Constrained maximum likelihood estimation for state space sampled-data models

F. Ávila *, J.I. Yuz*, A. Donaire†, J.C. Agüero*
* Electronic Engineering Department, Universidad Técnica Federico Santa María, Chile
felipe.avilac@alumnos.usm.cl, {juan.yuz,juan.aguero}@usm.cl

† Queensland University of Technology, Australia.
alejandro.donaire@qut.edu.au

***Abstract*—The Expectation-Maximization algorithm is applied in this paper to estimate state-space sampled-data models including constraints on the location of the poles. Linear quadratic matrix inequalities are used as constraints to obtain a model that preserves properties of the continuous time system, such as stability or damping characteristics. The results of the algorithm are shown in a simulation study.**

***Index Terms*—System identification, maximum likelihood, Expectation-Maximization, linear matrix inequalities, constraints.**

I. Introduction

Maximum likelihood (ML) estimation has been extensively applied to estimate parameters in time and frequency domains [1] [2] [3] [4] [5] [6]. However, the associated optimization problem may be non-convex. The expectation-maximization (EM) algorithm is an iterative procedure to obtain the ML estimate for the vector of parameters that define the true system [7].

The EM algorithm introduces the notion of complete data set, which includes the observed data and the hidden or missing data [7]. The algorithm transforms the problem of maximizing the likelihood function (usually non convex) into maximizing an auxiliary function which is iteratively updated and then maximized. For the case of state-space models, the more obvious choice for the hidden data is the state vector sequence.

The EM algorithm converges to stationary points of the likelihood function (that can be saddle points, or as local or global maxima) [8]. Conditions for the convergence of the algorithm were established in [9]. Numerically robust implementations of EM algorithm for linear state-space discrete time models have been studied in [10], and for fast and non-uniformly sampled data models in [11] [12] [13].

The ML estimator is consistent, however when estimating parameters there is always a limited amount of data. As a consequence, the estimated model may not reflect characteristics of the real system, such as, for example, stability or (non) oscillatory behavior. In particular, when estimating continuous time models from sampled data, prior knowledge of these characteristics of the physical system have to be reflected in the estimated model. In this paper, we consider this requirement and we propose a maximum likelihood estimation of the parameters of state-space models subject to constraints on the system poles.

Identification of state-space systems including constraints, has been developed using subspace methods. For example in [14] subspace estimation methods are applied and the matrix A is obtained as the product of a shift matrix with a pseudo-inverse. In that work, it is shown that if the shift matrix is formed by entering a block of zeros in a suitable position, then the resulting model is stable, however, at the cost of distorting the estimated observability matrix. Alternative approaches increase the data set, in order to guarantee marginal stability [15]. In [16], a regularization term is added to the cost of least squares, to force constraints on the eigenvalues of the matrix A. An interesting solution provided by [17] applies a subspace estimation algorithm, using the Lyapunov's Inequalities, $P - APA^T > 0$ and $P > 0$, as constraints. These inequalities ensure that the estimated model is asymptotically stable. The approach does not distort the extended observability matrix as in [14], nor modifies the estimated state or input sequences as in [15]. Moreover, in [18] a generalization of [17] is presented using constraints in the form of linear quadratic matrix inequalities. In fact, the unit circle and other regions of interest can be expressed as linear matrix inequalities (LMI).

In this paper we apply constraints based on LMI regions in ML estimation using the EM algorithm. In particular, we use LMI to force constraints on the location of the model poles.

The structure of the paper is as follows: Section II presents the EM algorithm for discrete-time state space models. LMI constraints are introduced in the EM algorithm in Section III, and then Section IV presents simulation results and comparison to the use of the unconstrained EM algorithm. Finally, conclusions are presented in Section V.

II. EM algorithm for sampled data

A. Continuous-time system description

In this paper, we consider the EM algorithm for state-space sampled data models. Firstly, we assume a continuous-time system subject to stochastic disturbances defined, as in [11] by the following stochastic differential equation (SDE) model

[19]

$$dx(t) = A_c x(t)dt + B_c u(t)dt + dw(t) \quad (1)$$
$$dz(t) = C_c x(t)dt + D_c u(t)dt + dv(t) \quad (2)$$

where $u(t) \in \mathbb{R}$, $y(t) \in \mathbb{R}$, and $x(t) \in \mathbb{R}^n$ are the input, output and state, respectively; the system matrices are $A_c \in \mathbb{R}^{n\times n}$, $B_c \in \mathbb{R}^{n\times 1}$, $C_c \in \mathbb{R}^{1\times n}$ and $D_c \in \mathbb{R}$; and the incremental state disturbance $dw(t)$ and incremental measurement disturbance $dv(t)$ are assumed to be independent, zero mean and with Gaussian distribution, such that

$$\mathcal{E}\left\{\begin{bmatrix} dw(t) \\ dv(t)\end{bmatrix}\begin{bmatrix} dw(s) \\ dv(s)\end{bmatrix}^T\right\} = \begin{bmatrix} Q_c & 0 \\ 0 & R_c\end{bmatrix}\delta_k(t-s) \quad (3)$$

Moreover, the initial state is assumed independent of $dw(t)$ and $dv(t)$, and also with Gaussian distribution having mean $\bar{x}_0$ and covariance P_0.

