CRAMÉR-RAO-TYPE BOUND FOR STATE ESTIMATION IN LINEAR DISCRETE-TIME SYSTEM WITH UNKNOWN SYSTEM PARAMETERS

Shahar Bar and Joseph Tabrikian

Dept. of Electrical and Computer Engineering
Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel
Email: shahba@post.bgu.ac.il, joseph@ee.bgu.ac.il

ABSTRACT

Tracking problems are usually investigated using the Bayesian approach. Many practical tracking problems involve some unknown deterministic nuisance parameters such as the system parameters or noise statistical parameters. This paper addresses the problem of state estimation in linear discrete-time dynamic systems in the presence of unknown deterministic system parameters. A Cramér-Rao-type bound on the mean-sqaure-error (MSE) of the state estimation is introduced. The bound is based on the concept of risk-unbiasedness and can be computed recursively. It allows evaluating the optimality of the estimation procedure. Some sequential estimators for this problem are proposed such that the estimation procedure can be considered an on-line technique. Simulation results show that the proposed bound is asymptotically achieved by the considered estimators.

***Index Terms*—** Kalman filter, sequential estimation, Cramér-Rao bound, risk-unbiased bound, MSE

1. INTRODUCTION

Since Kalman's breakthrough [1], the problem of state estimation in linear and nonlinear systems has been extensively investigated and various aspects of the problem have been analyzed. In a survey paper [2], Kailath revisited the problem of linear sequential estimation and presented a new approach to the problem based on innovations. Using this approach, Mehra [3] derived a test for the optimality of a particular Kalman filter, in sense of the whiteness of the innovations process, and showed how it can be adjusted when the noise variances are not precisely known. Since Mehra's work, different aspects of the problem of state estimation with unknown system parameters and statistics have been investigated. In [4, 5], the direct identification of stochastic linear systems has been reported using the steady-state innovations model. Estimation of the system state vector in the presence of unknown deterministic parameters was treated in [6, 7]. Parameter estimation was also treated as a special case of estimation of state variable via a maximum likelihood based criterion [8, 9]. The same criterion was used in [10–13] for estimating the parameters of a linear system. Furthermore, estimation of the covariances of the driving noise and observation noise has been considered under the steady-state assumption [14, 15].

In [16], the simultaneous estimation of the state and the parameters of a linear dynamic system subject to an arbitrary known input, a driving noise, and noisy output observations. The system parameters were treated as unknown deterministic parameters, interfering with the estimation of the state vector. The approach of [16] for parameter estimation was revisited in a general context [17, 18]. However, no analytical approach has been suggested for evaluating the MSE of state vector estimation in the presence of the unknown statistics. Thus, the optimality of the solution in [16] for state vector estimation has not been verified. Another drawback of this solution is that it requires the recomputation of the smoothed values of all the past states. Thus, this approach lacks the attractivity of sequential estimation. In [19] the scalar model of [16] was revisited as a special case of an auto-regressive (AR) model. Two alternatives to the system identification were applied based on recursive least squares (RLS) [20] and least mean square (LMS) [21] algorithms.

The problem of state estimation in presence of unknown system parameters is a special case of the problem of Bayesian estimation in presence of deterministic nuisance parameters. In a recent research [22, 23], this general problem has been approached using the concept of risk unbiasedness [24]. The basic ideas of this research has been presented in [25]. In [22] a new CR-type bound which was given the name risk-unbiased bound (RUB), is derived for risk-unbiased estimation of random parameters for a models with unknown deterministic nuisance parameters, which were treated as nuisance. The succeeding paper [23] deals with two alternatives to the conventional minimum MSE (MMSE) estimator and the maximum *a-posteriori* probability (MAP) estimator using MLE's for the nuisance parameters. The discussed estimators are the combined MMSE-maximum likelihood estimator and the joint MAP-maximum likelihood estimator [16, 18], which are denoted by MS-ML and JMAP-ML, respectively. These estimators were shown to be both asymptotically uniformly mean-unbiased and asymptotically uniformly risk-unbiased. Furthermore, the RUB was proven to be asymptotically tight for the MS-ML estimator, while mean-unbiased Cramér-Rao-type bounds, such as the hybrid Cramér-Rao bound [26, 27], are only tight in distinct cases.

