalization and an updator-shared implementation for multiple channel communication systems," *IEEE Trans. Signal Processing*, vol. 46, pp. 2372–2385, Sept. 1998.
[12] J. G. Proakis, *Digital Communications, 2nd ed.* New York: McGraw-Hill, 1989.
[13] S. Nagaraj, S. Gollamudi, S. Kapoor, and Y. F. Huang, "Set-membership filtering: A viable tool for nonlinear adaptive signal processing," *Proc. Int. Symp. Non-linear Circuits Syst.*, Hawaii, Dec. 1997.
[14] N. Al-Dhahir and J. M. Cioffi, "MMSE decision feedback equalizers: Finite length results," *IEEE Trans. Inform. Theory*, vol. 41, July 1995.
[15] A. Papoulis, *Probability, Random Variables and Stochastic Processes, 3nd ed.* New York; McGraw-Hill, 1991.
[16] C. R. Johnson, Jr. *et al.*, "On fractionally-spaced equalizer design for digital microwave radio channels," in *Proc. 29th Asilomar Conf.*, Pacific Grove, CA, Nov. 1995.
[17] S. Nagaraj, S. Gollamudi, S. Kapoor, and Y. F. Huang, "Bounded error estimation: Set-theoretic and least-squares formulations," *Proc. Conf. Inform. Sci. Syst.* Baltimore, MD, Johns Hopkins Univ., Mar. 1997.
[18] S. Kapoor, S. Gollamudi, S. Nagaraj, and Y. F. Huang, "Tracking of time-varying parameters using optimal bounding ellipsoid algorithms," in *Proc. 34th Ann. Allerton Conf. Commun., Contr. Comput.*, Oct. 2–4, 1996.

[19] M. Nayeri, M. S. Liu, and J. R. Deller, "Do optimal bounding ellipsoid algorithms converge? Parts I and II," in *Proc. IFAC/IFOSR Conf. Syst. Ident.*, Budapest, Hungary, 1994.
[20] G. C. Goodwin and K. S. Sin, *Adaptive Filtering Prediction and Control*. Englewood Cliffs, N.J.: Prentice-Hall, 1984.
[21] E. W. Bai, K. M. Nagpal, and R. Tempo, "Bounded-error parameter estimation: Noise models and recursive algorithms," *Automatica*, vol. 32, pp. 3–15, July 1996.
[22] S. V. Gusev, "Finite-memory convergent algorithm for estimating the regression function and its application in adaptive control," *Automat. Remote Sensing*, vol. 50, pp. 367–374, 1989.
[23] S. Gollamudi, S. Nagaraj, S. Kapoor, and Y. F. Huang, "Set-membership state estimation with optimal bounding ellipsoids," in *Proc. Int. Symp. Inform. Theory Appl.*, Victoria, B.C., Canada, Sept. 17–20, 1996, pp. 262–265.
[24] D. G. Maksarov and J. P. Norton, "State bounding with ellipsoidal set description of the uncertainty," *Int. J. Contr.*, vol. 65, pp. 847–866, 1996.
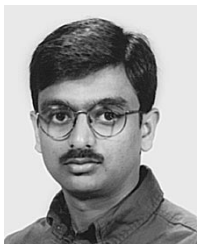
Shirish Nagaraj (S'97) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Bombay, in May 1995 and the M.S. degree from the University of Notre Dame, Notre Dame, IN, in May 1997, also in electrical engineering. He is currently pursuing the Ph.D. degree at the University of Notre Dame.

He worked as a research intern at Lucent Technologies–Bell Laboratories, Whippany, NJ, from June to December 1988 in the Wireless Signal Processing Laboratory. His research interests include adaptive signal processing, statistical estimation and detection theory, set-membership filtering, and multiuser detection for CDMA communications.

Mr. Nagaraj was a recipient of the Michael J. Birck fellowship from 1996 to 1997 and was a graduate fellow of the Center for Applied Mathematics, University of Notre Dame, from 1997 to 1998.

Sridhar Gollamudi (S'96) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Bombay, in 1994 and the M.S.E.E. degree from University of Notre Dame, Notre Dame, IN, in 1996. He is currently pursuing the Ph.D. degree at the Department of Electrical Engineering, University of Notre Dame.

From May to December 1997, he was an intern at Lucent Technologies–Bell Laboratories, Whippany, NJ, where he worked on interference suppression for wireless CDMA systems. His research interests are in the areas of statistical and adaptive signal processing, communication theory, multiuser communications, and antenna array processing.

Mr. Gollamudi was awarded the Center for Applied Mathematics Fellowship at the University of Notre Dame from 1996 to 1997. He is a member of Eta Kappa Nu.

Samir Kapoor (M'98) received the B.Tech., M.S., and Ph.D. degrees in electrical engineering from the Indian Institute of Technology, Bombay, in 1992, Washington State University, Pullman, in 1994, and University of Notre Dame, Notre Dame, IN, in 1998, respectively.

From 1994 to 1998, he was with Tellabs Research Center in Mishawaka, IN. He is currently with NEC USA, Inc., Princeton, NJ. His research interests include digital communications, adaptive signal processing, and array processing.

Dr. Kapoor is a member of Tau Beta Pi and Eta Kappa Nu and serves as an Associate Editor for the IEEE TRANSACTIONS ON MULTIMEDIA.

Yih-Fang Huang (F'95) received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, R.O.C., in 1976, the M.S.E.E. degree from University of Notre Dame, Notre Dame, IN, in 1980, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in 1982.

Since August 1982, he has been on the Faculty at University of Notre Dame, where he is currently Professor and Chair of Electrical Engineering. In Spring 1993, he received the Toshiba Fellowship and was Toshiba Visiting Professor at Waseda University, Tokyo, Japan, in the Department of Electrical Engineering. His research interests are in the general areas of statistical communications and signal processing. He also worked on image source coding and, more recently, joint source-channel coding.

Dr. Huang has served as Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS from 1989 to 1991 and for EXPRESS LETTERS for the same journal from 1992 to 1993. From 1995 to 1996, he was Chairman for the Digital Signal Processing Technical Committee of the IEEE Circuits and Systems Society. He was a Co-Chairman for Workshops/Short Courses for the 1997 *International Symposium on Circuits and Systems*. He served as Regional Editor of America for the *Journal of Circuits, Systems, and Computers*. He also served as Vice President—Publications, of the IEEE Circuits and Systems Society from 1997 to 1998.