

# RECURSIVE BLIND MINIMAX ESTIMATION: IMPROVING MSE OVER RECURSIVE LEAST SQUARES

*Guy Leibovitz, Asaf Elron, Zvika Ben-Haim, and Yonina C. Eldar*  
{ *sgl@t2, selron@t2, zvikabh@tx, yonina@ee* }.*technion.ac.il*

Faculty of Electrical Engineering, Technion - Israel Institute of Technology,  
Haifa 32000, Israel

## ABSTRACT

We consider the problem of on-line (or recursive) parameter estimation in which, at each moment, an unknown deterministic parameter vector must be re-estimated from measurements corrupted by additive noise. We present efficient algorithms for calculating two variants of the blind minimax estimator, which is a biased estimator proven to outperform least squares in terms of mean squared error. These operate in the same setting as the recursive least squares (RLS) method and utilize it. Both algorithms have a computational complexity in par with RLS. We discuss the advantages and shortcomings of the presented methods and demonstrate through simulations situations in which they produce substantial gain over RLS.

## 1. INTRODUCTION

The problem of on-line (or recursive) parameter estimation in the linear regression setting has numerous applications in engineering; these include, for example, the areas of system identification and array processing. We consider the case where the estimation parameter is deterministic and the measurements are corrupted by additive white Gaussian noise.

A classical way of performing such an estimate is the least squares (LS) method. This method produces the lowest mean squared error (MSE) amongst all unbiased estimators. Furthermore, by recognizing the dependencies between subsequent points in time, an efficient algorithm for calculating the LS estimator can be derived. The result, known as recursive least squares (RLS), involves only  $\mathcal{O}(m^2)$  scalar multiplications per step, where  $m$  is the number of parameters to be estimated.

Stein showed that LS can be dominated, i.e. that lower MSE can be achieved anywhere in the parameter space, by a biased estimator [3]. Since then, a number of estimation techniques that dominate LS were proposed [5][4][6].

In this paper we introduce a new method, based on the recently developed blind minimax estimator (BME) [6], for efficiently calculating an LS dominating estimator recursively. Previous work in this

field includes [7], in which a recursive realization of Bock's estimator [4] is proposed. However, the BME usually outperforms Bock's estimator [6].

We elaborate on the notion of on-line parameter estimation in Section 2. In Section 3 we give the required background, including a short review of the blind minimax estimator. Section 4 presents our proposed solution and explains how it is derived. In Section 5 we present a simulated example comparing RLS with our proposed solution. Finally, we discuss our findings in Section 6.

Throughout this paper, for a column vector  $\mathbf{a}$  and a positive definite matrix  $\mathbf{A}$ ,  $\|\mathbf{a}\|^2$  denotes the Euclidean square norm  $\mathbf{a}^H \mathbf{a}$  and  $\|\mathbf{a}\|_{\mathbf{A}}^2$  denotes the weighted square norm  $\mathbf{a}^H \mathbf{A} \mathbf{a}$ . The time index is  $n$  and the sample time for scalars, vectors and matrices will be displayed in subscript.  $\{\mathbf{A}\}_{i,:}$  is the  $i$ -th row of  $\mathbf{A}$ ,  $\{\mathbf{a}\}_i$  is the  $i$ -th element of the vector  $\mathbf{a}$ .  $\lambda_{\min}(\mathbf{A})$  and  $\lambda_{\max}(\mathbf{A})$  are the minimal and maximal eigenvalues of  $\mathbf{A}$ , respectively.

## 2. PROBLEM STATEMENT

Consider the problem of on-line parameter estimation in which, at each moment  $n$ , the unknown deterministic parameter vector  $\mathbf{x}$  must be re-estimated from the measurements

$$\mathbf{y}_n = \mathbf{H}_n \mathbf{x} + \mathbf{w}_n. \quad (2.1)$$

Here,  $\mathbf{w}_n$  is white Gaussian noise with known variance  $\sigma_w^2$ ,  $\mathbf{y}_n$  is a known vector and  $\mathbf{H}_n$  is a known matrix of full rank. At each moment, the measurement vector  $\mathbf{y}_{n-1}$  is extended by a new measurement  $y_n$  to form  $\mathbf{y}_n$  and a new row  $\boldsymbol{\rho}_n^T$  is added to the model matrix  $\mathbf{H}_{n-1}$ , modeling the relation between the new measurement  $y_n$  and the unknown parameter vector  $\mathbf{x}$ . Such a problem arises, for example, in a system identification scenario [1].

