Efficient Recursive Estimators for a Linear, Time-Varying Gaussian Model with General Constraints

Stefan Uhlich*, *Student Member, IEEE,* and Bin Yang, *Senior Member, IEEE,*

Abstract

The adaptive estimation of a time-varying parameter vector in a linear Gaussian model is considered where we a priori know that the parameter vector belongs to a known arbitrary subset. We consider a family of efficient recursive estimators for this problem: the recursive constrained maximum likelihood (ML) estimator, the recursive affine minimax, and the recursive minimum mean squared error (MMSE) estimator. We show that all three estimators can be substantially simplified by using the recursive weighted least squares (RWLS) algorithm in a first step as the RWLS computes the sufficient statistic for this estimation problem. The recursive constrained ML needs to solve an optimization problem in the second step for the case that the RWLS solution does not fulfill the constraint. In case of affine minimax, we have to solve an optimization problem and to perform an affine transform. The MMSE estimator needs to calculate the mean of a truncated Gaussian density in the second step which is done by Monte Carlo integration. A simple rejection scheme is used to take general constraints for the parameter vector into account. An example shows the superior performance of our proposed estimators in comparison to many other estimators.

Index Terms

Recursive constrained maximum likelihood, Recursive affine minimax, Recursive minimum mean squared error, Tracking, Sufficient statistic, Recursive weighted least squares

S. Uhlich and B. Yang are with the Chair of System Theory and Signal Processing at the Universität Stuttgart, Pfaffenwaldring 47, 70550 Stuttgart, Germany. (E-mail: stefan.uhlich@Lss.uni-stuttgart.de; bin.yang@Lss.uni-stuttgart.de)

I. INTRODUCTION

In this correspondence, we study the estimation of an unknown, time-variant parameter vector

$$\underline{\theta}_0(n) \in \Theta \subset \mathbb{R}^M \tag{1}$$

from observations $\underline{x}(n) \in \mathbb{R}^K$ of a linear, time-variant Gaussian model

$$\underline{x}(n) = \mathbf{H}(n)\underline{\theta}_0(n) + \underline{z}(n), \quad n \geq 0. \tag{2}$$

$\mathbf{H}(n) \in \mathbb{R}^{K\times M}$ is a known model matrix, $\underline{z}(n)$ is Gaussian noise with $\underline{z}(n) \sim \mathcal{N}(\underline{0}, \mathbf{C}(n))$ which is temporally uncorrelated, i.e. $\mathrm{E}[\underline{z}(n_1)\underline{z}(n_2)^T] = \mathbf{0}$ for all $n_1 \neq n_2$, and n denotes the discrete time. In addition, we assume to know a priori $\underline{\theta}_0(n) \in \Theta$, where Θ is an arbitrary subset of $\mathbb{R}^M$. Thus, the task is to estimate $\underline{\theta}_0(n)$ for all $n \geq 0$ subject to $\underline{\theta}_0(n) \in \Theta$ given all observations $\underline{x}(0), \ldots, \underline{x}(n)$.

Such a constrained time-varying parameter estimation problem is interesting for many applications. The constraints may either naturally arise from physical problems or are artificially introduced, e.g. in regularized regression problems. An example of natural constraints is a power limitation on $\underline{\theta}_0(n)$, i.e. $\|\underline{\theta}_0(n)\|^2 \leq E_{\text{max}}$ or an energy passivity constraint for the estimation of a channel impulse response. Another important class of constraints are l_1-norm constraints that are used to obtain sparse results. In [1], for example, Tibshirani introduced the LASSO (least absolute shrinkage and selection operator) which is a least squares approach with the l_1-norm constraint $\Theta = \{\underline{\theta}_0 : \|\underline{\theta}_0\|_{l_1} \leq t\}$.

The aim of this correspondence is to derive and compare three recursive estimators that reduce the mean squared error (MSE) for the estimation of $\underline{\theta}_0(n)$ from the observations $\underline{x}(0), \ldots, \underline{x}(n)$ by exploiting the a priori knowledge $\underline{\theta}_0(n) \in \Theta$. The first estimator is the constrained maximum likelihood (CML) estimator which maximizes the likelihood function over Θ. This estimator was e.g. used in [2], [3] for linear equality/inequality and quadratic constraints and in [4] for a l_1-norm constraint. The second estimator we consider is the affine minimax (AMX) estimator [5]–[7] which has attracted a lot of interest in the past. The AMX minimizes the worst-case MSE for $\underline{\theta}_0(n) \in \Theta$. So far, the AMX has either only be considered for a time-invariant estimation problem or for the blind minimax tracking problem [8] and we will extend the AMX to the problem given by (1) and (2). The third estimator is the Bayesian minimum mean squared error (MMSE) estimator with a uniform prior on Θ [9]–[11] which we formulate in a recursive way.

The main challenge is to apply the three estimators (CML, AMX, and MMSE) to a time-varying estimation problem with an increasing number of observations $\underline{x}(0), \ldots, \underline{x}(n)$. A direct application of these estimators would lead to a growing computational complexity as the time n proceeds. The use of a suitable sufficient statistic transforms the original estimation problem to an equivalent one with a fixed number of "observations", thus resulting in a fixed computational complexity at each time step n. It will turn out that the recursive weighted least squares (RWLS) algorithm provides an efficient way to calculate this sufficient statistic. Hence, the correspondence provides a uniform framework that allows the constrained tracking of $\underline{\theta}_0(n)$ from an increasing number of observations $\underline{x}(0), \ldots, \underline{x}(n)$. This unified framework, which has not been considered in the literature before, is the main contribution of this correspondence.

The correspondence is organized as follows: Sec. II-A formulates the time-varying estimation problem we consider in this correspondence. The sufficient statistic is derived in Sec. II-B. We show that the RWLS algorithm is an efficient way to compute the sufficient statistic. Sec. III is the main part of the correspondence. There, we consider the CML estimator and also generalize the concepts of AMX and MMSE estimation to the time-variant signal model (2). We show that all three estimators can be efficiently calculated by using RWLS as a preprocessing step. Finally, Sec. IV shows some simulation results and compares the performance of the estimators.

The following notations are used throughout this correspondence: $\underline{x}$ denotes a vector, $\mathbf{X}$ denotes a matrix and $\mathbf{I}$ is the identity matrix. $\mathbf{X} \succeq \mathbf{0}$ ($\mathbf{X} \succ \mathbf{0}$) means that $\mathbf{X}$ is symmetric and non-negative (positive) definite. $\underline{x} = \text{vec}\{\mathbf{X}\}$ is the vector we obtain after stacking the columns of $\mathbf{X}$.