In [16] the estimation is considered as optimal in the sense that the system identification is asymptotically efficient. In this paper, we wish to examine the optimality of estimators from the point of view of the state estimation. An estimator is considered as optimal if its MSE achieves the RUB as the number of observations increases. We consider the RUB as a benchmark for asymptotic performance, i.e. for a large number of observations, because of the properties mentioned above which are proven in [23]. Some of the proposed estimators are adopted from [19] to maintain sequential state estimation as much as possible.

In this paper, the estimation of the state of a scalar linear dynamic system subject to a driving noise and noisy output observations is addressed. A new Bayesian lower bound on the MSE of the state estimation is derived, based on the RUB. Several sequential estimation methods are examined and their optimality is tested, in the sense of their asymptotic risk-efficiency. That is, how fast (if at all) these estimators attain the proposed bound.

2. STATEMENT OF THE PROBLEM

Consider a discrete-time linear scalar system modeled by

$$x_{n+1} = ax_n + u_n,\ n = 0, \dots, N-1, \tag{1a}$$
$$y_n = x_n + v_n,\ n = 0, \dots, N, \tag{1b}$$

where x_n is the state of the system at time n, $\{u_n\}_{n=0}^{N}$ is a white Gaussian driving noise with $u_n \sim \mathcal{N}(0, \sigma_u^2)$, a is a deterministic unknown system parameter, y_n is the observation at time n, and $\{v_n\}_{n=0}^{N+1}$ is a white Gaussian observation noise with $v_n \sim \mathcal{N}(0, \sigma_v^2)$. $\{u_n\}_{n=0}^{N-1}$, $\{v_n\}_{n=0}^{N}$, and the system initial state, x_0 are assumed to be statistically independent, with $x_0 \sim \mathcal{N}(0, \sigma_0^2)$. Note that under these assumptions $\{x_n\}_{n=0}^{N}$ and $\{y_n\}_{n=0}^{N}$ are jointly Gaussian.

Due to the correlation of the measurements, one can verify that a closed-form expression to the marginal likelihood of the measurements after each step is not at hand. This fact has two consequences. First, a closed-form expression of the RUB is not available, as it is based on the derivative of the marginal likelihood of the measurements. Second, computing the MLE of a after each measurement is not feasible nor tractable. Thus, the MS-ML estimator is not obtainable. Hence, our goals are to obtain a performance bound on the MSE on the estimation of the state vector and optimal estimates of the state after each observation, in the absence of the value of a.

3. MODIFIED RUB

Following the results of [23], we present a modified version of the RUB in [22] which is based on the joint likelihood function of the measurements and the system states rather than the marginal likelihood of the measurements. The new bound is denoted by MRUB. The estimators for this problem in Section 4 do not require recomputation of the smoothed values of all the past states using the updated value of the estimate of a. Thus, the proposed bound better suits for predicting the performance of sequential estimators, as it is based on the joint likelihood function of the measurements and the system states.

Consider an unknown vector parameter $\boldsymbol{\psi} = [\varphi, \boldsymbol{\theta}_r^T, \boldsymbol{\theta}_d^T]^T$, where the parameter of interest, $\varphi \in \Omega_\varphi \subseteq \mathbb{R}$, is a random variable, and $\boldsymbol{\theta}_r \in \Omega_{\boldsymbol{\theta}_r} \subseteq \mathbb{R}^{M_r}$ and $\boldsymbol{\theta}_d \in \boldsymbol{\Theta}_d \subseteq \mathbb{R}^{M_d}$ are the vectors of random and deterministic nuisance parameters, respectively. We are interested to estimate the parameter of interest φ based on the random observation vector $\mathbf{y} \in \Omega_{\mathbf{y}}$. Let $f_{\mathbf{y},\boldsymbol{\varphi},\boldsymbol{\theta}_r}(\cdot,\cdot,\cdot;\boldsymbol{\theta}_d)$, $f_{\mathbf{y}}(\cdot;\boldsymbol{\theta}_d)$, and $f_{\boldsymbol{\varphi},\boldsymbol{\theta}_r|\mathbf{y}}(\cdot,\cdot|\mathbf{y};\boldsymbol{\theta}_d)$ denote the joint, the observation, and the posterior probability density functions (pdf's), respectively. The function $\hat{\varphi}(\mathbf{y})$ is an estimator of φ with estimation error $e_{\hat{\varphi}} = \hat{\varphi}(\mathbf{y}) - \varphi$. $\mathrm{E}_{\boldsymbol{\theta}}[\cdot]$ and $\mathrm{E}_{\boldsymbol{\theta}}[\cdot|\mathbf{y}]$ stand for the expectation operator w.r.t. $f_{\mathbf{y},\boldsymbol{\varphi},\boldsymbol{\theta}_r}(\cdot,\cdot,\cdot;\boldsymbol{\theta}_d)$ and $f_{\boldsymbol{\varphi},\boldsymbol{\theta}_r|\mathbf{y}}(\cdot,\cdot,\cdot;\boldsymbol{\theta}_d)$, respectively. The column vector of the gradient operator w.r.t. $\boldsymbol{\theta}$ is denoted by $\frac{\partial}{\partial^T\boldsymbol{\theta}}$, and the Hessian matrix operator w.r.t. $\boldsymbol{\theta}$ is denoted by $\frac{\partial^2}{\partial\boldsymbol{\theta}\partial^T\boldsymbol{\theta}}$.