In this setting, one would want to utilize the new data in order to get a more accurate estimation of  $\mathbf{x}$  while relying on previously calculated data in order to reduce computation time. A classical method for accomplishing this is using a recursive algorithm (known as RLS) to calculate the least squares estimate of  $\mathbf{x}$  [1]. Our method improves on this approach, by achieving lower mean squared error for every  $n$ , under certain regularity conditions.

---

This work was supported by the the Signal and Image Processing Lab, Faculty of Electrical Engineering, Technion - Israel Institute of Technology.

### 3. BACKGROUND

We now review some common definitions and the previous work on which we base our proposed solution. For ease of notation, the following non-recursive formulations will be presented in terms of the one-shot parameter estimation problem

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \quad (3.1)$$

where  $\mathbf{x}$  is an unknown deterministic parameter vector,  $\mathbf{w}$  is white Gaussian noise with known variance  $\sigma_w^2$ ,  $\mathbf{y}$  is a known measurement vector and  $\mathbf{H}$  is a known matrix.

A common measure for estimation quality is the MSE. Denoting the estimator of  $\mathbf{x}$  by  $\hat{\mathbf{x}} = \hat{\mathbf{x}}(\mathbf{y})$ , the MSE is

$$\begin{aligned} \text{MSE}(\hat{\mathbf{x}}) &\triangleq E\{\|\hat{\mathbf{x}} - \mathbf{x}\|^2\} \\ &= \text{var}(\mathbf{x}, \hat{\mathbf{x}}) + \|\text{bias}(\mathbf{x}, \hat{\mathbf{x}})\|^2 \end{aligned} \quad (3.2)$$

where  $\text{var}(\mathbf{x}, \hat{\mathbf{x}}) \triangleq \text{tr}(\text{cov}(\hat{\mathbf{x}}))$  and  $\text{bias}(\mathbf{x}, \hat{\mathbf{x}}) \triangleq E\{\hat{\mathbf{x}}\} - \mathbf{x}$  are the estimator variance and bias, respectively. Unlike the Bayesian setting, in which an estimator that minimizes the MSE can be formulated, it is impossible to minimize the MSE for all values of the unknown parameter vector  $\mathbf{x}$ .

#### 3.1. Least squares estimation

Least squares estimation is a long-standing, widespread solution for (3.1). It is derived by requiring unbiasedness and is proven to produce the lowest MSE amongst all unbiased estimators. The estimator is given by

$$\hat{\mathbf{x}}^{\text{LS}} = (\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H \mathbf{y} \quad (3.3)$$

and is a random vector distributed as

$$\hat{\mathbf{x}}^{\text{LS}} \sim \mathcal{N}(\mathbf{x}, \mathbf{Q}^{-1}) \quad (3.4)$$

where  $\mathbf{Q}^{-1} \triangleq \sigma_w^2 (\mathbf{H}^H \mathbf{H})^{-1}$ . This results in an MSE depending solely on  $\mathbf{Q}^{-1}$  and given by

$$\text{MSE}(\hat{\mathbf{x}}^{\text{LS}}) = \epsilon^{\text{LS}} \triangleq \text{tr}(\mathbf{Q}^{-1}). \quad (3.5)$$

Although it is seemingly very appealing to restrict oneself to unbiased estimators, and opting to minimize the estimator variance alone, such an approach does not necessarily minimize the MSE. In fact, it has been proven that, under some regularity conditions, LS can be dominated, i.e., lower MSE can be achieved for all values of  $\mathbf{x}$  by a biased estimator. This result is known as Stein's phenomenon. Two such estimators are the James-Stein estimator [5], which addresses the case in which  $\mathbf{H} = \mathbf{I}$ , and the blind minimax estimator (BME) [6].

#### 3.2. Blind minimax estimation

Suppose for a moment that  $\mathbf{x}$  is known to lie within a bounded set  $\mathcal{S}$ . In such a case, it is possible to construct a linear minimax estimator

$$\hat{\mathbf{x}}^{\text{M}} = \arg \min_{\hat{\mathbf{x}} = G\mathbf{y}} \max_{\mathbf{x} \in \mathcal{S}} E\{\|\hat{\mathbf{x}} - \mathbf{x}\|^2\}. \quad (3.6)$$

This estimator minimizes the worst-case MSE among all possible values of  $\mathbf{x}$  in  $\mathcal{S}$ . A closed-form formulation of (3.6) has been derived for many cases of interest [8].