II. Signal Model and Sufficient Statistics

A. Signal Model

The aim of this correspondence is to derive recursive estimators for the estimation of $\underline{\theta}_0(n)$ in (2) from the observations $\underline{x}(0), \ldots, \underline{x}(n)$ subject to the a priori constraint $\underline{\theta}_0(n) \in \Theta$. Assume that $\underline{\theta}_0(n)$ is slowly time-varying, such an adaptive estimation (tracking) is feasible. Given the instantaneous signal model (2) without any constraint on $\underline{\theta}_0(n)$, the following exponentially weighted least squares (WLS) estimator is popular in adaptive filtering (see [12], [13])

$$\hat{\underline{\theta}}_{\text{WLS}}(n) = \arg\min_{\underline{\theta}(n)} \sum_{i=0}^{n} \beta^{n-i} \left(\underline{x}(i) - \mathbf{H}(i)\underline{\theta}(n)\right)^T \mathbf{C}(i)^{-1} \left(\underline{x}(i) - \mathbf{H}(i)\underline{\theta}(n)\right) \tag{3}$$

where $0 < \beta \leq 1$ is the exponential forgetting factor. This estimator can be efficiently calculated by the recursive weighted least squares (RWLS) algorithm, see Sec. II-B. The solution $\hat{\underline{\theta}}_{\text{WLS}}(n)$ of (3) can also be interpreted as the ML estimator of the following stacked signal model

$$\underline{x}_n = \mathbf{H}_n \underline{\theta}(n) + \underline{z}_n. \tag{4}$$

$\underline{x}_n$, $\mathbf{H}_n$, and $\underline{z}_n$ are stacked versions of $\underline{x}(i)$, $\mathbf{H}(i)$, and $\underline{z}(i)$ for $i = 0, \ldots, n$, respectively:

$$\begin{aligned} \underline{x}_n &= \left[\underline{x}(n)^T \quad \cdots \quad \underline{x}(0)^T\right]^T, \\ \mathbf{H}_n &= \left[\mathbf{H}(n)^T \quad \cdots \quad \mathbf{H}(0)^T\right]^T, \\ \underline{z}_n &= \left[\underline{z}(n)^T \quad \beta^{-1/2}\underline{z}(n-1)^T \quad \cdots \quad \beta^{-n/2}\underline{z}(0)^T\right]^T. \end{aligned} \tag{5}$$

Note that the noise term $\underline{z}_n$ is not merely the stacking of $\underline{z}(i)$, but also incorporates an intentional weighting of $\underline{z}(i)$ with $\beta^{-(n-i)/2}$. This results in a tendentiously larger noise (as it is in reality) for past measurements, marking them to be less valuable than recent observations for the estimation of $\underline{\theta}_0(n)$. The effect is exactly the same as the exponential down-weighting of error terms caused by past observations in (3). As the noise $\underline{z}(n)$ is temporally uncorrelated, we have $\underline{z}_n \sim \mathcal{N}(\underline{0}, \mathbf{C}_n)$ where the covariance matrix $\mathbf{C}_n \in \mathbb{R}^{(n+1)K \times (n+1)K}$ is a block diagonal matrix

$$\mathbf{C}_n = \operatorname{diag}\left(\mathbf{C}(n), \beta^{-1}\mathbf{C}(n-1), \cdots, \beta^{-n}\mathbf{C}(0)\right). \tag{6}$$

The unconstrained maximum likelihood estimate to (4) is given by

$$\min_{\underline{\theta}(n)} \left(\underline{x}_n - \mathbf{H}_n\underline{\theta}(n)\right)^T \mathbf{C}_n^{-1} \left(\underline{x}_n - \mathbf{H}_n\underline{\theta}(n)\right). \tag{7}$$

It is identical to the WLS estimator (3). Using the stacked matrices $\mathbf{H}_n$, $\mathbf{C}_n$ and $\underline{x}_n$ defined in (5) and (6), the solution to (7) can be written as

$$\hat{\underline{\theta}}_{\text{WLS}}(n) = \mathbf{R}_n^{-1}\mathbf{H}_n^T\mathbf{C}_n^{-1}\underline{x}_n \quad \text{with} \quad \mathbf{R}_n = \mathbf{H}_n^T\mathbf{C}_n^{-1}\mathbf{H}_n \tag{8}$$

if $\mathbf{C}_n$ is invertible and $\mathbf{H}_n$ has a full column rank. It is obvious that $\underline{\theta}(n)$ in (4) is different from $\underline{\theta}_0(n)$ in (2) as (4) is an approximation of the original signal model (2) for a slowly varying parameter $\underline{\theta}_0(n)$. This approximation, however, has proved to be useful in many applications. If $\underline{\theta}_0(n) = \underline{\theta}_0$ is constant, $\hat{\underline{\theta}}_{\text{WLS}}(n)$ will approach $\underline{\theta}_0$ with increasing n. For the case that $\underline{\theta}_0(n)$ changes slowly, $\hat{\underline{\theta}}_{\text{WLS}}(n)$ will follow $\underline{\theta}_0(n)$. Therefore, we will use the stacked signal model (4) below and consider in addition the a priori constraint (1). We look for estimators for the signal model

$$\underline{x}_n = \mathbf{H}_n\underline{\theta}(n) + \underline{z}_n, \quad \underline{\theta}(n) \in \Theta \tag{9}$$

with a smaller MSE than the previous unconstrained WLS estimator in (8). Note that (9) describes a frequentist estimation problem as $\underline{\theta}(n)$ is deterministic. The length of $\underline{x}_n$, $\mathbf{H}_n$ and $\mathbf{C}_n$ grows linearly as the time n proceeds.

B. Sufficient Statistics

In order to avoid estimators with a growing computational complexity, we first perform a signal model transform by using the concept of sufficient statistic [14]. Let $\underline{x}$ be the random observation about the unknown, deterministic parameter $\underline{\theta}$. Furthermore, let $\underline{t}(\underline{x})$ be a statistic of $\underline{x}$. Then $\underline{t}(\underline{x})$ is a sufficient statistic if $p(\underline{x}|\underline{t}(\underline{x}) = \underline{t}_0, \underline{\theta}) = p(\underline{x}|\underline{t}(\underline{x}) = \underline{t}_0)$. Therefore, given $\underline{t}(\underline{x}) = \underline{t}_0$, we can not infer additional information about $\underline{\theta}$ from the observation $\underline{x}$. A convenient way to find the sufficient statistic is to use the Fisher-Neyman factorization theorem: A sufficient and necessary condition for $\underline{t}(\underline{x})$ being a sufficient statistic for $\underline{\theta}$ is the factorization

$$p(\underline{x}|\underline{\theta}) = g(\underline{t}(\underline{x}), \underline{\theta})h(\underline{x}). \tag{10}$$