Under the MSE criterion, the risk is defined as $L(\hat{\varphi}, \boldsymbol{\theta}_{d_t}) = \mathrm{E}_{\boldsymbol{\theta}_{d_t}}[e_{\hat{\varphi}}^2]$, where $\boldsymbol{\theta}_{d_t}$ denotes the true value of $\boldsymbol{\theta_d}$. If $\boldsymbol{\theta}_{d_t}$ is known, the MMSE estimator is given by the conditional mean, $\hat{\varphi}_{MS}(\mathbf{y}, \boldsymbol{\theta}_{d_t}) = \mathrm{E}_{\boldsymbol{\theta}_{d_t}}[\varphi|\mathbf{y}]$. The estimation error of the MMSE estimator is $e_{MS}(\mathbf{y}, \boldsymbol{\theta}_{d_t}) = \hat{\varphi}_{MS}(\mathbf{x}, \boldsymbol{\theta}_{d_t}) - \varphi$. If $\boldsymbol{\theta}_{d_t}$ is unknown, then $\hat{\varphi}_{MS}(\mathbf{x}, \boldsymbol{\theta}_{d_t})$ is not a valid estimator of φ and the conventional Bayesian MSE bounds are not tight.

In [25], the estimator $\hat{\varphi}(\mathbf{y})$ was said to be locally **risk-unbiased** at $\boldsymbol{\theta}_d$ if

$$\mathrm{E}_{\boldsymbol{\theta}_d}\left[z_{\hat{\varphi}}(\mathbf{y}, \boldsymbol{\theta}_d), \mathbf{d}(\mathbf{y}, \boldsymbol{\theta}_d)\right] = \mathbf{0}_M, \tag{2a}$$
$$\mathrm{E}_{\boldsymbol{\theta}_d}\left[z_{\hat{\varphi}}(\mathbf{y}, \boldsymbol{\theta}_d)\mathbf{H}(\mathbf{y}, \boldsymbol{\theta}_d)\right] = \mathbf{C_{dd}}(\boldsymbol{\theta}), \tag{2b}$$

where

$$z_{\hat{\varphi}}(\mathbf{y}, \boldsymbol{\theta}_d) \triangleq \hat{\varphi}(\mathbf{y}) - \hat{\varphi}_{MS}(\mathbf{y}, \boldsymbol{\theta}_d) \tag{3a}$$
$$\mathbf{d}(\mathbf{y}, \boldsymbol{\theta}_d) \triangleq \frac{\partial \hat{\varphi}_{MS}(\mathbf{y}, \boldsymbol{\theta}_d)}{\partial^T \boldsymbol{\theta}_d}, \tag{3b}$$
$$\mathbf{F}(\mathbf{y}, \boldsymbol{\theta}_d) \triangleq \frac{\partial^2 \hat{\varphi}_{MS}(\mathbf{y}, \boldsymbol{\theta}_d)}{\partial \boldsymbol{\theta}_d \partial^T \boldsymbol{\theta}_d}, \tag{3c}$$
$$\mathbf{H}(\mathbf{y}, \boldsymbol{\theta}_d) \triangleq \mathbf{F}(\mathbf{y}, \boldsymbol{\theta}_d) + \mathbf{d}(\mathbf{y}, \boldsymbol{\theta}_d)\frac{\partial \log f_{\mathbf{y}}(\mathbf{y}; \boldsymbol{\theta}_d)}{\partial \boldsymbol{\theta}_d}, \tag{3d}$$
$$\mathbf{C_{dd}}(\boldsymbol{\theta}_d) \triangleq \mathrm{E}_{\boldsymbol{\theta}_d}\left[\mathbf{d}(\mathbf{y}, \boldsymbol{\theta}_d)\mathbf{d}^T(\mathbf{y}, \boldsymbol{\theta}_d)\right], \tag{3e}$$

and vec$(\cdot)$ denotes the vectorization operation. Then, the following theorem provides a performance bound for risk-unbiased estimation.