The blind minimax estimator [6], on the other hand, first estimates a set  $\mathcal{S}$  in which  $\mathbf{x}$  is likely to lie. It then utilizes the minimax estimator for that set on the measurements to estimate a value for  $\mathbf{x}$ . This produces an estimator that requires no prior knowledge of  $\mathbf{x}$ .

##### 3.2.1. Spherical blind minimax estimation

Choosing  $\mathcal{S}$  to be the sphere  $\mathcal{S} = \{\mathbf{x} : \|\mathbf{x}\|^2 \leq \|\hat{\mathbf{x}}^{\text{LS}}\|^2\}$  results in the spherical blind minimax estimator (SBME)

$$\hat{\mathbf{x}}^{\text{SBM}} = \frac{\|\hat{\mathbf{x}}^{\text{LS}}\|^2}{\|\hat{\mathbf{x}}^{\text{LS}}\|^2 + \epsilon^{\text{LS}}} \hat{\mathbf{x}}^{\text{LS}} \quad (3.7)$$

where  $\hat{\mathbf{x}}^{\text{LS}}$  is the LS estimator (3.3) and  $\epsilon^{\text{LS}}$  is its MSE (3.5). One should note that this is a scalar shrinkage of the LS estimator, implying that  $\|\hat{\mathbf{x}}^{\text{LS}}\|^2$  tends to be an overestimate of  $\|\mathbf{x}\|^2$ .

It has been shown [6] that when

$$\text{tr}(\mathbf{Q}^{-1}) / \lambda_{\max}(\mathbf{Q}^{-1}) > 4 \quad (3.8)$$

the SBME (3.7) strictly dominates LS. The ratio in (3.8) is known as the effective dimension of the estimation problem.

##### 3.2.2. Ellipsoidal blind minimax estimation

Not all elements of the LS estimate  $\hat{\mathbf{x}}^{\text{LS}}$  are equally trustworthy, due to (3.4). Some elements have larger variance than others and so scalar shrinkage of  $\hat{\mathbf{x}}^{\text{LS}}$  may be suboptimal in treating these elements. The ellipsoidal blind minimax estimator (EBME) addresses this issue by shrinking each element of  $\hat{\mathbf{x}}^{\text{LS}}$  according to its variance.

This is accomplished by first estimating an ellipsoidal set  $\mathcal{S}$  likely to contain  $\mathbf{x}$

$$\mathcal{S} = \{\mathbf{x} : \|\mathbf{x}\|_{\mathbf{Q}^b}^2 \leq \|\hat{\mathbf{x}}^{\text{LS}}\|_{\mathbf{Q}^b}^2\} \quad (3.9)$$

for some real scalar  $b$  to be chosen below. Then, using the minimax estimator with  $\mathcal{S}$  produces the ellipsoidal blind minimax estimator (EBME) [6]

$$\hat{\mathbf{x}}^{\text{EBM}} = \mathbf{V} \text{diag}((1 - \alpha \sigma_1^{b/2})_+, \dots, (1 - \alpha \sigma_m^{b/2})_+) \mathbf{V}^H \hat{\mathbf{x}}^{\text{LS}} \quad (3.10)$$

where  $\mathbf{V}\mathbf{\Sigma}\mathbf{V}^H$  is the eigenvalue decomposition of  $\mathbf{Q}$  with  $\mathbf{V}$  unitary,  $\mathbf{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_m)$  and  $\sigma_1^b \geq \sigma_2^b \geq \dots \geq \sigma_m^b > 0$ . Here  $(\cdot)_+ = \max(\cdot, 0)$ ,

$$\alpha = \frac{r_1}{\|\hat{\mathbf{x}}^{\text{LS}}\|_{\mathbf{Q}^b}^2 + r_2} \quad (3.11)$$

$$r_1 = \sum_{i=k+1}^m \sigma_i^{b/2-1} \quad r_2 = \sum_{i=k+1}^m \sigma_i^{b-1}$$

and  $k$  is chosen as the smallest index  $0 \leq k \leq m-1$  such that

$$\alpha \sigma_{k+1}^{b/2} < 1. \quad (3.12)$$

It has been shown [6] that when  $\mathbf{Q}$  satisfies

$$\text{tr}(\mathbf{Q}^{b/2-1}) / \lambda_{\max}(\mathbf{Q}^{b/2-1}) > 4, \quad (3.13)$$

the EBME (3.10) strictly dominates LS (3.3). We refer to the ratio in (3.13) as the modified effective dimension.