For our linear signal model in additive Gaussian noise

$$\underline{x} = \mathbf{H}\underline{\theta} + \underline{z} \quad \text{with} \quad \underline{z} \sim \mathcal{N}(\underline{0}, \mathbf{C}), \tag{11}$$

a well-known sufficient statistic is $\underline{t}(\underline{x}) = \mathbf{H}^T\mathbf{C}^{-1}\underline{x}$ [15]. As any one-to-one function of $\underline{t}(\underline{x})$ is also a sufficient statistic, we use the following sufficient statistic

$$\underline{t}(\underline{x}) = \mathbf{Q}\mathbf{H}^T\mathbf{C}^{-1}\underline{x} \tag{12}$$

where $\mathbf{Q}$ is any square invertible matrix. In other words, the transform $\mathbf{T}^T\underline{x}$ of the signal model in (11) with $\mathbf{T} = \mathbf{C}^{-1}\mathbf{H}\mathbf{Q}^T$ does not result in an information loss for the estimation of $\underline{\theta}$. The new signal model is now

$$\underline{\tilde{x}} = \mathbf{T}^T\underline{x} = \mathbf{T}^T\mathbf{H}\underline{\theta} + \mathbf{T}^T\underline{z} = \tilde{\mathbf{H}}\underline{\theta} + \underline{\tilde{z}}. \tag{13}$$

We make use of this fact with the special choice of $\mathbf{Q} = \left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}$. This yields the transformed signal model $\underline{\tilde{x}} = \underline{\theta} + \underline{\tilde{z}}$ with $\underline{\tilde{z}} \sim \mathcal{N}(\underline{0}, \tilde{\mathbf{C}})$ and the new covariance matrix $\tilde{\mathbf{C}} = \mathbf{T}^T\mathbf{C}\mathbf{T} = \left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}$. Therefore, instead of the stacked signal model (9), we can look for estimators based on the simpler model $\underline{\tilde{x}}_n = \underline{\theta}(n) + \underline{\tilde{z}}_n$ with $\underline{\tilde{z}}_n \sim \mathcal{N}(\underline{0}, \tilde{\mathbf{C}}_n)$. The transformed observation vector $\underline{\tilde{x}}_n$ and the new covariance matrix $\tilde{\mathbf{C}}_n$ are given by

$$\underline{\tilde{x}}_n = \underline{t}(\underline{x}) = \left(\mathbf{H}_n^T\mathbf{C}_n^{-1}\mathbf{H}_n\right)^{-1}\mathbf{H}_n^T\mathbf{C}_n^{-1}\underline{x}_n = \underline{\hat{\theta}}_{\text{WLS}}(n), \tag{14a}$$

$$\tilde{\mathbf{C}}_n = \left(\mathbf{H}_n^T\mathbf{C}_n^{-1}\mathbf{H}_n\right)^{-1} = \mathbf{R}_n^{-1}. \tag{14b}$$

where $\underline{\hat{\theta}}_{\text{WLS}}(n)$ and $\mathbf{R}_n$ are defined in (8). The main advantage of this model transform is the fixed vector length M of $\underline{\tilde{x}}_n$ in comparison to the linearly growing vector length $(n+1)K$ of $\underline{x}_n$.

An efficient way to compute the sufficient statistic is to use the RWLS algorithm. According to the definition of $\mathbf{H}_n$ in (5) and $\mathbf{C}_n$ in (6), both matrices can be written recursively

$$\mathbf{H}_{n+1} = \begin{bmatrix} \mathbf{H}(n+1) \\ \mathbf{H}_n \end{bmatrix}, \qquad \mathbf{C}_{n+1} = \begin{bmatrix} \mathbf{C}(n+1) & \mathbf{0} \\ \mathbf{0} & \beta^{-1}\mathbf{C}_n \end{bmatrix}.$$

This also implies

$$\begin{aligned} \mathbf{R}_{n+1} &= \mathbf{H}_{n+1}^T \mathbf{C}_{n+1}^{-1} \mathbf{H}_{n+1} \\ &= \beta \mathbf{R}_n + \mathbf{H}(n+1)^T \mathbf{C}(n+1)^{-1} \mathbf{H}(n+1). \end{aligned} \tag{15}$$

By applying the matrix inversion lemma $(\mathbf{A}+\mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{D}\mathbf{A}^{-1}$, we can update $\mathbf{R}_n^{-1}$ and the parameter estimate $\hat{\underline{\theta}}_{\text{WLS}}(n)$ in (8) in a time recursive way. Introducing the gain matrix $\mathbf{G}_{n+1} = \mathbf{R}_{n+1}^{-1}\mathbf{H}(n+1)^T\mathbf{C}(n+1)^{-1}$, we finally obtain the RWLS algorithm

$$\hat{\underline{\theta}}_{\text{WLS}}(n+1) = \hat{\underline{\theta}}_{\text{WLS}}(n) + \mathbf{G}_{n+1}\left(\underline{x}(n+1) - \mathbf{H}(n+1)\hat{\underline{\theta}}_{\text{WLS}}(n)\right), \tag{16a}$$

$$\mathbf{R}_{n+1}^{-1} = \frac{1}{\beta}\left(\mathbf{R}_n^{-1} - \mathbf{G}_{n+1}\mathbf{H}(n+1)\mathbf{R}_n^{-1}\right). \tag{16b}$$

At each time step, RWLS updates the sufficient statistic $\hat{\underline{\theta}}_{\text{WLS}}(n)$ and the inverse correlation matrix $\mathbf{R}_n^{-1}$. Both values are needed by the three recursive estimators below subject to the constraint $\underline{\theta}(n) \in \Theta$.

III. Adaptive Estimation for a Linear Time-Variant Gaussian Model with General Constraints

A. Recursive Constrained Maximum Likelihood

The first estimator we consider is the constrained maximum likelihood (CML) estimator [16]. Given the linear signal model $\underline{x} = \mathbf{H}\underline{\theta} + \underline{z}$, $\underline{z} \sim \mathcal{N}(\underline{0}, \mathbf{C})$ with the constraint $\underline{\theta} \in \Theta$, the CML is given by the optimization problem

$$\hat{\underline{\theta}}_{\text{CML}} = \arg\max_{\underline{\theta}} p(\underline{x}|\underline{\theta}) \quad \text{s.t.} \quad \underline{\theta} \in \Theta \tag{17}$$

where $p(\underline{x}|\underline{\theta})$ is the probability density function (PDF) of the observation $\underline{x}$. Using the sufficient statistic $\hat{\underline{\theta}}_{\text{WLS}}$ in (12), we can rewrite (17) into

$$\hat{\underline{\theta}}_{\text{CML}} = \arg\min_{\underline{\theta}} (\underline{\theta} - \hat{\underline{\theta}}_{\text{WLS}})^T (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})(\underline{\theta} - \hat{\underline{\theta}}_{\text{WLS}}) \quad \text{s.t.} \quad \underline{\theta} \in \Theta. \tag{18}$$

As $\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H} \succ \mathbf{0}$, we have to check whether $\hat{\underline{\theta}}_{\text{WLS}}$ satisfies the constraint or not. If $\hat{\underline{\theta}}_{\text{WLS}} \in \Theta$ then $\hat{\underline{\theta}}_{\text{CML}} = \hat{\underline{\theta}}_{\text{WLS}}$. Otherwise we have to find the minimum of $(\underline{\theta} - \hat{\underline{\theta}}_{\text{WLS}})^T (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})(\underline{\theta} - \hat{\underline{\theta}}_{\text{WLS}})$ on the boundary of Θ.

