IDENTIFICATION OF LINEAR SYSTEMS IN CANONICAL FORM THROUGH AN EM FRAMEWORK.

Pavlos Papadopoulos, Vassilios Digalakis

Technical University of Crete
Department of Electronic & Computer Engineering
pavlos@telecom.tuc.gr, vas@telecom.tuc.gr

ABSTRACT
Least-squares estimation has always been the main approach when applying prediction error methods (PEM) in the identification of linear dynamical systems. Regardless of the estimation algorithm, if there are no restrictions on the form of the matrices we want to estimate, the matrices can be determined up to within a linear transformation and thus the result may be different than the true solution and the convergence of iterative algorithms may be affected. In this paper, we apply a new identification procedure based on the Expectation Maximization framework to a family of identifiable state-space models. To our knowledge, this is the first complete solution of Maximum-Likelihood estimation for general linear state-space models.

Index Terms— System Identification, MIMO State Space Models, Expectation Maximization, Prediction Error Methods

1. INTRODUCTION
System Identification (SI) has always been a very active and broad research topic. Identification procedures have been developed for various parametric models (transfer functions, state-space models, time-series models, etc) but also for non-parametric (impulse responses, frequency responses, spectral density functions, and so on). In this paper we will focus on the identification of discrete linear state-space models in the time domain, described by the following set of equations:

\[ x_{k+1} =Fx_k + w_k \]
\[ y_k =Hx_k + v_k \]

where \( x_k \in \mathbb{R}^n \) and \( y_k \in \mathbb{R}^m \) and \( w_k, v_k \) are zero-mean white additive Gaussian noise processes with

\[ E[w_kw_l^T] = Q\delta_{kl} \]
\[ E[v_kv_l^T] = R\delta_{kl} \]
\[ E[w_kv_l^T] = 0 \text{ for every } k, l \]

It is well known that if there are no constrains on the form of the matrices of (1), the identification will result into a linear transformation of the true solution([1][2]). In the case of iterative algorithms, convergence may be affected by iterating in a neighborhood of the true solution. In order to be able to determine uniquely the system parameters, the matrices must be structured in specific forms, called canonical forms, the exact parametrization of which has been under extensive study for some time(i.e. [3][4]). In most cases, the identification procedures do not apply directly to the initial model (1) but rather in its forward innovation representation([5][6]):

\[ x_{k+1} =Fx_k + Ke_k \]
\[ y_k =Hx_k + e_k \]

where \( K \) is the steady state Kalman Gain (SSKG).

There are two main streams of research in the area of SI, the first being subspace state-space system identification (SSID) which originated in the work of Ho and Kalman in [7] and is based on the realization theory. Our work follows the second approach, Prediction Error Methods (PEM), which has its roots in the work of Astrom and Bohlin in [8]. Classical PEM methods minimize some cost function using iterative procedures. Least-Squares Estimation and its variants dominate this family of algorithms.

In our work we follow the Maximum Likelihood (ML) approach and apply the Expectation-Maximization (EM) Algorithm ([9]) to a class of identifiable models presented in [10]. The rest of the paper is organized as follows. In Section 2 we present the form of the state-space models we have worked on and describe our algorithm. In Section 3 we show how we have applied the EM algorithm for the estimation, while in Section 4 we present our experimental results. Finally, in Section 5 we draw our conclusions.

2. CANONICAL FORMS FOR LIKELIHOOD IDENTIFIABILITY

There are multiple definitions of identifiability in the literature, depending on the criterion used for estimation. In this
work, since we use the ML criterion for estimation, we follow the definition in [10]. Two different parameter sets are maximum-likelihood unresolvable, if the conditional likelihood of a new observation given the previous observations is equal under the two different parameterizations with probability 1, as the number of observations goes to infinity. It follows that if two different parameterizations of a system are ML unresolvable, then the system is not ML identifiable.