Theorem 1. Let $\hat{\varphi}(\mathbf{y})$ be a locally risk-unbiased estimator of

φ. Then, under Assumptions 1-6 in [22],

$$L(\hat{\varphi}, \boldsymbol{\theta}_d) \geq B_{MRUB}(\boldsymbol{\theta}_d) \triangleq \mathrm{E}_{\boldsymbol{\theta}_d}\left[e^2_{MS}(\mathbf{y}, \boldsymbol{\theta}_d)\right] + \mathbf{c}^T_{\mathbf{d}}(\boldsymbol{\theta}_d)\left[\mathbf{C}_{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}}(\boldsymbol{\theta}_d) - \mathbf{C}_{\widetilde{\mathbf{h}}\mathbf{d}}(\boldsymbol{\theta}_d)\mathbf{C}^{-1}_{\mathbf{dd}}(\boldsymbol{\theta}_d)\mathbf{C}_{\mathbf{d}\widetilde{\mathbf{h}}}(\boldsymbol{\theta}_d)\right]^{-1}\mathbf{c}_{\mathbf{d}}(\boldsymbol{\theta}_d), \tag{4}$$

where

$$l(\mathbf{y}, \varphi, \boldsymbol{\theta}_r; \boldsymbol{\theta}_d) \triangleq \frac{\partial \log f_{\mathbf{y},\varphi,\boldsymbol{\theta}_r}(\mathbf{y}, \varphi, \boldsymbol{\theta}_r; \boldsymbol{\theta}_d)}{\partial \boldsymbol{\theta}_d}, \tag{5a}$$

$$\widetilde{\mathbf{H}}(\mathbf{y}, \boldsymbol{\theta}_r, \boldsymbol{\theta}_d) \triangleq \mathbf{F}(\mathbf{y}, \boldsymbol{\theta}_d) + \mathbf{d}(\mathbf{y}, \boldsymbol{\theta}_d) l(\mathbf{y}, \varphi, \boldsymbol{\theta}_r; \boldsymbol{\theta}_d), \tag{5b}$$

$$\widetilde{\mathbf{h}}(\mathbf{y}, \boldsymbol{\theta}_r, \boldsymbol{\theta}_d) \triangleq \mathrm{vec}\left(\mathbf{H}(\mathbf{y}, \boldsymbol{\theta}_r, \boldsymbol{\theta}_d)\right), \tag{5c}$$

$$\mathbf{c}_{\mathbf{d}}(\boldsymbol{\theta}_d) \triangleq \mathrm{vec}\left(\mathbf{C}_{\mathbf{dd}}(\boldsymbol{\theta}_d)\right), \tag{5d}$$

$$\mathbf{C}_{\widetilde{\mathbf{h}}\mathbf{d}}(\boldsymbol{\theta}_d) \triangleq \mathrm{E}_{\boldsymbol{\theta}_d}\left[\widetilde{\mathbf{h}}(\mathbf{y}, \boldsymbol{\theta}_r, \boldsymbol{\theta}_d)\mathbf{d}^T(\mathbf{y}, \boldsymbol{\theta}_r, \boldsymbol{\theta}_d)\right] \triangleq \mathbf{C}^T_{\mathbf{d}\widetilde{\mathbf{h}}}(\boldsymbol{\theta}), \tag{5e}$$

$$\mathbf{C}_{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}}(\boldsymbol{\theta}) \triangleq \mathrm{E}_{\boldsymbol{\theta}_d}\left[\widetilde{\mathbf{h}}(\mathbf{y}, \boldsymbol{\theta}_r, \boldsymbol{\theta}_d)\widetilde{\mathbf{h}}^T(\mathbf{y}, \boldsymbol{\theta}_r, \boldsymbol{\theta}_d)\right], \tag{5f}$$

Proof. *Under Assumptions 1 and 2 in [22], Lemma 2 in [28] implies that*

$$\mathrm{E}_{\boldsymbol{\theta}_d}\left[l(\mathbf{y}, \varphi, \boldsymbol{\theta}_r; \boldsymbol{\theta}_d)|\mathbf{y}\right] = \frac{\partial \log f_{\mathbf{y}}(\mathbf{y}; \boldsymbol{\theta}_d)}{\partial \boldsymbol{\theta}_d}. \tag{6}$$