Recall from (3.4) that  $\mathbf{Q}^{-1}$  provides information on the accuracy of  $\hat{\mathbf{x}}^{\text{LS}}$ . Each eigenvalue of  $\mathbf{Q}^{-1}$  determines the variance the LS estimator will have in the direction of the associated eigenvector. We refer to the eigenvectors of  $\mathbf{Q}^{-1}$  as the eigendirections of  $\hat{\mathbf{x}}^{\text{LS}}$  and to the projections of  $\hat{\mathbf{x}}^{\text{LS}}$  on the eigenvectors as the eigenelements of  $\hat{\mathbf{x}}^{\text{LS}}$ .

In the EBME, a different shrinkage factor of  $1 - \alpha\sigma_i^{b/2}$  is applied to each eigendirection of  $\hat{\mathbf{x}}^{\text{LS}}$ . The amount of shrinkage in each direction is dependent on the variance in that direction as well as on the scalar  $b$ , which we refer to as the discrimination factor. When  $b < 0$ , the EBME shrinks high-variance eigenelements of  $\hat{\mathbf{x}}^{\text{LS}}$  more than it shrinks low-variance ones. Furthermore, the more negative the value of  $b$ , the more eccentric the resulting ellipsoid  $\mathcal{S}$ . This, in turn, results in greater shrinkage of high-variance eigenelements relative to that of low-variance ones. When  $b = 0$ , the ellipsoid  $\mathcal{S}$  degenerates to a sphere, leading to the scalar shrinkage of the SBME (3.7).

In the estimation process, the EBME zeros out eigenelements of  $\hat{\mathbf{x}}^{\text{LS}}$  associated with eigenvalues satisfying  $\alpha\sigma_i^{b/2} > 1$ , an operation called the positive-part correction. The elements zeroed out are those associated with very large variances, compared to the other elements of  $\hat{\mathbf{x}}^{\text{LS}}$ .

### 3.3. Recursive least squares

For the on-line parameter estimation problem (2.1), the recursive least squares (RLS) algorithm accurately calculates the LS estimation of  $\mathbf{x}$  at each time  $n$ . To this end and remembering (3.3), it is useful to define

$$\mathbf{Q}_n \triangleq \sigma_w^{-2} \mathbf{H}_n^H \mathbf{H}_n. \quad (3.14)$$

In this on-line problem (2.1),  $\mathbf{Q}_n$  is given as a rank-1 update of  $\mathbf{Q}_{n-1}$

$$\mathbf{Q}_n = \sigma_w^{-2} (\mathbf{H}_{n-1}^H \mathbf{H}_{n-1} + \rho_n \rho_n^H). \quad (3.15)$$

Utilizing the rank-1 update (3.15), it is possible to efficiently update  $\mathbf{P}_n$ , the inverse of  $\sigma_w^2 \mathbf{Q}_n$ , given  $\mathbf{P}_{n-1}$ . Each update includes only matrix-vector multiplications, thus requiring only  $\mathcal{O}(m^2)$  computational operations.

For future reference, we state the algorithm as

$$\phi = \mathbf{P}_{n-1} \rho \quad (3.16)$$

$$\gamma = 1/(1 + \rho^H \phi) \quad (3.17)$$

$$\hat{\mathbf{x}}_n^{\text{LS}} = \hat{\mathbf{x}}_{n-1}^{\text{LS}} + \gamma(y_n - \rho^H \hat{\mathbf{x}}_{n-1}^{\text{LS}}) \phi \quad (3.18)$$

$$\mathbf{P}_n = (\mathbf{P}_{n-1} - \gamma \phi \phi^H) \quad (3.19)$$

where  $\hat{\mathbf{x}}_n^{\text{LS}}$  is the least squares estimate for  $\mathbf{x}$  at time  $n$  and  $\sigma_w^2 \mathbf{P}_n$  is its covariance matrix [2].

