ON THE JOINT ESTIMATION OF UNKNOWN PARAMETERS AND DISTURBANCES IN LINEAR STOCHASTIC TIME-VARIANT SYSTEMS

Stefano Perabò and Qinghua Zhang
IRISA, INRIA Rennes, Campus universitaire de Beaulieu, Avenue du General Leclerc, 35042 Rennes Cedex, France
sperabo@irisa.fr, zhang@irisa.fr

Keywords: Fault detection, parameters estimation, linear stochastic time varying systems, adaptive signal processing.

Abstract: Motivated by fault detection and isolation problems, we present an approach to the design of unknown parameters and disturbances estimators for linear time-variant stochastic systems. The main features of the proposed method are: (a) the joint estimation of parameters and disturbances can be carried out; (b) it is a full-stochastic approach: the unknown parameters and disturbances are random quantities and prior information, in terms of means and covariances, can be easily taken into account; (c) the estimator structure is not fixed *a priori*, rather derived from the optimal infinite dimensional one by means of a sliding window approximation. The advantages with respect to the widely used *parity space* approach are presented.

1 INTRODUCTION

The following discrete time linear stochastic system is considered in this brief paper:

$$\mathrm{x}_{k+1} = A_k \mathrm{x}_k + B_k \mathrm{u}_k + \Psi_k \mathrm{p} + E_k \mathrm{d}_k + \mathrm{w}_k \quad (1a)$$

$$\mathrm{y}_k = C_k \mathrm{x}_k + \mathrm{v}_k \quad (1b)$$

for $k \geq 0$, with $A_k \in \mathbb{R}^{n\times n}$, $B_k \in \mathbb{R}^{n\times m}$, $\Psi_k \in \mathbb{R}^{n\times q}$, $E_k \in \mathbb{R}^{n\times f}$ and $C_k \in \mathbb{R}^{l\times n}$ known time-variant matrices. The vector sequences $\{\mathrm{x}_k\}$, $\{\mathrm{u}_k\}$ and $\{\mathrm{y}_k\}$ denote respectively the state, input and output stochastic processes. The sequences $\{\mathrm{w}_k\}$ and $\{\mathrm{v}_k\}$ are assumed to be zero mean, white and uncorrelated wide-sense stochastic processes, with $\mathbb{E}[\mathrm{w}_k \mathrm{w}_k^T] = Q_k$ and $\mathbb{E}[\mathrm{v}_k \mathrm{v}_k^T] = R_k \succ O$ (positive definite), where $\mathbb{E}[\cdot]$ denotes the mathematical expectation operator. The initial condition x_0 has known mean $\mathbb{E}[\mathrm{x}_0] = \mu_0$ and covariance $\mathbb{E}[(\mathrm{x}_0 - \mu_0)(\mathrm{x}_0 - \mu_0)^T] = P_0$. Both the initial condition x_0 and the input process $\{\mathrm{u}_k\}$ are assumed uncorrelated with the noise sequences.
The term $E_k \mathrm{d}_k$ accounts for unknown disturbances acting on the system or faults, whence the sequence $\{\mathrm{d}_k\}$ is an unknown (and uncontrolled) input modeled as a wide-sense stochastic process, not necessarily stationary. The disturbances are further assumed uncorrelated with the initial state, the noise and the input processes, respectively x_0, $\{\mathrm{w}_k\}$, $\{\mathrm{v}_k\}$ and $\{\mathrm{u}_k\}$. Finally, the term $\Psi_k \mathrm{p}$ can account for the occurrence of parametric faults in the system (for instance with the meaning that when p is zero no faults are present) or for constant parameters that need to be estimated on-line. Here p is a random variable uncorrelated with the noise, input and disturbance processes.

The problem to be solved is the following: find for each $N \geq 0$ the *minimum variance unbiased linear estimators* of the disturbances sequence $\mathrm{d}_0^{N-1} = \{\mathrm{d}_k : 0 \leq k \leq N-1\}$ and of the parameters p, given the input and output sequences u_0^{N-1} and y_0^N, and the conditions guaranteeing the uniqueness of the corresponding estimates. These estimators will be denoted respectively by $\hat{\mathrm{d}}_{k|N}$ and $\hat{\mathrm{p}}_{|N}$ (since p does not depend on time).