The extension of the CML to our time-varying signal model (9) is now straightforward. All we have to do is to replace $\hat{\underline{\theta}}_{\text{WLS}}$, $\mathbf{C}$ and $\mathbf{H}$ by $\hat{\underline{\theta}}_{\text{WLS}}(n)$, $\mathbf{C}_n$ and $\mathbf{H}_n$ in (18). Therefore, for each time step we have to check $\hat{\underline{\theta}}_{\text{WLS}}(n) \in \Theta$. If this is satisfied then $\hat{\underline{\theta}}_{\text{RCML}}(n) = \hat{\underline{\theta}}_{\text{WLS}}(n)$. Otherwise, we have to find the minimum of $(\underline{\theta} - \hat{\underline{\theta}}_{\text{WLS}}(n))^T(\mathbf{H}_n^T\mathbf{C}_n^{-1}\mathbf{H}_n)(\underline{\theta} - \hat{\underline{\theta}}_{\text{WLS}}(n))$ on the boundary of Θ. We call this estimator the *recursive constrained ML (RCML)* estimator.

B. Recursive Affine Minimax Estimation

The second estimator we consider is the affine minimax (AMX) estimator in the form $\hat{\underline{\theta}}_{\text{AMX}} = \mathbf{U}\underline{x} + \underline{u}$. Here $\underline{\theta}$ is considered to be a deterministic unknown parameter vector from $\Theta \subset \mathbb{R}^M$. The matrix $\mathbf{U}$ and the offset $\underline{u}$ are chosen such to minimize the maximum (worst-case) MSE of $\hat{\underline{\theta}}_{\text{AMX}}$ for all $\underline{\theta} \in \Theta$.

The motivation for the use of the affine minimax estimator is as follows. In contrast to the traditional concept of Cramer-Rao bound (CRB) for the variance of unbiased estimators and efficient estimators achieving this bound, Eldar recently proposed a framework to derive estimators achieving the biased Cramer-Rao bound which is actually an MSE bound [6], [7], [17]. This bound can be made, by allowing for a suitable bias $\underline{b}$, often smaller than the CRB for all $\underline{\theta} \in \Theta$. In particular, Eldar presented a linear bias $\mathbf{M}\underline{\theta}$ in [6] and extended this idea to an affine bias $\mathbf{M}\underline{\theta} + \underline{u}$ in [7]. Moreover, it was shown that, given an unbiased and efficient estimator $\hat{\underline{\theta}}_{\text{eff}}$, the following affine transform of $\hat{\underline{\theta}}_{\text{eff}}$

$$\hat{\underline{\theta}}_{\text{AMX}} = (\mathbf{I} + \mathbf{M})\hat{\underline{\theta}}_{\text{eff}} + \underline{u} \tag{19}$$

with the bias $\mathbf{M}\underline{\theta} + \underline{u}$ achieves the MSE bound and has thus, if $\mathbf{M} \neq \mathbf{0}$ or $\underline{u} \neq \underline{0}$, a smaller MSE than $\hat{\underline{\theta}}_{\text{eff}}$ for all $\underline{\theta} \in \Theta$. The unknown matrix $\mathbf{M}$ and vector $\underline{u}$ can be determined by solving the minimax problem [7]

$$\min_{\mathbf{M},\underline{u}} \max_{\underline{\theta}\in\Theta} \quad \|\mathbf{M}\underline{\theta} + \underline{u}\|^2 + \operatorname{tr}\left\{(\mathbf{I}+\mathbf{M})\mathbf{J}(\underline{\theta})^{-1}(\mathbf{I}+\mathbf{M})^T\right\} - \operatorname{tr}\left\{\mathbf{J}(\underline{\theta})^{-1}\right\}, \tag{20}$$

where $\mathbf{J}(\underline{\theta})^{-1}$ is the inverse Fisher information matrix, i.e. the CRB of $\underline{\theta}$.

For a linear Gaussian model $\underline{x} = \mathbf{H}\underline{\theta} + \underline{z}$ with $\underline{z} \sim \mathcal{N}(\underline{0}, \mathbf{C})$ as in our case, the ML estimate

$$\hat{\underline{\theta}}_{\text{eff}} = \left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{C}^{-1}\underline{x} \tag{21}$$

is known to be unbiased and efficient. Since $\hat{\underline{\theta}}_{\text{eff}}$ is a linear function of $\underline{x}$, $\hat{\underline{\theta}}_{\text{AMX}}$ in (19) becomes an affine estimator

$$\hat{\underline{\theta}}_{\text{AMX}} = (\mathbf{I} + \mathbf{M})\left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{C}^{-1}\underline{x} + \underline{u}. \tag{22}$$

The Fisher information matrix for $\underline{\theta}$ in this case is $\mathbf{J}(\underline{\theta}) = \mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}$. For the calculation of the unknown matrix $\mathbf{M}$ and vector $\underline{u}$ under a special quadratic constraint $\Theta = \{\underline{\theta} : \underline{\theta}^T\mathbf{A}\underline{\theta} + 2\underline{b}^T\underline{\theta} + c \leq 0\}$, we can use a result from [7]. There, a SDP formulation is given and using standard convex optimization solvers, $\mathbf{M}$ and $\underline{u}$ can be calculated efficiently. For other constraints, standard techniques from convex optimization like Finsler's lemma or the S-procedure can be used [18], [19] after the problem has been rewritten in epigraphic form. However, the calculation of $\mathbf{M}$ and $\underline{u}$ is in general, i.e. for an arbitrary and possibly non-convex Θ as in this correspondence, difficult.

The extension of the AMX to our time-variant linear Gaussian model is straightforward. The unbiased efficient estimator $\hat{\underline{\theta}}_{\text{eff}}(n)$ in (21) is identical to $\hat{\underline{\theta}}_{\text{WLS}}(n)$ in (8) and can be calculated efficiently by the RWLS algorithm. After each update of $\hat{\underline{\theta}}_{\text{eff}}(n)$, we multiply it with $\mathbf{I}+\mathbf{M}(n)$ and add $\underline{u}(n)$ to obtain the *recursive AMX* (RAMX) estimator

$$\hat{\underline{\theta}}_{\text{RAMX}}(n) = (\mathbf{I} + \mathbf{M}(n))\,\hat{\underline{\theta}}_{\text{eff}}(n) + \underline{u}(n). \tag{23}$$

We have to solve the minimax problem (20) to obtain $\mathbf{M}(n)$ and $\underline{u}(n)$. The inverse Fisher information matrix $\mathbf{J}(\underline{\theta}(n))^{-1}$ in (20) is given by $\mathbf{R}_n^{-1} = \left(\mathbf{H}_n^T\mathbf{C}_n^{-1}\mathbf{H}_n\right)^{-1}$ which is also updated by the RWLS algorithm. Since $\mathbf{R}_n^{-1}$ is time-varying, the minimax problem (20) has to be solved repeatedly at each time step.