## 4. PROPOSED SOLUTION

In the following section we introduce an adaptation of the blind minimax estimators of Section 3.2 to the on-line estimation scenario

(2.1).

### 4.1. Spherical blind minimax

The SBME can readily be calculated on-line as a simple modification of the RLS algorithm. This is done by obtaining  $\hat{\mathbf{x}}_n^{\text{LS}}$  (3.18) and  $\mathbf{P}_n$  (3.19) using RLS, then substituting  $\hat{\mathbf{x}}^{\text{LS}}$  for  $\hat{\mathbf{x}}_n^{\text{LS}}$  and  $\epsilon_n^{\text{LS}} \triangleq \sigma_w^2 \text{tr}(\mathbf{P}_n)$  for  $\epsilon^{\text{LS}}$  in SBME (3.7). Thus, the SBME is very cheaply calculated, requiring an additional  $m + 1$  multiplications over RLS. At every moment  $n$ , dominance over the LS estimator is guaranteed, as long as (3.8) holds.

Similar to the SBME, the James-Stein RLS estimator [7] also applies scalar shrinkage to  $\hat{\mathbf{x}}^{\text{LS}}$ , recursively calculating Bock's estimator [4]. However, as demonstrated in [6], the SBME usually produces lower MSE than Bock's estimator.

### 4.2. Recursive full-rank ellipsoidal blind minimax

#### 4.2.1. Overview

As discussed in Section 3.2.2, the EBME applies non-scalar shrinkage to the LS estimator. We propose to use an adaptation of EBME to operate in the same computational complexity as RLS. The guiding principles in the choice of the derived estimator are achieving low computational complexity and non-scalar shrinkage, while maintaining dominance over LS.

To this end, observe that EBME (3.10) is the positive-part correction of the full-rank ellipsoidal blind minimax estimator (frEBME) which is the result of substituting  $k = 0$  in (3.10) and (3.11). At every moment  $n$ , the frEBME dominates the LS estimator under the same condition as EBME, i.e., as long as (3.13) holds [6].

Applying the positive-part correction to the above estimator requires obtaining the eigenvalue decomposition (EVD) of  $\mathbf{Q}_n$ , which necessitates many additional computations on top of RLS. In fact, most of the available EVD rank-1 update algorithms either do so in  $\mathcal{O}(m^2 r)$ , where  $r$  is the number of eigenpairs tracked, or rely on an iterative convergence method (e.g. see [9][10][11]). For the above reason, the choice of whether or not to apply the positive-part correction is left to the designer. A technique for applying the positive-part correction using a tracked eigenvector subspace is proposed in Section 4.2.3. In Section 4.2.2, we will use the frEBME as the basis for our proposed recursive estimator.

To facilitate low computational requirements, we choose a discrimination factor of  $b = -2$ . With this choice, and using parts of the already calculated RLS equations, calculation of the estimator is reduced to  $6m^2$  multiplications per time step. By contrast, RLS requires  $2m^2$  multiplications per time step. Further note that, for  $b = -2$ , no positive-part correction is needed if  $\|\hat{\mathbf{x}}^{\text{LS}}\|^2 \geq \text{tr}(\mathbf{P}^2 (\lambda_{\max}(\mathbf{P}) \mathbf{I} - \mathbf{P})) / \lambda_{\min}^2(\mathbf{P})$ . This includes cases where  $\mathbf{Q}$  is a scalar matrix and cases in which  $\mathbf{x}$  itself is very large. From various simulations conducted, we observed that almost always  $k = 0$ , with  $k > 1$  being an extreme rarity.