The following two related problems will also be discussed in this paper. First, how to weaken the uniqueness conditions by considering the quantities $\hat{\mathrm{d}}_{k|N+D}$ for $0 \leq k \leq N-1$ and some appropriate delay $D > 0$, which will be called, with an abuse of terms, "delayed estimators". Second, how to *recursively* and reliably compute the estimates $\hat{p}_{|N+D}$ and $\hat{d}_{k|N+D}$ once sample paths (measurements) u_0^{N+D-1} and y_0^{N+D} of the input and output processes are becoming available (by convention, italic characters will denote samples from the corresponding random variables which, in-

stead, will be denoted by roman characters).

The solution proposed in this work shares many similarities with the so called *parity space* method (Chow and Willsky, 1984; Gustafsson, 2001) which finds wide application in fault detection problems. However it has some advantageous features that will be presented at the end of the exposition.

Once the disturbances and parameters estimates have been computed, state estimation becomes straightforward and can also be easily performed *on demand*. This topic is discussed in (Perabò and Zhang, 2007).

2 BASIC EQUATIONS FOR ESTIMATION

Pretend for a while that the parameters and the disturbances sequence are known quantities, i.e. as if they were inputs of the system described by (1), and assume the following:

Assumption 1. (A_k, C_k) is *uniformly completely observable* and $(A_k, Q_k^{1/2})$ is *uniformly completely reachable*.

Assumption 2. The parameters p and the disturbance sequence $\{\mathrm{d}_k\}$ are uncorrelated from the initial state x_0 and the noise sequences $\{\mathrm{w}_k\}$ and $\{\mathrm{v}_k\}$.

Hence there is *no feedback* from the output to the parameters and disturbances (see (Gevers and Anderson, 1982) for details) and by applying well known results of the linear estimation theory (Kailath et al., 2000), the following *innovation representation* of the output process $\{\mathrm{y}_k\}$ can be derived:

$$\hat{\mathrm{x}}^*_{k+1|k} = A_k\hat{\mathrm{x}}^*_{k|k-1} + B_k\mathrm{u}_k + \Psi_k\mathrm{p} + E_k\mathrm{d}_k + K_k\mathrm{e}^*_k \quad (2a)$$

$$\mathrm{y}_k = C_k\hat{\mathrm{x}}^*_{k|k-1} + \mathrm{e}^*_k, \quad (2b)$$

the recursion being initiated setting $\hat{\mathrm{x}}^*_{0|-1} = \mathrm{x}_0$, where $\hat{\mathrm{x}}^*_{k+1|k}$ is the one step *minimum variance unbiased linear predictor* of the state. Each term of the *innovation sequence* $\{\mathrm{e}^*_k\}$ has zero mean and covariance Λ_k given by the recursive solution of the same Riccati equation which is solved in the standard Kalman filter (i.e. with no disturbances and unknown parameters). With respect to this one, however, the superscript * in (2) emphasizes that the "estimates" $\{\hat{x}^*_{k+1|k}\}$ cannot be computed because the realizations p and $\{d_k\}$, of p and $\{\mathrm{d}_k\}$ respectively, are not really available. Also the gains K_k are computed exactely as in the Kalman filter. By defining recursively the quantities,

$$\Upsilon_0 = O \quad \Upsilon_{k+1} = (A_k - K_kC_k)\Upsilon_k + \Psi_k \quad (3a)$$

$$\mathrm{s}_0 = 0 \quad \mathrm{s}_{k+1} = (A_k - K_kC_k)\mathrm{s}_k + E_k\mathrm{d}_k \quad (3b)$$

$$\mathrm{z}_0 = \mathrm{x}_0 \quad \mathrm{z}_{k+1} = (A_k - K_kC_k)\mathrm{z}_k + B_k\mathrm{u}_k + K_k\mathrm{y}_k \quad (3c)$$

and by using (2b) it is not difficult to check that the following is true:

$$C_k\mathrm{s}_k + C_k\Upsilon_k\mathrm{p} + \mathrm{e}^*_k = \mathrm{y}_k - C_k\mathrm{z}_k. \quad (4)$$