As we mentioned in Section 1, the model described by the equations (1) is not ML identifiable in the general case. There are many different groups of system matrices (F, H) which can produce a given data set. An alternative would be to study the associated Kalman filter of the model, but this approach also suffers from the same problem, since the mapping of a state-space model described by (1) to Kalman filter is many-to-one [11]. However, the parameters of the associated steady-state Kalman Filter are identifiable as long as \((F, H)\) is in a certain form specified below, \(F\) is stable, \((F, H)\) is an observable pair and \((F, K)\) is a controllable pair [10].

The form of the matrices we will apply to our work was first developed by Luenberger in [3] and Bucy in [12]. To describe the form, we will use some nonnegative numbers \(p_i\), where \(p_1 + p_2 + \ldots + p_m = n\). The way \(p_i\) are calculated can be found in [10]. Using these numbers, the structure of the matrix \(F\) is as follows:

\[
F = \begin{bmatrix}
F_{11} & 0 & 0 & \ldots & 0 \\
F_{21} & F_{22} & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
F_{m1} & F_{m2} & F_{m3} & \ldots & F_{mm}
\end{bmatrix}
\]

where \(F_{ii}\) is a \(p_i \times p_i\) matrix and \(F_{ij}, i > j\) is a \(p_i \times p_j\) matrix. The forms of \(F_{ii}\) and \(F_{ij}\) are:

\[
F_{ii} = \begin{bmatrix}
0 & 1 & 0 & \ldots & 0 \\
0 & 0 & 1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 1 \\
x & x & x & \ldots & x
\end{bmatrix},
F_{ij} = \begin{bmatrix}
0 & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
x & x & \ldots & x
\end{bmatrix}
\]

As shown above, \(F_{ii}\) has ones in its superdiagonal, its last row is filled with free parameters and every other element is set to zero, while \(F_{ij}\) is a zero matrix except from its last row which is filled with free parameters.

As for the structure of \(H\), each row has the following form

\[
h_i = [0, \ldots, 0, 1, 0, \ldots, 0] \text{ if } p_i > 0
\]

where the 1 is in the \(1 + p_1 + \ldots + p_{i-1}\) position and

\[
h_i = [x, x, \ldots, x, 0, \ldots, 0] \text{ if } p_i = 0
\]

where the free parameters occupy the first \(p_1 + \ldots + p_{i-1}\) elements of the row vector.

If the system matrices follow this form, then it can be shown [10] that the system is ML identifiable and in fact its associated Kalman Filter is also identifiable, which means that the mapping of the model to its steady state Kalman Filter is one-to-one. In the following section we derive the ML estimates of the system parameters using the EM algorithm.

### 3. ESTIMATION WITH THE EM ALGORITHM

The EM([9]) is a two-step iterative algorithm and is usually used for parameter estimation in problems with missing data. During the Expectation (E) step, the expected log-likelihood \(L\) of the complete data (both the observed and the missing) is calculated based on the observed data and the current parameter estimates \(\theta^k\). In the Maximization (M) step, the expected log-likelihood is maximized with respect to the new parameters \(\theta^{k+1}\).

To apply the EM algorithm for the estimation of a linear state-space system, we use the form (1) instead of (2), since it can be shown([13]) that the innovations representation is not uniquely identifiable through the auxiliary function of EM. On the contrary, the parameterization (1) can be uniquely determined by the auxiliary function.

In our case, the observed data are the observations \(Y = [y_0 \ldots y_N]\) and the missing data are the states \(X = [x_1 \ldots x_N]\). Since in (1) \(w_k, v_k\) are zero-mean gaussian noise processes and by use of the Markov property that \(x_k\) is conditionally independent from \(x_{k-2}\) given \(x_{k-1}\), the log-likelihood of the complete data for the current values of the parameters \(\theta\) is given by:

\[
L(X, Y, \theta) = -\sum_{k=1}^{N} \{ \log |Q| + (x_k - F x_{k-1})^T Q^{-1} (x_k - F x_{k-1}) \} + c
\]