Our main interest is to estimate the system parameters

$$\theta_c = \{A_c, B_c, C_c, D_c, Q_c, R_c\} \quad (4)$$

subject to the condition that the eigenvalues of the estimate of A_c yields in a area of interest in the complex plane. We assume prior knowledge of some characteristics of the system, e.g. the system is stable. We also assume that the sampling period is constant $T_s = \Delta$, and the sampling process is fast.

B. Sampled data-model

The problem of dealing with samples of continuous-time signals has been discussed in the literature, for example in [11] [20] [21] [22] [23]. In this paper, the system input $u(t)$ is assumed to be generated from an input sequence u_k using the usual zero order hold (ZOH), i.e.

$$u(t) = u_k \ , \ k\Delta < t < (k+1)\Delta \quad (5)$$

where Δ is the constant sampling period. Moreover, we assume an integrate and reset filter (IRF) before instantaneous sampling of the output as described in [11] [22] [24]. Under these assumptions, an incremental discrete model can be obtained, such that its output has the same second-order statistics the sampled continuous-time output of the original system [23]. That model is given by:

$$dx_k^+ = A_\delta x_k \Delta + B_\delta u_k \Delta + dw_k^+ \quad (6)$$
$$\bar{y}_{k+1}\Delta = dz_k^+ = C_\delta x_k \Delta_k + D_\delta u_k \Delta + dv_k^+ \quad (7)$$

where the increments are defined as

$$df_k^+ = f_{k+1} - f_k \quad (8)$$

The matrices are given by

$$A_\delta = \frac{e^{A_c\Delta} - I}{\Delta} \ , \ B_\delta = \left[\frac{1}{\Delta}\int_0^\Delta e^{A_c\eta}d\eta\right]B_c \quad (9)$$

$$C_\delta = C_c\left[\frac{1}{\Delta}\int_0^\Delta e^{A_c\eta}d\eta\right] \quad (10)$$

$$D_\delta = D_c + C_c\left[\frac{1}{\Delta}\int_0^\Delta \int_0^\xi e^{A_c\eta}d\eta d\xi\right]B_c \quad (11)$$

and the covariance structure of the noise vector is given by

$$\mathcal{E}\left\{\begin{bmatrix} dw_l^+ \\ dv_l^+\end{bmatrix}^T\begin{bmatrix} dw_k^+ \\ dv_k^+\end{bmatrix}\right\} = \begin{bmatrix} Q_\delta & S_\delta \\ (S_\delta)^T & R_\delta\end{bmatrix}\Delta\delta_k(l-k) \quad (12)$$

with

$$\begin{bmatrix} Q_\delta & S_\delta \\ (S_\delta)^T & R_\delta\end{bmatrix} = \frac{1}{\Delta}\int_0^\Delta \begin{bmatrix} e^{A_c\eta} & 0 \\ C_c\int_0^\eta e^{A_c\xi}d\xi & I\end{bmatrix} \times \begin{bmatrix} Q_c & 0 \\ 0 & R_c\end{bmatrix}\begin{bmatrix} e^{A_c\eta} & 0 \\ C_c\int_0^\eta e^{A_c\xi}d\xi & I\end{bmatrix}^T d\eta \quad (13)$$

The derivation of the expressions above can be found in [22]. Moreover, an implementation to obtain the integrals of matrix exponentials in a numerically stable way is presented in [25]. It can be noticed that when the sample period Δ goes to zero, the model (6)-(13) converges to the SDE model (1)-(2).

Alternatively, the model (6)-(13) can be expressed using the forward-shift operator q, i.e.

$$qx_k = x_{k+1} = A_q x_k + B_q u_k + \tilde{w}_k \quad (14)$$
$$\bar{y}_k = C_q x_k + D_q u_k + \bar{v}_k \quad (15)$$

where

$$\mathcal{E}\left\{\begin{bmatrix} \tilde{w}_l \\ \tilde{v}_l\end{bmatrix}^T\begin{bmatrix} \tilde{w}_k \\ \tilde{v}_k\end{bmatrix}\right\} = \begin{bmatrix} Q_q & S_q \\ (S_q)^T & R_q\end{bmatrix}\delta_k(l-k) \quad (16)$$

The transformation between incremental and shift form model matrices is given by

$$A_q = I + \Delta A_\delta \ , \ B_q = \Delta B_\delta \ , \ C_q = C_\delta \ , \ D_q = D_\delta \quad (17)$$
$$Q_q = \Delta Q_\delta \ , \ S_q = S_\delta \ , \ R_q = \frac{1}{\Delta}R_\delta \quad (18)$$

Remark 1: The sampled-data model written in incremental form may be preferred for fast sampling applications and for parameter estimation, in particular, due to its convergence properties as $\Delta \to 0$ [21] [22] [23].