#### 4.2.2. Mathematical formulation

In order to compute the recursive full-rank ellipsoidal blind mini-max estimator (RfrEBME) we first note that the matrix  $\mathbf{P}_n^2$  can be updated as

$$\mathbf{P}_n^2 = \mathbf{P}_{n-1}^2 - \gamma \phi \phi^H \mathbf{P}_{n-1} - \gamma \mathbf{P}_{n-1} \phi \phi^H + \gamma^2 \phi \phi^H \phi \phi^H$$

where  $\gamma$  and  $\phi$  are obtained from (3.17) and (3.16). After  $\mathbf{P}_n$  and  $\mathbf{P}_n^2$  have been calculated,  $\hat{\mathbf{x}}^{\text{RfrEBM}}$  is obtained by

$$\begin{aligned} \tilde{r}_1 &= \text{tr}(\mathbf{P}_n^2) \\ \tilde{r}_2 &= \text{tr}(\mathbf{P}_n^3) = \mathbf{1}^{1 \times N} \cdot (\mathbf{P}_n^2)^T * \mathbf{P}_n \cdot \mathbf{1}^{N \times 1} \\ \tilde{\alpha} &= \frac{\tilde{r}_1}{\tilde{r}_2 + \sigma_w^{-2} \|\hat{\mathbf{x}}^{\text{LS}}\|_{\mathbf{P}_n^2}^2} \\ \hat{\mathbf{x}}^{\text{RfrEBM}} &= (\mathbf{I} - \tilde{\alpha} \mathbf{P}_n) \hat{\mathbf{x}}^{\text{LS}} \end{aligned}$$

where  $\mathbf{1}^{A \times B}$  denotes an  $A \times B$  matrix whose elements are all ones and  $*$  denotes element-by-element matrix multiplication.

#### 4.2.3. Further improvement

At the cost of computation time, the estimator can be improved by applying some or all of the necessary positive-part corrections. This can be done even without a complete EVD of  $\mathbf{P}_n$ , as follows.

Given the  $r$  largest eigenvalues  $\lambda_1 \geq \dots \geq \lambda_r$  of  $\mathbf{P}_n$  and their corresponding normalized eigenvectors  $\mathbf{v}_1, \dots, \mathbf{v}_r$ , it is possible to apply the positive-part correction if  $k \leq r$ . The EBME then takes the form

$$\begin{aligned} \tilde{\mathbf{P}} &= (\mathbf{P} - \mathbf{W}(\mathbf{L} - \tilde{\alpha}^{-1} \mathbf{I}_k) \mathbf{W}^H) \\ \tilde{\alpha} &= \frac{\tilde{r}_1 - \sum_{i=1}^k \lambda_i^2}{\tilde{r}_2 - \sum_{i=1}^k \lambda_i^3 + \sigma_w^{-2} \|\hat{\mathbf{x}}^{\text{LS}}\|_{\tilde{\mathbf{P}}}^2} \\ \hat{\mathbf{x}}^{\text{REBM}} &= (\mathbf{I} - \tilde{\alpha} \tilde{\mathbf{P}}) \hat{\mathbf{x}}^{\text{LS}} \end{aligned}$$

where  $k$  is determined by (3.12)

$$k = \arg \max \lambda_{k+1}, \text{ subject to } \alpha \lambda_{k+1} \leq 1. \quad (4.1)$$

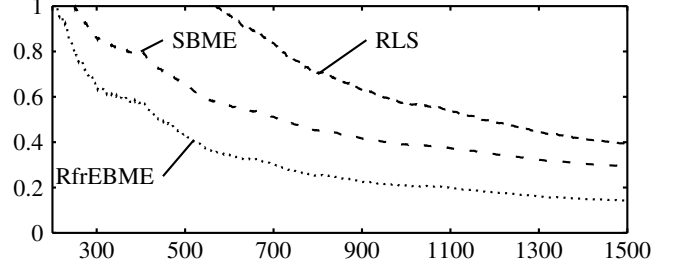
Here  $\mathbf{W} = [\mathbf{v}_1, \dots, \mathbf{v}_k]$  and  $\mathbf{L} = \text{diag}(\lambda_1, \dots, \lambda_k)$ .

It is possible to utilize a low-complexity major-subspace tracker (e.g. [9][10][11]) to obtain  $\mathbf{W}$  and  $\mathbf{L}$ . Note that it may be preferable to start from  $r = 0$  and increment it at each step if (4.1) does not hold.