In the stationary case with $\mathbf{H}(n) = \mathbf{H}$ and $\mathbf{C}(n) = \mathbf{C}$, we can avoid the repeated solution of the minimax problem (20) by approximating $\mathbf{R}_n$ by its steady state value

$$\mathbf{R}_\infty = \lim_{n\to\infty}\mathbf{R}_n = \lim_{n\to\infty}\sum_{k=0}^{n}\beta^{n-k}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H} = \frac{1}{1-\beta}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H},\ (0<\beta<1). \tag{24}$$

In this case, we only need to compute $\mathbf{M}$ and $\underline{u}$ once from $\mathbf{R}_\infty$ and use them to change the efficient estimator $\hat{\underline{\theta}}_{\text{eff}}(n)$ to the RAMX estimator $\hat{\underline{\theta}}_{\text{RAMX}}(n)$ according to (23). This approach reduces the computational complexity of RAMX at the expense of an increased mean squared error for small n.

C. Recursive MMSE Estimation

The third approach we use after the calculation of the sufficient statistic is the MMSE estimation. In contrast to the two frequentist approaches (ML and AMX), we now assume $\underline{\theta}$ to be a random vector and recast the constraint $\underline{\theta} \in \Theta$ as an a priori probability density function (PDF) $p(\underline{\theta})$. In this way, we have reformulated the frequentist estimation problem ($\underline{\theta}$ deterministic) to a Bayesian one. The MMSE estimator for $\underline{\theta} \in \Theta \subset \mathbb{R}^M$ is given by [14]

$$\hat{\underline{\theta}}_{\text{MMSE}} = \frac{\int\limits_{\Theta}\underline{\theta}\,p(\underline{x}|\underline{\theta})p(\underline{\theta})d\underline{\theta}}{\int\limits_{\Theta}p(\underline{x}|\underline{\theta})p(\underline{\theta})d\underline{\theta}} \tag{25}$$

where $p(\underline{x}|\underline{\theta})$ is the likelihood function of the observation $\underline{x} \in \mathbb{R}^K$ conditioned on $\underline{\theta}$. We assume a uniform a priori PDF (a natural choice as we have no information about $\underline{\theta}$ except for $\underline{\theta} \in \Theta$)

$$p(\underline{\theta}) = \begin{cases} \text{const} & \underline{\theta} \in \Theta \\ 0 & \text{otherwise} \end{cases}. \tag{26}$$

This estimator is sometimes called the mean likelihood estimator (MELE) [9], [10] as (25) for this special prior calculates the mean of the normalized likelihood function $p(\underline{x}|\underline{\theta})/\int_\Theta p(\underline{x}|\underline{\theta})d\underline{\theta}$. The main advantage of MMSE is that it gives the best results with respect to a minimum MSE. However, the disadvantage of MMSE is the calculation of two multidimensional integrals. A closed-form calculation of these Bayesian integrals is in general, i.e. for arbitrary $p(\underline{x}|\underline{\theta})$ and/or arbitrary $\Theta \subset \mathbb{R}^M$, impossible. A numerical integration is only feasible for a small value of M as well as a small number of observations K and low signal-to-noise ratio (SNR) because otherwise a sharp peak of $p(\underline{x}|\underline{\theta})$ would make the numerical integration difficult. Beside the numerical integration, some other methods have been proven to be useful [20], [21]. Later in this correspondence, we will use Monte Carlo integration to approximate the MMSE estimate.

To derive the MMSE estimator for the time-varying model (9), we make use of the Bayesian sufficient statistic [15], [22]. Let $\underline{x}$ be the observation about the unknown, random parameter $\underline{\theta}$. Furthermore, let $\underline{t}(\underline{x})$ be a statistic of $\underline{x}$. Then $\underline{t}(\underline{x})$ is a Bayesian sufficient statistic if for all priors $p(\underline{\theta})$ the a posteriori density $p(\underline{\theta}|\underline{x})$ can be written as $p(\underline{\theta}|\underline{x}) = p(\underline{\theta}|\underline{t}(\underline{x}))$. This concept was introduced by Kolmogorov in [23] and is called Bayesian sufficiency. Therefore, given $\underline{t}(\underline{x})$, we can not infer additional information about $\underline{\theta}$ from the observations $\underline{x}$. The concept of Bayesian sufficiency is directly related to the classical sufficiency in the deterministic case introduced by Fisher [24]. Using the Fisher-Neyman factorization theorem, it is easy to proof that the classical sufficiency implies Bayesian sufficiency. The MMSE estimator $\hat{\underline{\theta}}_{\text{MMSE}}$ can then be written as $\hat{\underline{\theta}}_{\text{MMSE}} = \int_\Theta \underline{\theta}\, g(\underline{t}(\underline{x}), \underline{\theta})p(\underline{\theta})d\underline{\theta}/\int_\Theta g(\underline{t}(\underline{x}), \underline{\theta})p(\underline{\theta})d\underline{\theta}$.

This significantly simplifies the adaptive MMSE estimation which is now split into two steps: First, we apply the RWLS algorithm to simplify the signal model and to compute the sufficient statistic $\underline{\tilde{x}}_n = \hat{\underline{\theta}}_{\text{WLS}}(n)$. Then, we calculate the MMSE estimator from the simplified signal model $\underline{\tilde{x}}_n = \underline{\theta}(n) + \underline{\tilde{z}}_n$ with a uniform a priori distribution of $\underline{\theta}(n)$ in Θ. We refer to this estimator as *recursive MMSE* (RMMSE) estimator.

Note that the MMSE estimation in the second step still involves two M-dimensional integrals which we in general can not compute analytically. In [25], we considered the case of ellipsoidal constraints on $\underline{\theta}$. The computational complexity was reduced to the calculation of M one-dimensional integrals. In this correspondence, we apply a more general statistical approach which is not restricted to ellipsoidal constraints only. We use the Monte Carlo method to approximate the integrals. Since $\underline{\tilde{x}}_n|\underline{\theta} \sim \mathcal{N}(\underline{\theta}, \tilde{\mathbf{C}}_n)$ after the signal model transform and $\underline{\theta}$ is uniformly distributed in Θ, we only need to generate random samples from a truncated multivariate Gaussian distribution, i.e. of a Gaussian distribution which is set to zero for $\underline{\theta} \notin \Theta$. This becomes clear if we consider the MMSE estimator for the simplified model

$$\hat{\underline{\theta}}_{\text{RMMSE}}(n) = \frac{\int_{\Theta} \underline{\theta} \, \exp\{-\frac{1}{2}(\underline{\theta} - \underline{\tilde{x}}_n)^T \tilde{\mathbf{C}}_n^{-1} (\underline{\theta} - \underline{\tilde{x}}_n)\} d\underline{\theta}}{\int_{\Theta} \exp\{-\frac{1}{2}(\underline{\theta} - \underline{\tilde{x}}_n)^T \tilde{\mathbf{C}}_n^{-1} (\underline{\theta} - \underline{\tilde{x}}_n)\} d\underline{\theta}}. \tag{27}$$