Thus, using the total law of expectation, the left hand side of (2b) *takes the form*

$$\begin{aligned}\mathrm{E}_{\boldsymbol{\theta}_d}\left[z_{\hat{\varphi}}(\mathbf{y}, \boldsymbol{\theta}_d)\mathbf{H}(\mathbf{y}, \boldsymbol{\theta}_d)\right] &= \mathrm{E}_{\boldsymbol{\theta}_d}\left[\mathrm{E}_{\boldsymbol{\theta}_d}\left[z_{\hat{\varphi}}(\mathbf{y}, \boldsymbol{\theta}_d)\mathbf{H}(\mathbf{y}, \boldsymbol{\theta}_d)|\mathbf{y}\right]\right] \\ &= \mathrm{E}_{\boldsymbol{\theta}_d}\Bigg[z_{\hat{\varphi}}(\mathbf{y}, \boldsymbol{\theta}_d)\mathbf{F}(\mathbf{y}, \boldsymbol{\theta}_d) + \\ &\quad z_{\hat{\varphi}}(\mathbf{y}, \boldsymbol{\theta}_d)\mathbf{d}(\mathbf{y}, \boldsymbol{\theta}_d)\mathrm{E}_{\boldsymbol{\theta}_d}\left[\frac{\partial \log f_{\mathbf{y}}(\mathbf{y}; \boldsymbol{\theta}_d)}{\partial \boldsymbol{\theta}_d}\bigg|\mathbf{y}\right]\Bigg] = \\ &\quad \mathrm{E}_{\boldsymbol{\theta}_d}\left[z_{\hat{\varphi}}(\mathbf{y}, \boldsymbol{\theta}_d)\widetilde{\mathbf{H}}(\mathbf{y}, \boldsymbol{\theta}_r; \boldsymbol{\theta}_d)\right].\end{aligned} \tag{7}$$

The rest of the proof follows the lines of equations (9)-(13) in [25] with $\mathbf{y}$, $\boldsymbol{\theta}_d$ *and* $\widetilde{\mathbf{H}}(\mathbf{y}, \boldsymbol{\theta}_r; \boldsymbol{\theta}_d)$ *taking the role of* $\mathbf{x}$, $\boldsymbol{\theta}$, *and* $\mathbf{H}(\mathbf{y}; \boldsymbol{\theta})$*, respectively.* ■

In terms of the estimation problem in this paper, in order to obtain the MRUB for estimation of x_n from $\{y_m\}^n_{m=0}$, we set $\varphi = x_n$, $\boldsymbol{\theta}_r = [x_0, \ldots, x_{n-1}]^T$, $\theta_d = a$, and $\mathbf{y} = [y_0, \ldots, y_n]^T$. This stems from the fact that after y_n is obtained, the former system states, $\{x_m\}^{n-1}_{m=0}$, turn into random nuisance parameters. The obtained expressions for the matrices $\mathbf{C}_{\widetilde{\mathbf{d}}\mathbf{d}}(\boldsymbol{\theta}_d)$, $\mathbf{C}_{\widetilde{\mathbf{h}}\mathbf{d}}(\boldsymbol{\theta}_d)$, and $\mathbf{C}_{\widetilde{\mathbf{h}}\mathbf{h}}(\boldsymbol{\theta}_d)$ are recursive, but are omitted due to the lack of space.

4. ESTIMATION OF THE STATE

In this section, we propose three schemes for estimation of the state. According to [2], if a is known, the MMSE estimator of x_n is obtained using the following set of equations:

$$\hat{x}_{n+1} = a\hat{x}_n + K[n+1](y_{n+1} - a\hat{x}_n) \tag{8a}$$

$$K_{n+1} = (a^2 P_n + \sigma^2_u)((a^2 P_n + \sigma^2_u) + \sigma^2_v)^{-1} \tag{8b}$$

$$P_{n+1} = (1 - K_n)(a^2 P_n + \sigma^2_u), \tag{8c}$$

where $\hat{x}_n$ is the MMSE estimator of x_n given $\{y_m\}^n_{m=0}$, K_n is the Kalman gain, and P_n is the MMSE for estimation of x_n given $\{y_m\}^n_{m=0}$. If a is unknown, we wish to continue exploiting (8) by substituting an estimate of a at each step. To update the estimate of a, information from both the old state and the new observation has to be integrated. For this purpose, three adaptive approaches are suggested. The first two are based on [19]. The last approach is suggested to ensure consistency of the estimate.