We are interested in obtaining an estimate of the the system parameters, i.e.

$$\theta = \{A_\delta, B_\delta, C_\delta, D_\delta, Q_\delta, R_\delta\} \quad (19)$$

such that the eigenvalues of the estimate of matrix A_δ lies in a specific region of the complex plane.

Remark 2: Notice that constraints used when estimating A_q or A_δ arise from prior knowledge of A_c.

- If A_c is stable, then A_q and A_δ are stable.
- If A_c has complex conjugate eigenvalues (resonant poles), then A_q and A_δ have resonant poles.

C. Maximum Likelihood estimation and the EM algorithm

The likelihood function is given by the probability of the data Y given the parameter vector θ. If the measurements are Gaussian distributed, usually we consider the natural logarithm of the likelihood function, i.e.

$$L(\theta) = \log p(Y|\theta) \quad (20)$$

The maximum (log) likelihood estimate of θ is then given by

$$\hat{\theta}_{ML} = \underset{\theta\in\Theta}{\text{argmax}}\ L(\theta) \tag{21}$$

The maximum likelihood of θ is efficient and asymptotically consistent. However, the associated optimization problem is, in general, non-convex.

Is well known that (20) can be rewritten as (see e.g. [2])

$$\begin{aligned} L(\theta) = &-\frac{1}{2}\sum_{t=1}^{N}\log\ \det\ (C_q P_{t|t-1} C_q^T + R_q) \\ &-\frac{1}{2}\sum_{t=1}^{N}\epsilon_t^T(\theta)[C_q P_{t|t-1} C_q^T + R_q]^{-1}\epsilon_t(\theta) \end{aligned} \tag{22}$$

where θ are the system parameters and (assuming $y_0 = 0$)

$$\epsilon_t(\theta) \triangleq y_t - \hat{y}_{t|t-1}(\theta)\ ,\ \hat{y}_{t|t-1} \triangleq \mathcal{E}\left\{y_t|y_0 : y_{t-1}, \theta\right\} \tag{23}$$

with N the amount of data available and $\hat{y}_{t|t-1}$ the mean square optimal one-step ahead prediction of the system output and the notation $y_0 : y_k = [y_0, y_1, ..., y_k]$. Also

$$P_{t|t-1} \triangleq \mathcal{E}\left\{(x_t - \hat{x}_{t|t-1})(x_t - \hat{x}_{t|t-1})^T|\theta\right\} \tag{24}$$

is the covariance matrix associated to the estimate $\hat{x}_{t|t-1} \triangleq \mathcal{E}\left\{x_t|y_0 : y_{t-1}, \theta\right\}$. Both of these may be computed using Kalman filtering algorithms.

The (log)likelihood function can be rewritten as

$$\log\ p(Y|\theta) = \mathcal{Q}(\theta, \hat{\theta}_i) - \mathcal{H}(\theta, \hat{\theta}_i) \tag{25}$$

where $\hat{\theta}_i$ is an estimate of θ available at the i-th iteration, and

$$\mathcal{Q}(\theta, \hat{\theta}_i) = \mathcal{E}\left\{\ \log\ P(X, Y|\theta)|Y, \hat{\theta}_i\right\} \tag{26}$$

$$\mathcal{H}(\theta, \hat{\theta}_i) = \mathcal{E}\left\{\ \log\ P(X|Y, \theta)|Y, \hat{\theta}_i\right\} \tag{27}$$

where Y is the available data, and X is the hidden data. Using Jensen's inequality it can shown that $\mathcal{H}(\theta, \hat{\theta}_i) \leq \mathcal{H}(\hat{\theta}_i, \hat{\theta}_i)$ [7]. As a consequence, if we maximize (or increase) the value of $\mathcal{Q}(\theta, \hat{\theta}_i)$ as a function of θ, then we increase the log-likelihood function, leading to an iterative algorithm. The EM algorithm is then given by the following steps:

- *E-step*: given $\hat{\theta}_i$, we obtain $\mathcal{Q}(\theta, \hat{\theta}_i)$
- *M-step*: Find θ that maximize $\mathcal{Q}(\theta, \hat{\theta}_i)$, getting a new parameter estimate $\hat{\theta}_{i+1}$
- Return to the $E - step$, increasing the index $i \to i + 1$, and iterate until a convergence criteria is satisfied.