Using Monte Carlo integration [26], we have the approximation

$$\hat{\underline{\theta}}_{\text{RMMSE}}(n) \approx \frac{1}{I} \sum_{i=1}^{I} \underline{\theta}_i \tag{28}$$

where $\underline{\theta}_i$ is a random sample drawn from $\exp\{-\frac{1}{2}(\underline{\theta} - \underline{\tilde{x}}_n)^T \tilde{\mathbf{C}}_n^{-1} (\underline{\theta} - \underline{\tilde{x}}_n)\}$ satisfying $\underline{\theta}_i \in \Theta$. Therefore, $\underline{\theta}_i$ is a sample of a truncated Gaussian. To generate $\underline{\theta}_i$, we first draw a sample $\underline{\psi}$ from $\mathcal{N}(\underline{0}, \mathbf{I})$ and then transform it by $\underline{\varphi} = \mathbf{U}_n \underline{\psi} + \underline{\tilde{x}}_n$ with $\mathbf{U}_n \mathbf{U}_n^T = \tilde{\mathbf{C}}_n$. It is accepted as a new value for $\underline{\theta}_i$ if $\underline{\varphi} \in \Theta$. Otherwise, it is rejected.

By doing so, we can include any constraints on $\underline{\theta}$ easily into the generation of samples from $p(\underline{\theta})p(\underline{\tilde{x}}_n|\underline{\theta})$. This rejection sampling works well when the truncated Gaussian PDF has enough "probability mass" in Θ as otherwise we would do many rejections before a sample $\underline{\varphi}$ is accepted. This happens, for example, for equality constraints. Such constraints can not be handled by a rejection scheme, see [25] for a series expansion approach for this kind of constraints.

Note that there are also more sophisticated methods to draw multivariate samples from a truncated Gaussian, see e.g. [27]. They often yield a smaller rejection rate than the simple scheme presented here but have the drawback that they are only applicable if it is possible to easily compute the lower and upper truncation bounds of the PDF of the truncated Gaussian random variable.

D. Comparison of the Different Approaches

In this section, we compare all three approaches to solve our time-variant estimation problem (9). We have four options:

1) Simply ignore the additional information $\underline{\theta}(n) \in \Theta$ and use the ML estimator for (9). This yields the WLS estimator (3) and the RWLS algorithm (16) for an efficient calculation.
2) A second option is to consider the constrained ML estimator for (9) which is given in (18). The advantage of this method is that it will ensure $\hat{\underline{\theta}} \in \Theta$, i.e. the a priori information is taken into account as a hard constraint. However the drawback of this method is that there are no efficient

recursive algorithms for the case that $\hat{\underline{\theta}}_{\text{WLS}} \notin \Theta$. Then, we have the original problem (18) at each time step. An exception is the case of linear equality/inequality, quadratic inequality constraints as considered in [2], [3] or l_1-norm constraints as in [4].

3) Restrict the attention to the class of affine estimators only and find that affine estimator that has the smallest worst-case MSE for all $\underline{\theta}(n)$ in Θ. By doing this, the a priori information is taken into account as a soft constraint by using the minimax approach. Although this recursive AMX has in general an even higher computational complexity than CML, it is in our view still to be favored due to the following three reasons: First, the AMX estimator often shows a smaller MSE compared to RCML. Second, if $\mathbf{H}(n)$ and $\mathbf{C}(n)$ are constant over time, $\mathbf{M}(n)$ and $\underline{u}(n)$ reach their steady state value after some time steps and hence, no update of $\mathbf{M}(n)$ and $\underline{u}(n)$ is needed anymore. Third, we can restrict $\mathbf{M}(n)$ and $\underline{u}(n)$ to have a special structure, e.g. $\mathbf{M}(n)$ is only allowed to be diagonal. By this we can reduce the number of optimization variables and consequently the compuational complexity.
4) Another possibility is to recast the frequentist problem to a Bayesian one by interpreting the a priori information $\underline{\theta} \in \Theta$ as a priori PDF $p(\underline{\theta})$ and thus $\underline{\theta} \in \Theta$ is used as a soft constraint (see discussion below). In this correspondence, we choose $p(\underline{\theta})$ as uniform in Θ and zero otherwise as we do not have any other information about $\underline{\theta}$ except for $\underline{\theta} \in \Theta$. This is a natural choice as it does not prefer particular values of $\underline{\theta}$. As we use rejection sampling to incorporate the a priori information, the RMMSE estimator has the advantage compared to RCML and RAMX that it can handle quite complicated constraints.

In summary, all three approaches (RCML, RAMX and RMMSE with a uniform prior) can be viewed as different methods to model the a priori information $\underline{\theta} \in \Theta$ of the frequentist problem (9). Especially the RAMX and RMMSE estimator allow to efficiently incorporate the a priori information into the estimation problem.

Note that the RAMX and the RMMSE estimator will in general not ensure $\hat{\underline{\theta}} \in \Theta$. The a priori knowledge is only exploited to achieve a smaller mean squared error of $\hat{\underline{\theta}}$. Interestingly, the RMMSE estimator satisfies the constraint $\hat{\underline{\theta}} \in \Theta$ automatically if Θ is a convex set [18].

IV. Example

The following example compares the introduced estimators in terms of the mean squared error. We consider the estimation of $\underline{\theta}(n) \in \Theta \subset \mathbb{R}^2$, i.e. $M = 2$, under the ellipsoidal constraint $\Theta = \{\underline{\theta} : \underline{\theta}^T \mathbf{A} \underline{\theta} \leq 1\}$. Fig. 1 shows the trajectory of $\underline{\theta}(n)$ (solid line) consisting of $N = 45$ discrete-time points and the

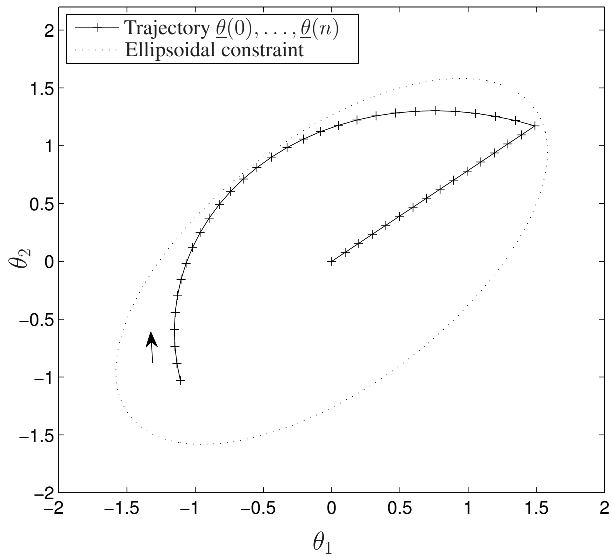

Fig. 1: True trajectory of $\underline{\theta}(n)$

boundary of Θ (dotted line). The signal model is given by $\underline{x}(n) = \underline{\theta}(n) + \underline{z}(n)$, with $\underline{z}(n) \sim \mathcal{N}(\underline{0}, 0.1\mathbf{I})$. This means $\mathbf{H}(n) = \mathbf{I}$.