1. **RLS Estimation**-The RLS estimate of a is obtained from equations (25)-(27) in [19] and is given by

$$w_n = w_{n-1} + \frac{\hat{x}^2_{n-1}}{\sigma^2_u + \sigma^2_v}, \tag{9a}$$

$$\hat{a}_{n_{RLS}} \triangleq \hat{a}_{n-1_{RLS}} + \frac{\hat{x}_{n-1}(y_n - \hat{a}_{n-1_{RLS}}\hat{x}_{n-1})}{w_n(\sigma^2_u + \sigma^2_v)}, \tag{9b}$$

where w_n is a weight factor and $\hat{a}_{n_{RLS}}$ is the RLS estimate of a given $\{y_m\}^n_{m=0}$.

2. **LMS Estimation**-The LMS estimate of a is obtained from equations (27) and (29) in [19] and is given by

$$\hat{a}_{n_{LMS}} \triangleq \hat{a}_{n-1_{RLS}} + s\hat{x}_{n-1}(y_n - \hat{a}_{n-1_{LMS}}\hat{x}_{n-1}), \tag{10}$$

where s is the step size and $\hat{a}_{n_{LMS}}$ is the LMS estimate of a given $\{y_m\}^n_{m=0}$.

3. **CC Estimation**-The correlation coefficient (CC) estimate of a is based on equation (3.13) in [16]. This equation implies that the estimation of a after each step requires recomputation of the smoothed values of all the past states. Thus, the method of [16] is computationally demanding. However, we suggest using the following substance. Let

$$p_n \triangleq p_{n-1} + y_n y_{n-1}, \tag{11a}$$

$$q_n \triangleq q_{n-1} + y^2_{n-1} - \sigma^2_v, \tag{11b}$$

$$\hat{a}_{n_{CC}} \triangleq \frac{p_n}{q_n}, \tag{11c}$$

where p and q are two recursively updated sums and $\hat{a}_{n_C C}$ is the CC estimate of a given $\{y_m\}^n_{m=0}$. It is based on the correlation coefficient of each two successive observations. Let the symbols $\underset{p}{\rightarrow}$ and $\underset{as}{\rightarrow}$ denote convergence in probability and almost sure convergence, respectively. The next proposition suggests the CC estimate is strongly consistent.

Proposition 2.

$$\hat{a}_{n_{CC}} \underset{as}{\rightarrow} a \tag{12}$$

Proof. *Equation* (1) *implies that for* $k = 1, \ldots, n-1$

$$\begin{aligned} y_{k+1}y_k =& (ax_k + u_k + v_{k+1})(x_k + v_k) \\ =& ax_k^2 + ax_kv_k + u_kx_k + u_kv_k + \\ & v_{k+1}x_k + v_{k+1}v_k, \quad (13a) \\ y_ky_k =& (x_k + v_k)^2 = x_k^2 + v_k^2 + 2x_kv_k. \quad (13b) \end{aligned}$$

Following (11a)*, by taking the time-average of* (13a) *one obtains*

$$\frac{1}{n}p_n = \frac{1}{n}\sum_{k=0}^{n-1} y_{k+1}y_k \underset{as}{\rightarrow} \frac{a}{n}\sum_{k=0}^{n-1} x_k^2, \tag{14}$$

where the convergence stems from the whiteness of $\{v_k\}_{k=0}^{n-1}$ *and* $\{u_k\}_{k=0}^{n-1}$*, from their statistical independence, and from the Gaussian character of the model in* (1)*. Similarly, following* (11b)*, by taking the time-average of* (13b) *one obtains*

$$\frac{1}{n}q_n = \frac{1}{n}\sum_{k=0}^{n-1}(y_ky_k - \sigma_v^2) \underset{as}{\rightarrow} \frac{1}{n}\sum_{k=0}^{n-1} x_k^2, \tag{15}$$

where the convergence stems from the whiteness and the Gaussian distribution of $\{v_k\}_{k=0}^{n-1}$*, which implies that* $\frac{1}{n}\sum_{k=0}^{n-1} v_k^2 \underset{as}{\rightarrow} \sigma_v^2$*. The properties of almost sure convergence in [29] implies that subtraction of* σ_v^2 *from both sides of* (15) *and division of* (14) *by the obtained expression yields* (12). ■

By substituting the estimates of a from (9b), (10), and (11) after each step into (8a) instead of a, the MS-RLS, the MS-LMS, and the MS-CC estimators are obtained.

