Multi-Model System Parameter Estimation

Ivan Markovsky, Sabine Van Huffel, and Bart De Moor

SCD-SISTA, Electrical Engineering department of K.U. Leuven
Kasteelpark 10, B-3001 Leuven-Heverlee, Belgium
{ivan.markovsky, sabine.vanhuffel, bart.demoor}@esat.kuleuven.ac.be

Abstract— **We pose a multi-model system parameter estimation problem. A multi-model system is a linearly parameterized system $H(z,p) = \sum_{i=1}^{n_p} p_i H_i(z)$. The parameter estimation problem is: given the set of systems $\{H_i(z)\}_{i=1}^{n_p}$, describing the multi-model system, find a causal system that assumes as an input the input/output signals of the multi-model system and produces as an output the parameter estimate.**

We propose an easy to implement suboptimal solution. The algorithm that realizes it selects the best linear combination of the estimates produced by the Kalman filters designed for the models $\{H_i(z)\}_{i=1}^{n_p}$. "Best" is defined in the sense of minimization of the output error of estimation covariance.

The algorithm is appropriate for fault detection and can be viewed as an observer for the discrete state of a hybrid system.

Keywords— **Multi-model systems, Kalman filtering, Recursive parameter estimation, Fault detection, Hybrid systems.**

I. Introduction: multi-model system

LET $\mathcal{S}$ be a set of systems and let p be an n_p-dimensional real vector of parameters. The *parameter space* is $\mathbb{R}^{n_p}$, the n_p-dimensional real vector space. We define abstractly a *parameterized system* S as a mapping from the parameter space to the set of systems, *i.e.*, $S : \mathbb{R}^{n_p} \to \mathcal{S}$.

The parameter vector can be viewed as a *selector* of a particular dynamical system in a subset of $\mathcal{S}$ (the image of S). In practice, we think of the parameter as a "macro state" or a "supervisory control" of the parameterized system in the sense that p can be an "interface" to a higher level control system or to a human being that supervises the system.

We will consider *linearly parameterized systems* in the class of the discrete-time linear systems. A linearly parameterized system S, also called a *multi-model systems*, is a parameterized systems such that

$$S(p) = \sum_{i=1}^{n_p} p_i S_i, \quad \text{for some} \quad S_i \in \mathcal{S}, \quad i = 1, \dots, n_p.$$

The multi-model system is completely described by the set of systems $\{S_i\}_{i=1}^{n_p}$.

Let the set of systems $\mathcal{S}$ be the set of discrete-time linear time-invariant systems. For a fixed value p of the parameter vector, $S(p)$ is a discrete-time linear time-invariant system described by the transfer function

$$H(z,p) = \sum_{i=1}^{n_p} p_i H_i(z),$$

where $H_i(z)$ is the transfer function describing the system S_i. We refer to [1] for more information about linearly parameterized systems.

II. Problem formulation: multi-model system parameter estimation

Given is the multi-model system $H(z,p)$, described by the set of systems $\{H_i\}_{i=1}^{n_p}$. The input of $H(z,p)$ is partitioned into an unmeasurable input or *disturbance* w_d and a *measurable input* u. The output is measured with *additive noise* w_n and the parameter p is unmeasurable. We consider the problem to design a causal system, called an *estimator*, that assumes as an input the measured input/output signals from the multi-model system and produces as an output the parameter estimate $\hat{p}$, see Fig. 1. We will assume that the

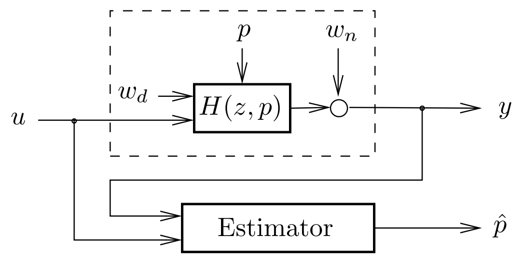

Fig. 1. Multi-model system parameter estimation problem setup.