## 5. USAGE EXAMPLE AND NUMERICAL RESULTS

A classical application of the recursive parameter estimation scenario (2.1) is channel estimation for an autoregressive model with exogenous input (ARX) [2]. Given the known input signal  $u_n$ , an ARX model of orders  $(p, q)$  is

$$y_n = \sum_{i=1}^p a_i y_{n-i} + \sum_{i=1}^q b_i u_{n-i} + w_n \quad (5.1)$$



**Fig. 1.**  $\text{MSE}/\|\mathbf{x}\|^2$  as a function of the number of observations for RLS, SBME and RfrEBME. The input is a square wave with a period of  $2\pi$ . The filter is a 19th order FIR whose coefficients are proportional to the eigendirection having minimal variance. The effective dimension and modified effective dimension rise steadily with  $n$  and converge to 9.3 and 6.9, respectively.

where  $\{a_i\}_{i=1}^p$  and  $\{b_i\}_{i=1}^q$  are unknown filter coefficients and  $w_n$  is additive white Gaussian noise with variance  $\sigma_w^2$ . Casting the above as an on-line parameter estimation problem (2.1) gives

$$\begin{aligned} \mathbf{x} &= [a_1, \dots, a_p, b_1, \dots, b_q]^T \\ \rho_n &= \{\mathbf{H}_n\}_{:,n} = [y_{n-1}, \dots, y_{n-p}, u_{n-1}, \dots, u_{n-q}]^T. \end{aligned} \quad (5.2)$$

We assume that  $u_i = y_i = 0$  for  $i < 0$ . The length of the parameter vector is  $m = p + q$ .

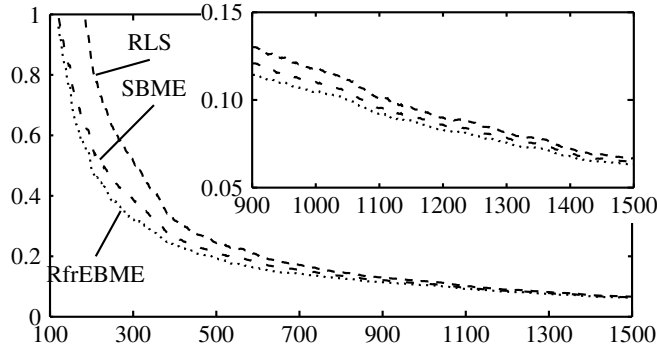
For the special case  $p = 0$ , the model (5.1) reduces to an FIR filter with additive noise. The problem is then to estimate the filter coefficients based on the known input and known (but noisy) output. In this case, the matrix  $\mathbf{H}_n$  contains only deterministic values  $u_n$ , and (5.2) corresponds precisely to the on-line setting (2.1). Thus, the proposed recursive BME methods are guaranteed to achieve lower MSE than the standard RLS solution for all values of  $n$ .

This is demonstrated in Fig. 1, in which the performance of RLS, SBME, and RfrEBME are compared in an FIR estimation scenario. In this simulation, an SNR of 5 dB was chosen, where the SNR is defined as

$$\text{SNR} \triangleq \frac{\|\mathbf{z}_n\|^2}{E\{\|\mathbf{g}_n\|^2\}}. \quad (5.3)$$

with  $\mathbf{z}_n \triangleq \mathbf{y}_n|_{w_n=0}$  and  $\mathbf{g}_n \triangleq \mathbf{y}_n|_{u_n=0}$ . The MSE was calculated by averaging the performance over 100 noise realizations. The plotted MSE is normalized by  $\|\mathbf{x}\|^2$  to demonstrate that reasonable performance is achieved relative to the trivial estimator  $\hat{\mathbf{x}} \equiv 0$ . This is because the low SNR value is compensated by the large number of observations.

In the general ARX case, the model matrix  $\mathbf{H}_n$  contains measurements of both the input signal  $u_n$  and the output signal  $y_n$ . Therefore,  $\mathbf{H}_n$  is random, so that  $\hat{\mathbf{x}}_n^{\text{LS}}$  is no longer distributed as  $\mathcal{N}(\mathbf{x}, \mathbf{Q}_n^{-1})$ ; consequently, the dominance results of the blind mini-max estimators no longer hold in this case. Furthermore, in this case, the effective dimension (3.13) also depends on the measurements, and thus is not known in advance. However, even in the general ARX case,  $\hat{\mathbf{x}}_n^{\text{LS}}$  converges in distribution to  $\mathcal{N}(\mathbf{x}, \mathbf{Q}_n^{-1})$  as  $n \rightarrow \infty$ ,