5. EXPERIMENTAL VALIDATION

The MSE of the MS-RLS and the MS-CC, the MMSE, and the MRUB versus n are presented in Fig. 1. The MSE of the MS-LMS appeared to diverge. Thus, it is not presented. The MSE was evaluated using 10,000 Monte-Carlo simulations with $\sigma_u = 1, \sigma_v = 0.5, \sigma_0 = 1$, and $a = 1$. While the MMSE is achievable only if a is known, the MRUB provides lower bound for the MSE of the estimators for all sample sizes. Even though the system is instable, the consistency of $\hat{a}_{n_{CC}}$ yields convergence of the MSE to the MRUB. The MS-RLS yields similar results.

6. CONCLUSION

In this paper, the problem of state estimation in a linear discrete-time dynamic system with an unknown system parameter is explored. A Cramér-Rao-type bound, named MRUB, is developed for the MSE of random parameter estimation in the presence of detreministic and random nuisance parameters was . The proposed bound assumes risk-unbiasedness which is more appropriate for the case of unknown system parameters. Three adaptive sequential estimators are explored and their performance is compared to the MRUB through simulations.

The simulations illustare the potential of the MRUB and the parameter estimation methodology of this paper. As a topic of further research, the MRUB can be implemented for models of higher complexity, such as state estimation in vector models and higher order AR models.

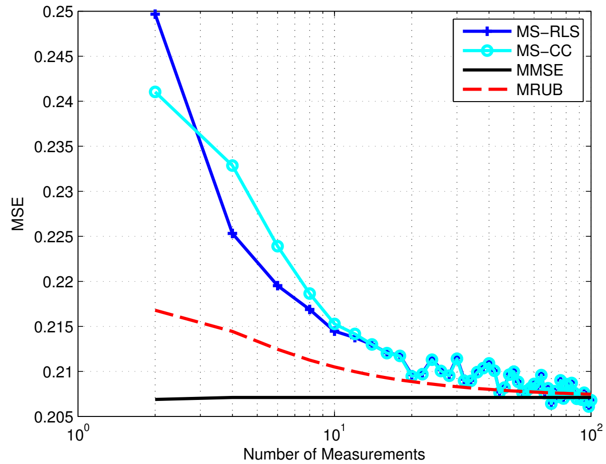

Fig. 1. The MMSE, MRUB, and MSE of MS-RLS and MS-CC.

7. ACKNOWLEDGMENTS

This research was partially supported by THE ISRAEL SCIENCE FOUNDATION (grant No. 1392/11) and by the Yaakov ben Yitzhak Hacohen scholarship.

8. REFERENCES

[1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. of the ASME Journal of Basic Engineering*, , no. 82 (Series D), pp. 35–45, 1960.

[2] T. Kailath, "An innovations approach to least-squares estimation–Part I: Linear filtering in additive white noise," *IEEE Trans. Automat. Contr.*, vol. 13, no. 6, pp. 646–655, Dec. 1968.

[3] R. K. Mehra, "On the identification of variances and adaptive Kalman filtering," *IEEE Trans. Automat. Contr.*, vol. 15, no. 2, pp. 175–184, Apr. 1970.

[4] D. S. Spain, "Identification and modelling of discrete stochastic linear systems," in *Proc. CDC 1972 and 11th Symposium on Adaptive Processes*, Dec. 1972, pp. 386–390.

[5] R. K. Mehra, "Identification of stochastic linear dynamic systems using Kalman filter representation," *AIAA Journal*, vol. 9, no. 1, pp. 28–31, Jan. 1971.

[6] C. G. Hilborn and D. G. Lainiotis, "Optimal estimation in the presence of unknown parameters," *IEEE Trans. Syst. Sci. Cybernetics*, vol. 5, no. 1, pp. 38–43, Jan. 1969.

[7] C. G. Hilborn and D. G. Lainiotis, "Optimal adaptive filter realizations for sample stochastic processes with an unknown parameter," *IEEE Trans. Automat. Contr.*, vol. 14, no. 6, pp. 767–770, Dec. 1969.