We use the following five instantaneous estimators to estimate $\underline{\theta}(n)$ given only $\underline{x}(n)$:

- Least squares (LS): The LS estimator is given by $\hat{\underline{\theta}}(n) = \left(\mathbf{H}(n)^T\mathbf{H}(n)\right)^{-1}\mathbf{H}(n)^T\underline{x}(n) = \underline{x}(n)$. This means, LS accepts each observation $\underline{x}(n)$ as a new estimate and does not take the old observations $\underline{x}(n-1), \ldots, \underline{x}(0)$ into account.
- Constrained maximum likelihood (CML): The CML estimate can be found by solving a quadratically constrained quadratic program (QCQP). Using the method of Lagrange multipliers, a simple recursive rule can be derived which yields the CML solution.
- Affine minimax (AMX): Similar to LS, this estimator only uses the actual observation $\underline{x}(n)$. As the model matrix $\mathbf{H}(n)$ and the covariance matrix $\mathbf{C}(n)$ of the noise are time-invariant in our example, we have $\mathbf{M}(n) = \mathbf{M}$ and $\underline{u}(n) = \underline{u}$ for our affine estimator. Furthermore, as the center of the ellipsoid Θ is the origin, we have $\underline{u} = \underline{0}$. The optimization problem (20) for the special case of an ellipsoidal constraint $\underline{\theta}^T\mathbf{A}\underline{\theta} \leq 1$ was considered in [28] and a SDP formulation was given there.
- Projected AMX: This estimator is identical to the AMX but conditionally does a projection onto the boundary if $\hat{\underline{\theta}}_{\text{AMX}}(n) \notin \Theta$. The projection is done with an algorithm taken from [29].
- MMSE: The MMSE solution is calculated by means of Monte Carlo integration as described in Sec. III-C using $I = 200$ samples for approximating the integral.

Furthermore, we use the following five time-recursive estimators to estimate $\underline{\theta}(n)$ given $\underline{x}(0), \ldots, \underline{x}(n)$:

- Recursive weighted least squares (RWLS): We solve the weighted least squares problem (9) where old observations are weighted with a forgetting factor β. The RWLS does not care about the constraint $\underline{\theta}(n) \in \Theta$.
- Recursive constrained ML (RCML): Similar to the CML, we have to solve a QCQP where $\hat{\underline{\theta}}_{\text{WLS}}$ is now replaced by $\hat{\underline{\theta}}_{\text{WLS}}(n)$.
- Recursive affine minimax (RAMX): As we have seen in Sec. III-B, RAMX is given by a matrix multiplication of the RWLS solution with $\mathbf{I} + \mathbf{M}(n)$ plus an addition of $\underline{u}(n)$. Thus, we have to solve the minimax optimization problem (20) repeatedly at each time step from $\mathbf{R}_n$. Similar to the AMX estimator, we use the SDP formulation from [28].
- Projected RAMX: Similar to the projected AMX, this estimator does a conditional projection if the RAMX estimate is not in Θ.
- Recursive MMSE (RMMSE): The RMMSE estimator combines RWLS and MMSE as described in Sec. III-C. We use again Monte Carlo integration with $I = 200$ samples.

Fig. 2 shows the averaged mean squared error $\frac{1}{N}\sum_{n=1}^{N}\|\hat{\underline{\theta}}(n) - \underline{\theta}(n)\|^2$ averaged over $1\,500$ trials of this experiment for all estimators as a function of the forgetting factor β. Note that the five instantaneous estimators do not depend on β. The RMMSE estimator and the projected RAMX have the smallest mean squared error for all values of β among the recursive estimators. They also outperform the instantaneous estimators for $0 < \beta < 0.75$. The optimal forgetting factor for this time-variant problem is $\beta_{\text{opt}} \approx 0.5$. If the forgetting factor is smaller, the effective number $\frac{1}{1-\beta}$ of observations (degrees of freedom) is too small for a reliable estimation of $\underline{\theta}(n)$. If, however, the forgetting factor is too large, the estimator is not able to track the time-varying parameter $\underline{\theta}(n)$.

The choice of the number of samples I in the Monte Carlo integration also plays an important role in the design of the estimator. To analyze the influence of I, we varied its value. Fig. 3 shows the mean squared error of the RMMSE for different values of I and the noise variance σ^2. We see that it is sufficient to use $I \geq 100$ samples. To be on the safe side, we choose $I = 200$.

Finally, we compare the RAMX and its steady-state version. Instead of calculating $\mathbf{M}(n)$ at each time step, we calculate a fixed $\mathbf{M}$ from the steady-state correlation matrix $\mathbf{R}_\infty$ given in (24). Fig. 4 shows the ratio of the mean squared errors of both estimators as a function of the time index n, i.e. $\gamma = \text{MSE(RAMX)}/\text{MSE(steady-state RAMX)}$. Clearly, the MSE values of the steady-state approximation are larger than the corresponding values of the RAMX. The difference between both estimators vanishes for

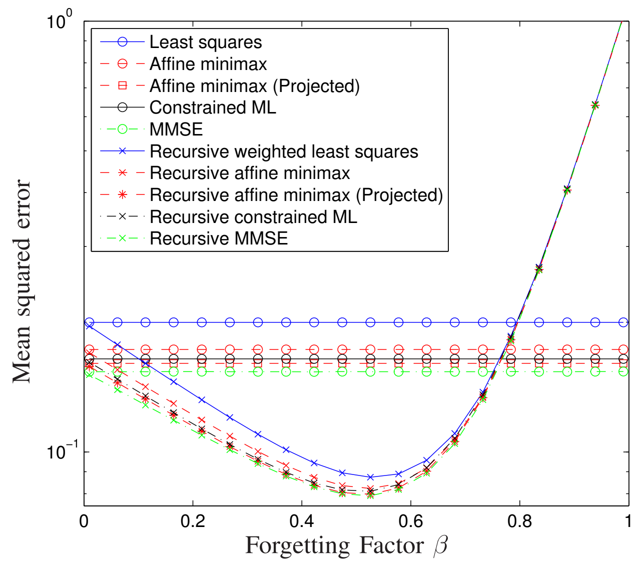

Fig. 2: Mean squared error vs. the forgetting factor β

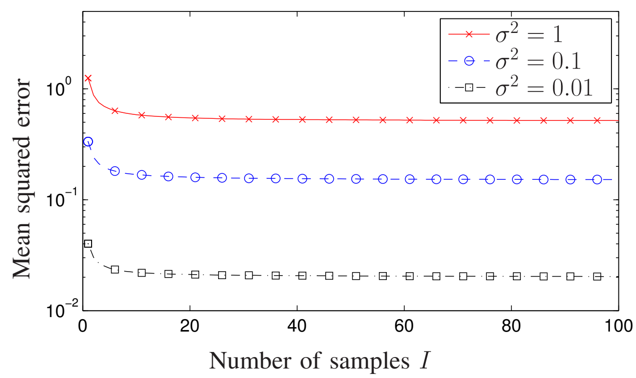

Fig. 3: Mean squared error for varying number of Monte Carlo samples

increasing time indices n.