Note that a realization of the sequence $\{\mathrm{z}_k\}$ can be computed from available data only, i.e. system matrices, input and output sequences. As a matter of fact, (3c) is exactly the Kalman filter equation that would be obtained if $\mathrm{p} \equiv 0$ and $\mathrm{d}_k \equiv 0$ for all k.
It is possible to arrange in matrix form the set of equations obtained from (4) when $k = 1, 2, \ldots, N$. For example for $N = 4$ one obtains

$$\begin{bmatrix} C_4\Phi_1^4E_0 & C_4\Phi_2^4E_1 & C_4\Phi_3^4E_2 & C_4E_3 & C_4\Upsilon_4 \\ C_3\Phi_1^3E_0 & C_3\Phi_2^3E_1 & C_3E_2 & O & C_3\Upsilon_3 \\ C_2\Phi_1^2E_0 & C_2E_1 & O & O & C_2\Upsilon_2 \\ C_1E_0 & O & O & O & C_1\Upsilon_1 \end{bmatrix} \begin{bmatrix} \mathrm{d}_0 \\ \mathrm{d}_1 \\ \mathrm{d}_2 \\ \mathrm{d}_3 \\ \mathrm{p} \end{bmatrix} + \begin{bmatrix} \mathrm{e}^*_4 \\ \mathrm{e}^*_3 \\ \mathrm{e}^*_2 \\ \mathrm{e}^*_1 \end{bmatrix} = \begin{bmatrix} \mathrm{y}_4 - C_4\mathrm{z}_4 \\ \mathrm{y}_3 - C_3\mathrm{z}_3 \\ \mathrm{y}_2 - C_2\mathrm{z}_2 \\ \mathrm{y}_1 - C_1\mathrm{z}_1 \end{bmatrix}, \quad (5)$$

where the transition matrices Φ_h^k are defined by

$$\Phi_h^h = I, \qquad \Phi_h^{k+1} = (A_k - K_kC_k)\Phi_h^k. \quad (6)$$

For an arbitrary N, left multiply the above system by the block diagonal matrix $\mathrm{blkdiag}\{\Lambda_N^{-1/2}, \ldots, \Lambda_1^{-1/2}\}$ in such a way that the covariance of the zero mean vector $\mathrm{e}^* = \mathrm{vec}[\Lambda_N^{-1/2}\mathrm{e}^*_N \ \ldots \ \Lambda_1^{-1/2}\mathrm{e}^*_1]$ is equal to the identity matrix. A system of the form

$$\mathsf{A}\mathsf{g} + \mathsf{e}^* = \mathsf{r} \quad (7)$$

is thus obtained, where the matrix $\mathsf{A} \in \mathbb{R}^{lN\times(fN+q)}$ has the same structure as in (5), $\mathsf{g} = \mathrm{vec}[\mathrm{d}_0 \ \ldots \ \mathrm{d}_{N-1} \ \mathrm{p}]$ is the unknown term, and the vector $\mathsf{r} = \mathrm{vec}[\mathrm{r}_N \ \ldots \ \mathrm{r}_1]$ contains the computable *residuals*

$$\mathrm{r}_k = \Lambda_k^{-1/2}(\mathrm{y}_k - C_k\mathrm{z}_k). \quad (8)$$

If $\mathrm{d}_k \equiv 0$ for each k and $\mathrm{p} \equiv 0$, then $\mathsf{r} = \mathsf{e}^*$, i.e. the vector of residuals has zero mean and its covariance equals the identity matrix. Any statistical test indicating a deviation from this condition can be used to detect the presence of non-null disturbances and/or parameters.