The natural choice of the hidden variables X for state space models is the state sequence. Thus, in the E-step, given the state-space discrete-time linear model, the function $\mathcal{Q}(\theta, \hat{\theta})$ can be expressed as

$$\begin{aligned} -2\mathcal{Q}(\theta, \hat{\theta}_i) =&\log\ \det(P_0) + \\ &\text{Tr}\{P_0^{-1}\mathcal{E}\left\{(x_0 - \mu)(x_0 - \mu)^T|Y_N\right\}\} \\ &N\ \log\ \det\Pi + N\text{Tr}\{\Pi^{-1}[\Phi \\ &- \Psi\Gamma^T - \Gamma\Psi^T + \Gamma\Sigma\Gamma^T]\} \end{aligned} \tag{28}$$

where the augmented matrices can be expressed in terms of the incremental or the shift operator models matrices

$$\Pi \triangleq \begin{bmatrix}\Delta Q_\delta & 0 \\ 0 & \frac{1}{\Delta}R_\delta\end{bmatrix}\begin{bmatrix}Q_q & 0 \\ 0 & R_q\end{bmatrix} \tag{29}$$

$$\Gamma \triangleq \begin{bmatrix}I + \Delta A_\delta & \Delta B_\delta \\ C_\delta & D_\delta\end{bmatrix} = \begin{bmatrix}Aq & Bq \\ Cq & Dq\end{bmatrix} \tag{30}$$

$$z_t^T \triangleq \begin{bmatrix}x_t^T & u_t^T\end{bmatrix}\ ,\ \xi_t^T \triangleq \begin{bmatrix}x_{t+1}^T & y_t^T\end{bmatrix} \tag{31}$$

$$\Phi \triangleq \frac{1}{N}\sum_{t=1}^{N}\mathcal{E}\left\{\xi_t\xi_t^T|Y_N, \hat{\theta}_i\right\} \tag{32}$$

$$\Psi \triangleq \frac{1}{N}\sum_{t=1}^{N}\mathcal{E}\left\{\xi_t z_t^T|Y_N, \hat{\theta}_i\right\} \tag{33}$$

$$\Sigma \triangleq \frac{1}{N}\sum_{t=1}^{N}\mathcal{E}\left\{z_t z_t^T|Y_N, \hat{\theta}_i\right\} \tag{34}$$

where $Y_N = [y_1 : y_N]$. Notice that, based on the convergence properties as $\Delta \to 0$, we have neglected $S_q = S_\delta = 0$ (13). The details of the proof can be found in [10] [11]. The estimated hidden data is the sequence state, that for linear system with Gaussian noise can be readily obtained using Kalman smoothing, both in shift-operator or incremental form (see e.g. [10] [11]).

In the M-step, we need to maximize the auxilary function $\mathcal{Q}(\theta, \hat{\theta}_i)$ in (28). As discussed in Remark 2, if we aim at preserving properties of the matrix A_c, we need to include constraints in the estimation of A_q or A_δ. Analytic expressions for $\hat{\theta}_{i+1}$ that maximizes $\mathcal{Q}(\theta, \hat{\theta}_i)$ are given in [10] [11]. However, maximizing the likelihood function may lead to A_q or A_δ estimates that do not preserve stability or damping properties arising from A_c for short data sets or low signal-to-noise ratios. To overcome this problem, in the next section we propose to introduce constraints in the M-step based in LMI regions in order to maximize the likelihood function subject to the constraints of interest in this paper..

III. LMI CONSTRAINTS

Linear matrix inequalities have been used in the system identification literature, for example, for pole placement and subspace parameters estimation with eigenvalues constraints (see, e.g, [26] [27] [18] [28]). In this section, we present a way to include constraints based on LMI regions in the M-step. These constraints represent the location of the eigenvalues of the estimated model.

A general LMI region is defined as

$$\mathcal{D} = \{z \in \mathbb{C} : f_\mathcal{D}(z) < 0\} \tag{35}$$

where

$$f_\mathcal{D}(z) = L + Mz + M^T\bar{z} \tag{36}$$

with $L, M \in \mathbb{R}^{n\times n}$ and $L = L^T$ and $\bar{z}$ denotes the conjugate of z. Therefore, the eigenvalues of $A \in \mathbb{R}^{m\times m}$ are within the LMI region $\mathcal{D}$ if and only if there exists a real matrix $P \in \mathbb{R}^{m\times m}$ such that

$$P = P^T > 0 \tag{37}$$

$$L \otimes P + M \otimes AP + M^T \otimes PA^T < 0 \tag{38}$$

where $\otimes$ denotes the kronecker product. If a matrix A (and P) satisfies (37)-(38), then we call A a $\mathcal{D}$-*stable* matrix. In [26], it is shown that any LMI region is convex and symmetric respect to the real axis, and that the intersection of LMI regions is also an LMI region. Therefore, in this paper we define LMI regions of interest, such as stability region, i.e. the unit circle in the z-complex plane, or (a region containing) the real axis within the stability region.