The free-parameter vector \(\theta\) is “incorporated” in the system matrices \(F, H, Q, R\) as described in the canonical form presented above, and \(c\) is a constant quantity. Throughout the paper the base of \(\log\) is \(e\). The quantity we need to maximize in the EM procedure at each iteration \(p\) is

\[
Q(\theta^{(p+1)} | \theta^{(p)}) = E_{\theta^{(p)}} \{ L(X, Y, \theta^{(p+1)}) | Y \} \]

In the E-step of the EM algorithm, we need to calculate the following quantities [14]:

\[
E_{\theta^{(p)}} \{ x_k^T | Y \} = \hat{x}_{k|N}
\]

\[
E_{\theta^{(p)}} \{ x_k x_k^T | Y \} = E_{\theta^{(p)}} \{ x_k | Y \} x_k^T + \hat{x}_{k|N} \hat{x}_{k|N}^T
\]

\[
E_{\theta^{(p)}} \{ x_k x_{k-1}^T | Y \} = \Sigma_{k-1|N} + \hat{x}_{k|N} \hat{x}_{k-1|N}^T
\]

By applying the fixed interval smoothing form of the Kalman Filter (RTS Smoother), we can estimate the conditional mean and variance of the state given the observations.
\[ \hat{x}_{k|N} \text{ and } \Sigma_{k|N}. \] The estimation of the conditional cross-covariance of the state process \( \Sigma_{k,k-1|N} \) is not included in the classic equations of the Kalman Filter. In [14] it has been shown though that \( \Sigma_{k,k-1|N} = (I - K_k^1 H) F \Sigma_{k-1,k-1}. \) Since we use the steady state Kalman Filter, we solve the Discrete Algebraic Riccati Equation (DARE)

\[
\Sigma = F \Sigma H T - F \Sigma H T (H \Sigma H T + R)^{-1} H \Sigma F T + Q \quad (7)
\]

and then using the Kalman filter we get the quantities we need to estimate in the E-step. Existence of a unique solution in equation (7) is guaranteed, since we examine only stable systems.

To find new estimates of the parameters in the M-step, note that the parameters that appear in a system matrix are independent from the parameters that appear in any other. Therefore we can examine the two terms of equation (4) separately. For simplicity, we will first show the derivation straight on \( L(X,Y,\theta) \) instead of \( E \{ L(X,Y,\theta) \} \) and only for the first term, since the derivation of the second term is similar.

Under light of the proposition 5.1 in [14] the likelihood is maximized by:

\[
\hat{\theta} = \arg \min_{\theta} \left( \text{logdet} \left[ \frac{1}{N} \sum_{k=1}^{N} e_k(\theta) e_k^T(\theta) \right] \right) = \arg \min_{\theta} \left( \text{tr} \left\{ \log \left[ \frac{1}{N} \sum_{k=1}^{N} e_k(\theta) e_k^T(\theta) \right] \right\} \right) \quad (8a)
\]

\[
\hat{\theta} = \frac{1}{N} \sum_{k=1}^{N} e_k(\theta) e_k^T(\theta) \quad (8b)
\]

where \( e_k(\theta) = x_k - F x_{k-1} \). We assume that \( \sum_{k=1}^{N} e_k(\theta) e_k^T(\theta) \) is full rank. Since we want to preserve the structure of \( F \), we differentiate the trace in (8a) only with respect to the free elements of matrix \( F \). The derivation is quite elaborate so we will present the results with the key insights that led to them.