[8] H. Cox, "On the estimation of state variables and parameters for noisy dynamic systems," *IEEE Tran. Automat. Contr.*, vol. 9, no. 1, pp. 5–12, Jan. 1964.

[9] A. V. Balakrishnan, "Stochastic system identification techniques," in *Stochastic Optimization and Control*, H. F. Karreman, Ed., vol. 9, pp. 65–89. Wiley, New York, Jan. 1968.

[10] M. J. Levin, "Estimation of a system pulse transfer function in the presence of noise," *IEEE Tran. Automat. Contr.*, vol. 9, no. 3, pp. 229–235, Jul. 1964.

[11] Karl-Johan Åström and Torsten Bohlin, "Numerical identification of linear dynamic systems from normal operating records," in *Theory of Self-Adaptive Control Systems*, P.H. Hammond, Ed., pp. 96–111. Springer US, 1966.

[12] A. Rogers and K. Steiglitz, "Maximum likelihood estimation of rational transfer function parameters," *IEEE Tran. Automat. Contr.*, vol. 12, no. 5, pp. 594–597, Oct. 1967.

[13] M. Aoki and P. Yue, "On certain convergence questions in system identification," *SIAM Journal on Control*, vol. 8, no. 2, pp. 239–256, 1970.

[14] R. Kashyap, "Maximum likelihood identification of stochastic linear systems," *IEEE Tran. Automat. Contr.*, vol. 15, no. 1, pp. 25–34, Feb. 1970.

[15] R. K. Mehra, "On-line identification of linear dynamic systems with applications to Kalman filtering," *IEEE Trans. Automat. Contr.*, vol. 16, no. 1, pp. 12–21, Feb. 1971.

[16] Y. Bar-Shalom, "Optimal simultaneous state estimation and parameter identification in linear discrete-time systems," *IEEE Trans. Autom. Control*, vol. 17, no. 3, pp. 308–319, Jun. 1972.

[17] A Yeredor and E. Weinstein, "The extended least-squares and the joint maximum-a-posteriori maximum-likelihood estimation criteria," in *Proc. ICASSP 1999*, Mar. 1999, vol. 4, pp. 1813–1816.

[18] A. Yeredor, "The joint MAP-ML criterion and its relation to ML and to extended least-squares," *IEEE Trans. Signal Processing*, vol. 48, no. 12, pp. 3484–3492, Dec. 2000.

[19] S. Korl, H.-A Loeliger, and Allen G. Lindgren, "AR model parameter estimation: from factor graphs to algorithms," in *Proc. ICASSP 2004*, May 2004, vol. 5, pp. 509–512.

[20] T. Söderström and P. Stoica, *System identification*, Prentice-Hall, Inc., 1988.

[21] S. Haykin and B. Widrow, *Least-mean-square adaptive filters*, vol. 31, John Wiley & Sons, 2003.

[22] S. Bar and J. Tabrikian, "Bayesian estimation in the presence of deterministic nuisance parameters—Part I: Performance bounds," submitted for publication in the *IEEE Trans. Signal Processing*, 2014.

[23] S. Bar and J. Tabrikian, "Bayesian estimation in the presence of deterministic nuisance parameters—Part II: Risk-unbiased estimators," submitted for publication in the *IEEE Trans. Signal Processing*, 2014.

[24] T. K. Nayak and M. Qin, "The concept of risk unbiasedness in statistical prediction," *Journal of Statistical Planning and Inference*, vol. 140, no. 7, pp. 1923–1938, Jul. 2010.

[25] S. Bar and J. Tabrikian, "Bayesian Cramér-Rao type bound for risk-unbiased estimation with deterministic nuisance parameters," in *Proc. ICASSP 2014*, May 2014, pp. 2962–2966.

[26] Y. Noam and H. Messer, "The hybrid Cramér-Rao bound and the generalized gaussian linear estimation problem," in *Sensor Array and Multichannel Signal Processing Workshop, 2008. SAM 2008. 5th IEEE*, Jul. 2008, pp. 395–399.

[27] H. Messer, "The hybrid Cramér-Rao lower bound-from practice to theory," in *Proc. 2006 Fourth IEEE Workshop on Sensor Array and Multichannel Processing*, Jul. 2006, pp. 304–307.

[28] Y. Noam and H. Messer, "Notes on the tightness of the hybrid Cramér-Rao lower bound," *IEEE Trans. Signal Processing*, vol. 57, no. 6, pp. 2074–2084, Jun. 2009.

[29] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge University Press, 1998.