A region of interest is the stability region for the eigenvalues of $\hat{A}_\delta$. This region is a circle with center at $(-\frac{1}{\Delta}, 0)$ and radius $\frac{1}{\Delta}$ [21]. The LMI region for a circle with center $(q, 0)$ and radius r defined for (35)-(36) is

$$L = \begin{bmatrix} -r & -q \\ -q & -r \end{bmatrix} \ , \ M = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \tag{39}$$

We can also define a semi-plane of the complex plane such that Re$\{z\} > \alpha$, which is described in the form of (35)-(36) by

$$L = -2\alpha \ , \ M = -1 \tag{40}$$

Another region of interest is $|\text{Im}\{z\}| < \epsilon$, with $\epsilon > 0$, which is characterized by

$$L = \epsilon I_2 \ , \ M = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \tag{41}$$

with I_2 the 2×2 identity matrix. Notice that if we choose an arbitrarily small ϵ, we can get a region really arbitrarily close to the real axis.

Notice that the left-hand-side of the inequalities (37)-(38) are symmetric matrices and thus, for the constraints, we can use the Sylvester's criterion to verify if these matrices are positive definite [29]. Sylvester's criterion states that a symmetric matrix $M \in \mathbb{R}^{n\times n}$ is positive definite if and only if all the $\ell \times \ell$ upper left sub matrices of M have a positive determinant (i.e. the principal minors).

Sylvester's criterion can be used to reduce the computational complexity of the corresponding optimization problem in the M-step of the estimation algorithm. Indeed, one could compute the principal minors instead of computing their eigenvalues.

Hence, the M-step is defined as follows

$$\hat{\theta}_{i+1} \triangleq \underset{\theta}{\text{argmin}} - \mathcal{Q}(\theta, \hat{\theta}_i) \tag{42}$$

subject to

$$-\left(L \otimes P + M \otimes A_\delta P + M^T \otimes P(A_\delta)^T\right) > 0 \tag{43}$$

$$P = P^T \tag{44}$$

$$P > 0 \tag{45}$$

If $\hat{A}_\delta, P \in \mathbb{R}^{n\times n}$ and $L, M \in \mathbb{R}^{k\times k}$, then the left hand side of the inequality (43) is a symmetric matrix $\mathbb{R}^{nk\times nk}$, therefore we can apply Sylvester's criterion to the nk principal minor. These determinants are functions of the parameter estimation, hence the M-step is defined as

$$\hat{\theta}_{i+1} \triangleq \underset{\theta}{\text{argmin}} - \mathcal{Q}(\theta, \hat{\theta}_i) \tag{46}$$

subject to

$$c_\ell(\theta) < 0 \ , \ \textit{with} \ \ell = 1, ..., nk \tag{47}$$

$$P = P^T \tag{48}$$

$$b_k(\theta) < 0 \ , \ \textit{with} \ k = 1, ..., n \tag{49}$$

where c_ℓ is the ℓ-th principal minor of the left-hand-side of inequality (43), and b_k is the k-th principal minor of P.

A. Log-barrier function

The non-linear optimization problem defined in the M-step of the estimation algorithm with LMI constraints could be solved, for example, using $fmincon(\cdot)$ in MatLab. However, this solver does not guarantee that the constraints are (strictly) satisfied at each iteration. According to [30] only *"bound"* constraints are strictly satisfied by such solver, where *"bound"* constraints are of the type $lb \leq x \leq ub$, with x the vector parameter of optimization and lb and ub are constant vectors. As a consequence, if any non-linear constraint is not satisfied in one step of the EM algorithm, then our estimation problem of $\hat{A}_\delta$ will not satisfy the LMI constraints that represent the desired location of the eigenvalues. In order to overcome this issue we introduce *log-barrier* functions in the non-linear optimization problem.

In particular, for our problem (46) - (49) we define the *log-barrier* function as

$$\phi(\theta) \triangleq -\sum_{i=1}^{nk} \log\left(-c_\ell(\theta)\right) - \sum_{i=1}^{n} \log\left(-b_k(\theta)\right) \tag{50}$$

Then, the optimization problem results

$$\theta_{i+1} \triangleq \underset{\theta}{\text{argmin}} - t \cdot \mathcal{Q}(\theta, \hat{\theta}_i) + \phi(\theta) \tag{51}$$

subject to

$$P = P^T \tag{52}$$

where t is a parameter chosen by the user, typically a large number.