V. Conclusions

We considered three recursive estimators – the recursive constrained ML, the recursive affine minimax and the recursive MMSE – for the adaptive estimation of a time-varying parameter vector $\underline{\theta}(n)$ in a linear Gaussian signal model with the a priori knowledge $\underline{\theta}(n) \in \Theta$. All three estimators use the RWLS algorithm in a preprocessing step to transform a dimensionally growing estimation problem to a fixed-dimension task. As the estimators take care of the history of measurements and the a priori knowledge $\underline{\theta}(n) \in \Theta$, they outperform the instantaneous estimators (least squares, constrained ML, affine minimax, MMSE) as well

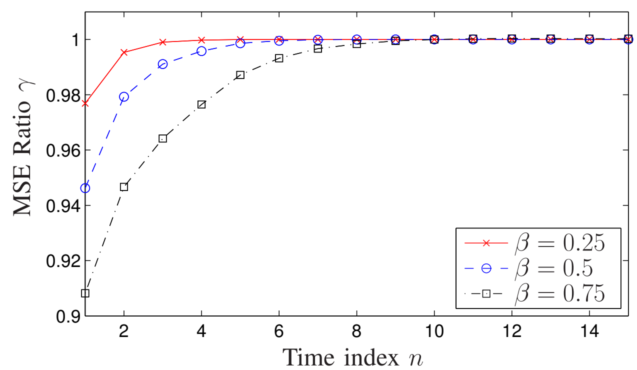

Fig. 4: Ratio of the mean squared errors for RAMX and its steady-state approximation

as the unconstrained recursive estimator RWLS. Especially the derived RMMSE estimator is interesting as the proposed rejection sampling allows a simple handling of general constraints. Furthermore, the computational complexity is moderate as we only have to generate samples from a truncated Gaussian density.

REFERENCES

[1] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal. Statist. Soc., Series B.*, vol. 58, no. 1, pp. 267–288, 1996.

[2] H. L. Van Trees, *Optimum Array Processing. Part IV of Detection, Estimation and Modulation Theory*, Wiley, 2002.

[3] Y. Zhu and X. R. Li, "Recursive least squares with linear constraints," *Communications in Information and Systems*, vol. 7, no. 3, pp. 287–312, 2007.

[4] B. Babadi, N. Kalouptsidis, and V. Tarokh, "Comparison of SPARLS and RLS algorithms for adaptive filtering," *Proceedings of the 2009 IEEE Sarnoff Symposium*, 2009.

[5] J. Pilz, "Minimax linear regression estimation with symmetric parameter restrictions," *Journal of Statistical Planning and Inference*, vol. 13, pp. 297–318, 1986.

[6] Y. C. Eldar, "Uniformly improving the Cramer-Rao Bound and Maximum-Likelihood estimation," *IEEE Transactions on Signal Processing*, vol. 54, no. 8, pp. 2943–2956, 2006.

[7] Y. C. Eldar, "MSE bounds with affine bias dominating the Cramer-Rao bound," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3824–3836, 2008.

[8] A. Elron, G. Leibovitz, Z. Ben-Haim, and Y. C. Eldar, "Recursive blind minimax estimation: Improving MSE over recursive least squares," *Proc. IEEEI*, 2008.

[9] A. I. Mcleod and B. Quenneville, "Mean likelihood estimators," *Statistics and Computing*, vol. 11, no. 1, pp. 57–65, 2001.

[10] G. M. Jenkins and D. G. Watts, *Spectral analysis and its applications*, Emerson-Adams Press, 1969.

[11] C. P. Robert, *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, Springer, 2001.

[12] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, 4th edition, 2002.

[13] J. A. Farrell and M. M. Polycarpou, *Adaptive Approximation Based Control*, Wiley, 2006.

[14] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 1: Estimation Theory*, Prentice-Hall, 1993.

[15] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation and Time Series Analysis*, Addison-Wesley, 1990.

[16] Y. Bard, *Nonlinear Parameter Estimation*, Academic Press, 1973.

[17] H. L. Van Trees and K. L. Bell, Eds., *Bayesian bounds for parameter estimation and nonlinear filtering/tracking*, Wiley-Interscience, 2007.

[18] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge Univ. Pr., 2007.

[19] A. Ben-Tal, L. E. Ghaoui, and A. Nemirovski, *Robust Optimization*, Princeton University Press, 2009.

[20] M. Evans and T. Swartz, "Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems," *Statistical Science*, vol. 10, no. 3, pp. 254–272, 1995.

[21] J. E. H. Shaw, A. Genz, J. Monahan, M. J. Schervish, L. Wasserman, and R. Wolfinger, "Comments on: "Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems" by M. Evans and T. Swartz," *Statistical Science*, vol. 11, no. 1, pp. 54–64, 1996.

[22] J. K. Ghosh, M. Delampady, and T. Samanta, *An Introduction to Bayesian Analysis: Theory and Methods*, Springer, 2006.

[23] A. N. Kolmogorov, "Definition of the center of scattering and the measure of accuracy by a limited number of observations," *Izvestia Akademii Nauk SSSR*, vol. 6, pp. 3–32, 1942.

[24] R. A. Fisher, "On the mathematical foundations of theoretical statistics," *Philosophical Transactions of the Royal Society of London. Series A*, vol. 222, pp. 309–368, 1922.

[25] S. Uhlich and B. Yang, "MMSE estimation in a linear signal model with ellipsoidal constraints," *Proc. IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 3249–3252, 2009.

[26] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, Springer, 2008.

[27] C. P. Robert, "Simulation of truncated normal variables," *Statistics and Computing*, vol. 5, no. 2, pp. 121–125, 1995.

[28] Y. C. Eldar, "Comparing between estimation approaches: Admissible and dominating linear estimators," *IEEE Transactions on Signal Processing*, vol. 54, no. 5, pp. 1689– 1702, 2006.

[29] Y. N. Kiseliov, "Algorithms of projection of a point onto an ellipsoid," *Lithuanian Mathematical Journal*, vol. 34, no. 2, pp. 141–159, 1994.