Since samples of r are available but instead e cannot be observed, the most appealing approach to estimate g is to compute its minimum variance linear estimator $\hat{\mathsf{g}}$ given the random vector r. Thanks to the Assumption 2, the following holds:

$$\mathbb{E}[\mathrm{e}^*_k\mathrm{d}_h^T] = O \quad \mathbb{E}[\mathrm{e}^*_k\mathrm{p}^T] = O \quad \forall k, h \geq 0. \quad (9)$$

As a result, g and e^* in (7) are in fact uncorrelated. Provided that *prior information* on the random vector g is given in terms of its mean μ_g and covariance Σ_g (assume Σ_g invertible and the factorization $\Sigma_g^{-1} = B^T B$), a straightforward application of linear estimation formulas shows that $\hat{g}$ and the covariance of the error $\tilde{g} = g - \hat{g}$ can be obtained from

$$\left(A^T A + B^T B\right)(\hat{g} - \mu_g) = A^T(r - A\mu_g) \qquad (10a)$$

$$\Sigma_{\tilde{g}} = \left(A^T A + B^T B\right)^{-1}. \qquad (10b)$$

One could suspect, at this point, that the *information* about the unknown terms which is available from knowledge of the input and output sequences, is not fully exploited if the only quantities that are used for the estimation of the disturbances and parameters are the residuals defined in (8). However, as long as linear estimators are considered, it is possible to prove that the proposed method is *optimal* in the sense that, by estimating g from (7) (instead of a different linear relation with the measurable sequences $\{u_0^{N-1}, y_0^N\}$) one in fact minimizes the estimation error variance.

When sample paths of the input and output sequences, say $\{u_0^{N-1}\}$ and $\{y_0^N\}$, are available, one is faced to the problem of computing numerically the estimate $\hat{g} = \mathrm{vec}[\hat{d}_{0|N} \ \dots \ \hat{d}_{N-1|N} \ \hat{p}_{|N}]$ from the vector r denoting the realization of r. To this end, the availability or lack of prior information makes a difference. In the following the latter case is discussed.

3 NO PRIOR INFORMATION

3.1 Estimability Conditions

The absence of prior information about g can be dealt with by setting $\mu_g = 0$ and letting $\Sigma_g \to \infty$ (or equivalently $\Sigma_g^{-1} \to 0$) which corresponds to a very large uncertainty. Formula (10a) becomes $\left(A^T A\right)\hat{g} = A^T r$ which is the system of *normal equations* for computing the unique *least squares solution* of

$$Ag = r. \qquad (11)$$

in the unknown g, provided that the matrix A has full column rank. From a practical point of view, It should be noted that the proposed method requires simply checking the rank of matrices and solving least squares problems, for which efficient numerical tools are readily available. But, unfortunately, finding general estimability condition in analytic form, is a very complex task. The following is not difficult to prove:

Proposition 1. *For a given $N \geq 1$, the estimates $\hat{p}_{|N}$ and $\hat{d}_{k|N}$ for $0 \leq k \leq N-1$ are unique if and only if the matrix* A *in* (11) *has full column rank. Moreover, the uniqueness holds only if the following necessary conditions are satisfied:*

$$\text{(C1)} \quad \mathrm{rank} \begin{bmatrix} E_0 & O & \cdots & O & \Psi_0 \\ O & E_1 & \cdots & O & \Psi_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & O & E_{N-1} & \Psi_{N-1} \end{bmatrix} = rN + q \quad (12a)$$

$$\text{(C2)} \quad \mathrm{rank}\left(\sum_{k=1}^{N} \Upsilon_k^T C_k^T C_k \Upsilon_k\right) = q. \qquad (12b)$$

If $\mathrm{rank}(E_k) = f$ *for all* $k \geq 0$ *and* (C1) *is true for a value* $N = N_{min}$*, then it is satisfied for all values* $N \geq N_{min}$*. Analogously, if* (C2) *is true for a value* $N = N_{min}$*, then it is satisfied for all values* $N \geq N_{min}$*.*

3.2 Delayed Estimation

Consider first the case when there are *no unknown parameters* ($q = 0$). A *sufficient* (but *not necessary*) condition to ensure that A has full column rank for all $N \geq 1$, hence the uniqueness of the estimates $\hat{d}_{k|N}$ for $0 \leq k \leq N-1$, is the following:

$$\text{(C3)} \qquad \mathrm{rank}(C_{k+1}E_k) = f \qquad \forall k \geq 0. \qquad (13)$$