IV. Simulations

In this section, we show the results of including constraints in the M-step of the EM algorithm. Let us consider the parameter estimation problem for a continuous time system given by:

$$G(s) = \frac{1.15(s + 5.609)}{(s + 6)(s + 3)} \tag{53}$$

The sample time is chosen as $T_s = 0.008[s]$. The state-space sampled-data model matrices described in shift operator model (14) - (16) are

$$A_q = \begin{bmatrix} 0.9763 & 0 \\ 0 & 0.9531 \end{bmatrix} \ , \ B_q = \begin{bmatrix} 0.0079 \\ 0.0008 \end{bmatrix} \tag{54}$$

$$C_q = \begin{bmatrix} 0.9881 & 1.4646 \end{bmatrix} \ , \ D_q = 0.0046 \tag{55}$$

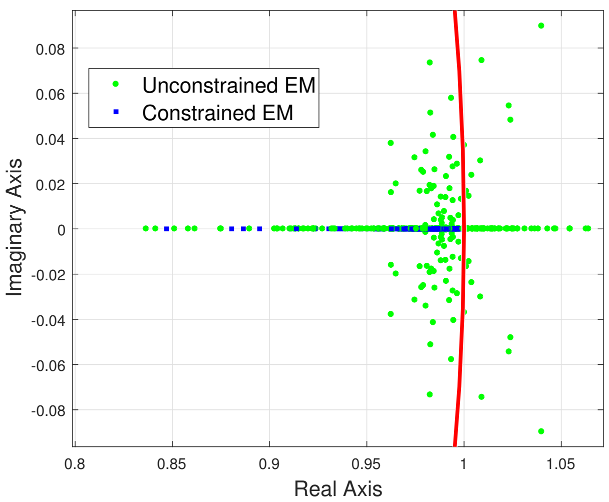

Fig. 1. Eigenvalues of estimated matrix A_q in shift operator model for 200 realizations. The red line corresponds to the unit circle (stability boundary).

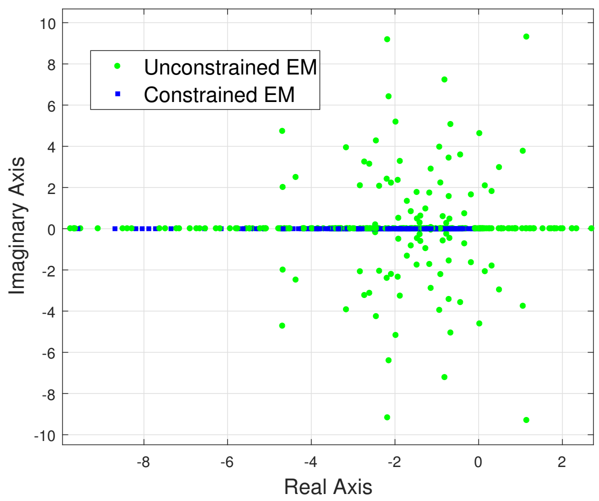

Fig. 2. Eigenvalues of estimated matrix A_δ in incremental model for 200 realizations.

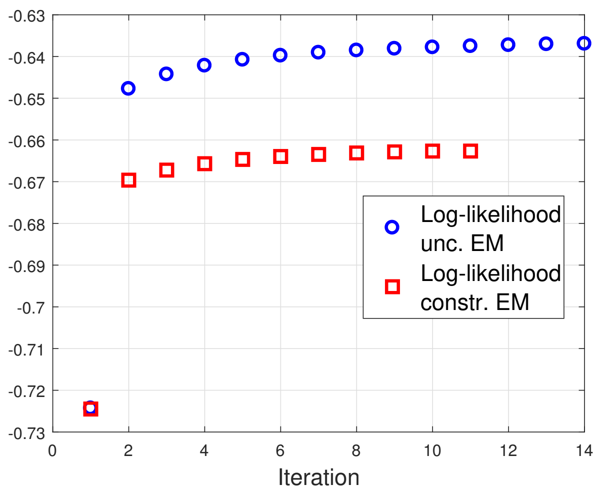

Fig. 3. Log-Likelihood evolution for estimated parameters at each iteration.

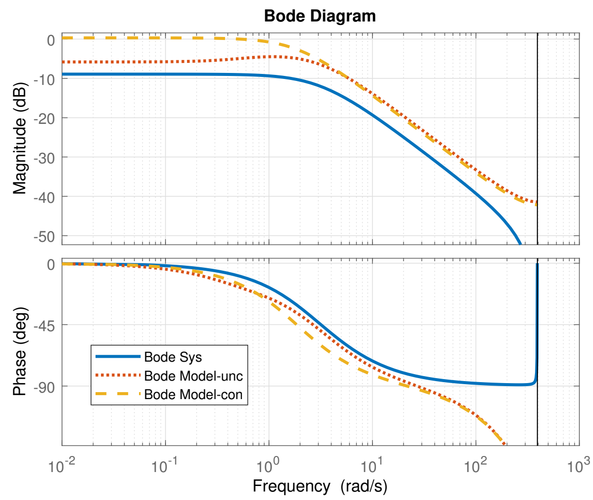

Fig. 4. Bode system and model.