**Fig. 2.**  $\text{MSE}/\|x\|^2$  as a function of the number of observations for RLS, SBME and RfrEBME. The input is a randomly generated IID signal. The filter is a 10th order Butterworth low-pass with a cut-off frequency of  $\pi/2$ . The SNR is  $-20$  dB. The average effective dimension and average modified effective dimension converge to 9.5 and 7.5, respectively. The inlaid graph displays the same results in a different scale.

so that dominance over RLS can be expected for large  $n$ . In fact, simulations show that when the effective dimension satisfies the respective dominance conditions, the SBME and RfrEBME typically produce lower MSE than RLS. One example is provided in Fig. 2, in which the different methods are used to estimate the parameters of an IIR filter. The SNR in this case is  $-20$  dB. Apart from the filter and input signal, the settings are identical to those of Fig. 1.

While the dominance of the BMEs is clearly visible for all  $n$ , the difference becomes less pronounced as the number of measurements increases. This is because the LS estimator is optimal in terms of MSE in the limit  $n \rightarrow \infty$ , so that all methods approach this optimal estimate for large  $n$ .

## 6. DISCUSSION

In this paper we surveyed several possible uses of the recently proposed BME [6] framework in an on-line model (2.1). The first is the easily calculated SBME, performing a scalar shrinkage of the LS estimator  $\hat{x}^{\text{LS}}$ , requiring additional  $m + 1$  multiplications on top of RLS's  $2m^2$ . However, the use of scalar shrinkage is sub-optimal in that all components of  $\hat{x}^{\text{LS}}$  receive equal treatment, regardless of their variance. To improve performance, we introduced the RfrEBME, which allows for non-scalar shrinkage, and requires  $6m^2$  multiplications. Although RfrEBME poses, in our opinion, a reasonable trade-off between computational requirements and performance, we suggest a means for improving its performance, at the cost of additional computations. We recommend considering this modification when only a few noisy directions of  $\hat{x}^{\text{LS}}$  cause a large MSE [6].

We observed from simulations that, in many system identification scenarios, SBME and RfrEBME produce a lower MSE even when the dominance conditions do not hold. Considerable improve-

ment in MSE is achieved under a variety of operating conditions, with the difference being most pronounced for low SNR.

In summary, we have presented several methods for on-line computation of a deterministic parameter, which are guaranteed to improve on the standard RLS approach. For applications in which the SNR is relatively low, we recommend using the RfrEBME for it generally produces better results than SBME at a reasonable computational cost.

## 7. REFERENCES

- [1] T. Söderström, P. Stoica, *System Identification*, Prentice Hall, 1989.
- [2] S. Haykin, *Adaptive Filter Theory*, 4rd ed., Prentice-Hall, 2001.
- [3] C. Stein, "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," *Proc. 3rd Berkeley Symp. Mathematical Statistics and Probability*, Berkeley, CA, 1956, vol. 1, pp. 197-206.
- [4] M. E. Bock, "Minimax estimators of the mean of a multivariate normal distribution," *Ann. Statist.*, vol. 3, no. 1, pp. 209-218, Jan. 1975.
- [5] W. James and C. Stein, "Estimation with quadratic loss," *Proc. 4th Berkeley Symp. Mathematical Statistics and Probability*, Berkeley, CA, 1961, vol. 1, pp. 311-319.
- [6] Z. Ben-Haim and Y. C. Eldar, "Blind minimax estimation," *IEEE Trans. Information Theory*, vol. 53, no. 9, pp. 3145-3157, Sep. 2007.
- [7] J.H. Manton, V. Krishnamurty and H.V. Poor, "James-Stein state filtering algorithms," *IEEE Trans. Signal Process.*, vol. 46, no. 9, pp. 2431-2447, Sep. 1998.
- [8] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 168-181, Jan. 2005.
- [9] R. Badeau, G. Richard, and B. David, "Fast and Stable YAST Algorithm for Principal and Minor Subspace Tracking," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3437-3446, Aug. 2008.
- [10] Kai-Bor Yu, "Recursive Updating the Eigenvalue Decomposition of a Covariance Matrix", *IEEE Trans. Signal Process.*, vol. 39, no. 5, pp. 1136-1145, May 1991.
- [11] R. D. Degroot, and R. A. Roberts, "Efficient, Numerically Stabilized Rank-One Eigenstructure Updating", *IEEE Trans. Signal Process.*, vol. 38, no. 2, pp. 301-315, Feb. 1990.