However, when (C3) is not satisfied, it could still be possible to compute, for some delay $D > 0$, unique *delayed estimates* $\hat{d}_{k|N+D}$ for $0 \leq k \leq N-1$. To exemplify what has been just asserted, consider the case $C_{k+1}E_k = O$ and thus (C3) is not satisfied (this situation may happen typically when C_{k+1} and E_k have both some zero entries, for example $C_{k+1} = [1\ 0]$ and $E_k = [0\ 1]^T$). Then the zero blocks appear in the term Ag in (7) as shown in the following scheme (suppose, for example, that $N = 4$):

$$\begin{array}{r|c|c|} 5 & \times & \times & \times & * & O \\ \hline N=4 & \times & \times & * & O & O \\ 3 & \times & * & O & O & O \\ 2 & * & O & O & O & O \\ 1 & O & O & O & O & O \end{array} \quad \begin{bmatrix} \mathrm{d}_0 \\ \mathrm{d}_1 \\ \mathrm{d}_2 \\ \mathrm{d}_3 \\ \hline \mathrm{d}_4 \end{bmatrix}.$$

It is evident that d_{N-1} (d_3 in the example above) is not estimable from measurements collected till time N (in other words $\hat{d}_{3|4}$ is not unique). However, if the blocks marked with a $*$, i.e. the matrices $C_{k+2}\Phi_{k+1}^{k+2}E_k$ in (7), have full column rank, it is sufficient to add the measurements at time $N+1$ (at time 5 to continue the example) so that the unique estimates $\hat{d}_{k|N+1}$ for $k = 0, \dots, N-1$ and, in particular $\hat{d}_{N-1|N+1}$ (in the example $\hat{d}_{3|5}$), could be computed. The above argument can be generalized as follows: if for some $D > 0$ the conditions

$$\text{(C4a)} \quad \mathrm{rank}(C_{k+D+1}\Phi_{k+1}^{k+D+1}E_k) = f \ \ \forall k \geq 0 \quad (14a)$$

$$\text{(C4b)} \quad \begin{bmatrix} C_{k+D}\Phi_{k+1}^{k+D}E_k \\ \cdots \\ C_{k+2}\Phi_{k+1}^{k+2}E_k \\ C_{k+1}E_k \end{bmatrix} = O \qquad (14b)$$

are satisfied, then the estimates $\hat{d}_{k|N+D}$ for $0 \leq k \leq N-1$ are unique (even if A has not full rank).

When there are unknown parameters ($q > 0$), the conditions in (13) or (14) are no longer sufficient and, in general, the rank of the matrix A has to be checked numerically. However note the following result:

Proposition 2. *(a) Assuming that condition* (C3) *in* (13) *is satisfied, if the estimates $\hat{p}_{|N}$ and $\hat{d}_{k|N}$ for $0 \leq k \leq N-1$ are unique (i.e. the matrix A has full column rank) for a value $N = N_{min}$, then they are unique also for all $N \geq N_{min}$.*
(b) Analogously, assuming that conditions (C4) *in* (14) *are satisfied, if the delayed estimates $\hat{p}_{|N+D}$ and $\hat{d}_{k|N+D}$ for $0 \leq k \leq N-1$ are unique for a value $N = N_{min}$, then they are unique also for all $N \geq N_{min}$.*

3.3 Approximate Recursive Estimation

In order to compute the estimates from (11), a growing size least squares problem as to be solved as N increases. Observe, however, that the upper left blocks of the matrix A tend to zero as N grows, because the uniform observability and reachability assumption guarantees that the transition matrices Φ_h^k defined in (6) tend to the null matrix as the difference $k-h \to \infty$. Hence, it is natural to consider an approximate problem by replacing A with $\mathsf{A}+\mathsf{E}$, where E annihilates the blocks $\Lambda_k^{-1/2} C_k \Phi_h^k E_{h-1}$ such that $k-h \geq L \geq L_{min}$, where $L_{min} \geq 1$ is the minimum value guaranteeing that $\mathrm{rank}(\mathsf{A}) = \mathrm{rank}(\mathsf{A}+\mathsf{E})$ for all N, so that the estimability properties of the original problem are conserved also in the approximate one. Obviously, the accuracy of the approximate solution increases as $L \to \infty$. The system $(\mathsf{A}+\mathsf{E})g = r$ has thus the *banded* structure shown in the following scheme (for $N = 5$ and $L = 3$):