$$Q_q = \begin{bmatrix} 0.0039 & 0 \\ 0 & 0.0038 \end{bmatrix}, \; R_q = 1.5242 \qquad (56)$$

Using (17) - (18) we can have the sampled-data model in incremental form defined in (6) - (13). We consider a short data set of length $N = 300$. The requirements are that the estimated sampled-data model has to be stable and should have no oscillatory behavior. These requirements are met if the eigenvalues lie in the segment $(0, 1)$ of the real line for $\hat{A}_q$ or, equivalently, in the segment $\left(\frac{-1}{\Delta}, 0\right)$ of the real line for $\hat{A}_\delta$ (see the transformations in (17)).

Figure 1 shows the location of the eigenvalues of the estimated matrix $\hat{A}_q$ using both the unconstrained EM (green) and constrained EM (blue) for a Monte-Carlo study with 200 realizations. As can be seen, the eigenvalues of the estimated matrix considering the unconstrained EM are sometimes unstable or complex conjugate. However, including constraints into the EM algorithm, the estimated model is stable and non oscillatory (i.e. real poles). Figure 2 shows the the location of eigenvalues fore the matrix $\hat{A}_\delta$ in the same Monte-Carlo study.

Figure 3 shows the evolution of the log-likelihood function for one particular realization using the unconstrained EM algorithm and also when LMI constraints are introduced. For the computation of the log-likelihood, we use (22) with a scale factor of $\frac{1}{N}$. As one would expect, the introduction of constraints in the parameter space implies that the likelihood function is smaller than in the unconstrained EM case. Nevertheless, for this particular realization, the unconstrained ML estimate leads to an oscillatory model and the real (continuous-time) system is not oscillatory and stable.

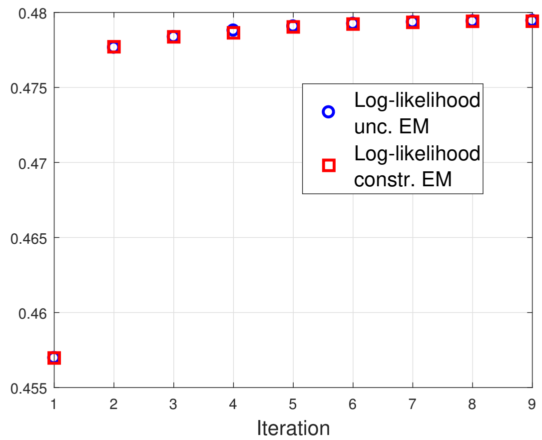

Fig. 5. Log-Likelihood evolution for estimated parameters at each iteration.

Figure 4 shows the evolution of the likelihood using a large data set ($N = 30.000$). We can see almost no difference between both log-likelihood function.

Finally, Figure 5 shows the Bode diagram of the real system and the estimated models obtained via unconstrained EM and constrained EM using LMI for one specific realization. One would think that unconstrained EM provides a more accurate model (both in magnitude and phase), however, it corresponds to an unstable model. In contrast, the discrete-time estimated model obtained using constrained EM is stable and non oscillatory.

V. Conclusions

The EM algorithm has been used in this paper to estimate sampled data models with constraints in the location of the system eigenvalues. LMI constraints have been used for the eigenvalues of the estimated discrete-time system matrix ($\hat{A}_q$ or $\hat{A}_\delta$), to guarantee that they lie on a specific region of interest. The results show that by introducing these constraints, the model obtained have the same characteristics than the original system. One would expect that introducing additional knowledge of the original system would improve the likelihood function, however, in our approach the knowledge of the system is implemented as constraint in the parameter space and, thus, the likelihood function is smaller that the unconstrained case. The approach presented here can be useful, in particular, for short data sets or low signal to noise ratio.

Acknowledgments

This work has been supported by CONICYT-Chile through grants CONICYT-PCHA Magíster Nacional 2016-22161323, Basal Project FB0008, and FONDECYT 1181090.