By expanding the trace of (8a) and taking the partial derivative with respect to \( f_{ij} \), we get a quantity which includes elements of the same row. By construction, the rows of \( F \) may contain free parameters in their left-most positions, while the remaining elements are zero. This is equivalent to taking the partial derivatives of the first \( r \) elements of the \( i \)-th row, and using some elementary linear algebra we get the following result:

\[
[f_{11}, f_{12}, \ldots, f_{1r}]^{(\text{new})} = \frac{N}{\left[ \sum_{k=1}^{N} x_k x_k^T \right]}_{[i,1:r]} \left[ \sum_{k=1}^{N} x_{k-1} x_{k-1}^T \right]_{[1:r,1:r]} \quad (9)
\]

where \([i,1:r]\) denotes the first \( r \) elements of row \( i \) and \([1:r,1:r]\) denotes the upper square \( r \times r \) matrix. Contrary to \( F \), the covariance matrix \( Q \) is filled with free parameters, so we take partial derivatives with respect to all of its elements. By expanding (8b) and based on our new estimate for \( F \) we have:

\[
\hat{Q} = \frac{1}{N} \left( \sum_{k=1}^{N} x_k x_k^T - \sum_{k=1}^{N} x_{k-1} x_{k-1}^T \right) \hat{F}^T \quad (10)
\]

Returning to the problem of maximizing \( E \{ L(X,Y,\theta) \} \), since the processes \( X,Y \) are jointly Gaussian and belong to the exponential family, we can replace in equations (9), (10) the sufficient statistics of the state process \( X \) with their expected values from equations (6). The derivation of \( \hat{R} \) and \( \hat{Q} \) is analogous with \( \hat{F} \) and \( \hat{Q} \), respectively.

In summary, our algorithm is described by the following steps:

1. Initialize \( F,H,Q,R \)
2. Solve the DARE equation (7) to find \( \Sigma_{k|k-1} \) as \( k \to \infty \) and the SSKG.
3. Apply the steady state Kalman Filter to the data set \( Y \) and then the RTS smoother
4. Collect sufficient statistics
5. reestimate \( F,H,Q,R \) through the statistics gathered in previous step
6. return to step 2 and solve the DARE with the new matrices

Steps 2-4 are the E-step of the EM algorithm, while step 5 is the M-step. In the next section we will present the experimental results of our method.

4. EXPERIMENTAL RESULTS

In our experiments we have applied our algorithm to a number of different cases of the canonical form we presented in Section 2, varying the dimensions of \( m \) and \( n \). In each case, we generated 10000 observation vectors from which we estimate the system matrices. One such example is the following, where the system matrices and the true SSKG \( K \) were:

\[
F = \begin{bmatrix}
0 & 1 & 0 \\
0 & 0 & 1 \\
-0.2 & 0.1 & -0.1
\end{bmatrix}, \quad H = \begin{bmatrix}
1 & 0 \\
0.3 & 0.3 \\
0 & 0
\end{bmatrix},
\]

\[
Q = R = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix},
\]

\[
K = \begin{bmatrix}
2.6856 & -0.1408 & 0.2499 \\
-0.1408 & 2.0025 & -0.0973 \\
0.2499 & -0.0973 & 1.0167
\end{bmatrix}
\]
Notice that $F$ is stable, $F, H$ is an observable pair and $F, K$ is a controllable pair, as we have described in Section 2. Figure 1 shows the convergence of the system matrices, by plotting the Frobenius norm of the difference of the true matrix with the estimated in each step of the algorithm. In order to check if our procedure converges to the true solution, we also examined the convergence of the SSKG. Since the mapping of a model under the specific structure we have presented with its associate steady state Kalman Filter is one-to-one, if the SSKG converges to the true solution then the system converges too to the true solution. As we can see in figure 2, the SSKG converges to its true value thus ensuring that the identification procedure estimates the true system matrices and not a linear transformation of those.

![Fig. 1. Convergence of the system matrices](image1)

![Fig. 2. Convergence of Steady State Kalman Gain](image2)

5. CONCLUSIONS

We have presented a method for ML identification of general state-space models in canonical form. To the best of our knowledge, this is the first work that solves the problem of maximum-likelihood identification of linear state-space models for arbitrary state and observation dimensions. ML identifiability is ensured by using canonical forms and the solution is obtained using the EM framework. The experimental results show good convergence properties in all cases that we examined.

6. REFERENCES