$$\begin{bmatrix} & & \times & \times & \times \\ & \times & \times & \times & \\ \boxed{\times} & \boxed{\times} & \boxed{\times} & & \boxed{\times} \\ \boxed{\times} & \boxed{\times} & & & \boxed{\times} \\ \boxed{\times} & & & & \boxed{\times} \end{bmatrix} \cdot \begin{bmatrix} \boxed{d_0} \\ \boxed{d_1} \\ \boxed{d_2} \\ d_3 \\ d_4 \\ \boxed{p} \end{bmatrix} = \begin{bmatrix} r_5 \\ r_4 \\ \boxed{r_3} \\ \boxed{r_2} \\ \boxed{r_1} \end{bmatrix}$$

In the above, also an initial data window has been indicated with a solid line box. Using the numerical techniques described for example in (Björck, 1996, Chapter 6.2), this approximate least squares problem can then be solved recursively using a sliding window procedure.

3.4 Comparison with the Parity Space Approach

In the parity space method, the parameters and disturbances are estimated from a set of relations which can be cast in the form $\bar{\mathsf{A}}\mathsf{g} + \mathsf{w} = \bar{\mathsf{r}}$. The matrix $\bar{\mathsf{A}}$ differs from A in (7) only because the transition matrices Φ_h^k defined in (6) are replaced by $\Gamma_h^k = A_{k-1} \ldots A_{h+1} A_h$. Moreover, the covariance of the noise term w does not equal the identity matrix and the residuals $\bar{\mathsf{r}}$ are built in a different way.
The approach proposed here is new in that it makes explicit reference to the innovation representation of the system (1), with the following advantages:

(a) The components of the noise term e^* are independent and normalized, while an important drawback of the parity space approach is that the covariance of the noise term w has to be *whitened* before computing the least squares estimate, thus increasing the computational load, especially for large scale problems.

(b) If the matrices A_k are not stable, as it can happen typically in control problems, the matrix $\bar{\mathsf{A}}$ could be largely ill-conditioned, thus making numerically harder the process of computing reliably the estimate, especially for large window sizes.

(c) The initial condition x_0 affects the residuals r through the sequence $\{\mathrm{z}_k\}$. However the transition matrices Φ_h^k are stable. Hence the effect of the initial condition is asymptotically forgot as $k \to +\infty$. As a consequence, when using the sliding window estimation procedure, one has not to take care of the estimation or rejection of the state at the initial time of the window as happens for the parity space approach (Törnqvist and Gustafsson, 2006).

REFERENCES

Björck, A. (1996). *Numerical Methods for Least Squares Problems*. SIAM.

Chow, E. Y. and Willsky, A. S. (1984). Analytical redundancy and the design of robust failure detection systems. *IEEE T. Automat. Contr.*, AC-29(7):603–614.

Gevers, M. R. and Anderson, B. D. O. (1982). On jointly stationary feedback-free stochastic processes. *IEEE T. Automat. Contr.*, AC-27(2):431–436.

Gustafsson, F. (2001). *Adaptive Filtering and Change Detection*. John Wiley & Sons, Ltd.

Kailath, T., Sayed, A. H., and Hassibi, B. (2000). *Linear Estimation*. Prentice Hall.

Perabò, S. and Zhang, Q. (2007). Adaptive observers for linear time-variant stochastic systems with disturbances. In *Proceedings of the European Control Conference 2007*, Kos, Greece. (accepted for publication).

Törnqvist, D. and Gustafsson, F. (2006). Eliminating the initial state for the generalized likelihood ratio test. In *Proceedings of the 6^{th} IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*, pages 643–648, Beijing, P. R. China.