References

[1] K. J. Åström and B. Torsten, “Numerical identification of linear dynamic systems from normal operating records,” *IFAC Proceedings Volumes*, vol. 2, no. 2, pp. 96–111, 1965.
[2] L. Ljung, *System identification*. Wiley Online Library, 1999.
[3] T. Söderström and P. Stoica, “System identification,” 1989.
[4] J. C. Agüero, J. I. Yuz, G. C. Goodwin, and R. A. Delgado, “On the equivalence of time and frequency domain maximum likelihood estimation,” *Automatica*, vol. 46, no. 2, pp. 260–270, 2010.
[5] G. C. Goodwin and R. L. Payne, “Dynamic system identification: experiment design and data analysis,” 1977.
[6] R. Pintelon and J. Schoukens, “Box–Jenkins identification revisited – part i: theory,” *Automatica*, vol. 42, no. 1, pp. 63–75, 2006.
[7] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38, 1977.
[8] G. J. McLachlan and T. Krishnan, “The EM algorithm and extensions,” 2008.
[9] C. J. Wu, “On the convergence properties of the EM algorithm,” *The Annals of statistics*, pp. 95–103, 1983.
[10] S. Gibson and B. Ninness, “Robust maximum-likelihood estimation of multivariable dynamic systems,” *Automatica*, vol. 41, no. 10, pp. 1667–1682, 2005.
[11] J. Yuz, J. Alfaro, J. Agüero, and G. Goodwin, “Identification of continuous-time state-space models from non-uniform fast-sampled data,” *IET Control Theory & Applications*, vol. 5, no. 7, pp. 842–855, 2011.
[12] R. P. Aguilera, B. I. Godoy, J. C. Agüero, G. C. Goodwin, and J. I. Yuz, “An EM-based identification algorithm for a class of hybrid systems with application to power electronics,” *International Journal of Control*, vol. 87, no. 7, pp. 1339–1351, 2014.
[13] F. Chen, J. C. Agüero, M. Gilson, H. Garnier, and T. Liu, “EM-based identification of continuous-time arma models from irregularly sampled data,” *Automatica*, vol. 77, pp. 293–301, 2017.
[14] J. M. Maciejowski, “Guaranteed stability with subspace methods,” *Systems & Control Letters*, vol. 26, no. 2, pp. 153–156, 1995.
[15] N. L. C. Chui and J. M. Maciejowski, “Realization of stable models with subspace methods,” *Automatica*, vol. 32, no. 11, pp. 1587–1595, 1996.
[16] T. Van Gestel, J. A. Suykens, P. Van Dooren, and B. De Moor, “Identification of stable models in subspace identification by using regularization,” *IEEE Transactions on Automatic control*, vol. 46, no. 9, pp. 1416–1420, 2001.
[17] S. L. Lacy and D. S. Bernstein, “Subspace identification with guaranteed stability using constrained optimization,” *IEEE Transactions on automatic control*, vol. 48, no. 7, pp. 1259–1263, 2003.
[18] D. N. Miller and R. A. De Callafon, “Subspace identification with eigenvalue constraints,” *Automatica*, vol. 49, no. 8, pp. 2468–2473, 2013.
[19] B. Øksendal, “Stochastic differential equations,” in *Stochastic differential equations*, pp. 65–84, Springer, 2003.
[20] G. C. Goodwin, J. C. Aguero, M. E. Cea-Garridos, M. E. Salgado, and J. I. Yuz, “Sampling and sampled-data models: The interface between the continuous world and digital algorithms,” *IEEE Control Systems*, vol. 33, no. 5, pp. 34–53, 2013.
[21] R. H. Middleton and G. C. Goodwin, *Digital Control and Estimation: A Unified Approach (Prentice Hall Information and System Sciences Series)*. Prentice Hall Englewood Cliffs, NJ, 1990.
[22] A. Feuer and G. Goodwin, *Sampling in digital signal processing and control*. Springer Science & Business Media, 2012.
[23] J. I. Yuz and G. C. Goodwin, *Sampled-data models for linear and nonlinear systems*. Springer, 2016.
[24] K. J. Åström, *Introduction to stochastic control theory*. Courier Corporation, 2012.
[25] L. Ljung and A. Wills, “Issues in sampling and estimating continuous-time models with stochastic disturbances,” *Automatica*, vol. 46, no. 5, pp. 925–931, 2010.
[26] M. Chilali and P. Gahinet, “H/sub/spl infin//design with pole placement constraints: an lmi approach,” *IEEE Transactions on automatic control*, vol. 41, no. 3, pp. 358–367, 1996.
[27] M. Chilali, P. Gahinet, and P. Apkarian, “Robust pole placement in lmi regions,” *IEEE transactions on Automatic Control*, vol. 44, no. 12, pp. 2257–2270, 1999.
[28] F. Demourant and C. Poussot-Vassal, “A new frequency-domain subspace algorithm with restricted poles location through lmi regions and its application to a wind tunnel test,” *International Journal of Control*, vol. 90, no. 4, pp. 779–799, 2017.
[29] G. T. Gilbert, “Positive definite matrices and sylvester's criterion,” *The American Mathematical Monthly*, vol. 98, no. 1, pp. 44–46, 1991.
[30] MathWorks contributors, “Iterations can violate constraints - mathworks.” [Online; accessed 1-June-2018].