disturbance and the measurement noise are white, zero mean random processes with known constant covariance matrices V_{w_d} and V_{w_n}, respectively. Denote by $\tilde{p}$ the error of estimation,

$$\tilde{p} := p - \hat{p}.$$

The estimation criterion is minimum estimation error covariance, *i.e.*,

$$\min_p \operatorname{tr}\left(\operatorname{Cov}\left(\tilde{p}\right)\right). \tag{1}$$

As defined, the multi-model system parameter estimation problem is a *filtering problem*. In the corresponding *smoothing problem* we consider a finite time interval $0,\ldots,t_f-1$ and solve problem (1) for the minimizing sequence $\{\hat{p}(t)\}_{t=0}^{t_f-1}$. The solution of the smoothing problem is a *structured total least squares* problem, see [2].

Partition $H_i(z)$ according to the disturbance channel and the measured input channel

$$H_i(z) = [H_{w_d,i}(z) \quad H_{u,i}(z)].$$

We denote by $H_{w_d,i}w_d$ the response of the system with transfer function $H_{w_d,i}(z)$ to the signal w_d under zero initial conditions. Similarly for the measured input channel. Note that these mappings are causal convolution operators, so that they are linear, represented by lower triangular Toeplitz matrices. Thus the notation $H_{w_d,i}w_d$ can be interpreted as a matrix-vector multiplication.

Assume zero initial conditions. The response of the i-th model to a disturbance w_d and measured input u is

$$y_i = H_{u,i}u + H_{w_d,i}w_d.$$

The response of the multi-model system to the same signals is

$$y = \sum_{i=1}^{n_p} p_i y_i + w_n.$$

Combining these equations, we have

$$[H_{u,1}u + H_{w_d,1}w_d, \ldots, H_{u,n_p}u + H_{w_d,n_p}w_d]\, p = y - w_n,$$

which expresses the smoothing problem as a "static" regression problem. The vectors u and y are known, w_d and w_n are unknown, and p is unknown and to-be-estimated.

While the disturbance enters the right-hand-side of the regression equation in a structured way, the problem is not a standard least-squares problem. This problem is called structured total least squares problem and is known to be difficult non-convex optimization problem [3]. Moreover, in the context of the filtering problem, we are interested in a recursive algorithm that updates the solution. Next we present a suboptimal but simpler solution.

III. Proposed solution

The reason why the multi-model system parameter estimation problem is difficult is that we try to filter the noise signals and estimate the parameter vector simultaneously. The problem is simplified if we separate it into two independent phases: first, filter the input/output data, and second, estimate the parameters from the filtered measurements. Clearly this approach leads to a suboptimal solution.

In the filtering stage, we process the input/output signals with a *bank of Kalman filters*, designed for the set of models $\{H_i\}_{i=1}^{n_p}$. Let $\hat{y}_i$ be the output estimate of the i-th Kalman filter. For any $p \in \mathbb{R}^{n_p}$ we interpret

$$\hat{y}(p) := \sum_{i=1}^{n_p} p_i \hat{y}_i$$

as the *predicted output*. The *output error of estimation* is

$$\tilde{y}(p) := y - \hat{y}.$$

In the estimation stage, we select as an estimate $\hat{p}$, the parameter vector that minimizes the covariance of the output error of estimation, *i.e.*,

$$\min_p \operatorname{tr}\left(\operatorname{Cov}\left(\tilde{y}(p)\right)\right). \tag{2}$$

Denote by $\hat{Y}$ the matrix of the stacked one next to the other filtered outputs

$$\hat{Y} := [\hat{y}_1 \cdots \hat{y}_{n_p}].$$

Then

$$\tilde{y}(p) = y - \hat{Y}p$$

and the solution of problem (2) is

$$\hat{p} = \mathbf{E}\{\hat{Y}\hat{Y}^T\}^{-1}\mathbf{E}\{\hat{Y}^T y\}. \tag{3}$$

While $\hat{Y}$ is computed and y is measured, the quantities

$$F := \mathbf{E}\{\hat{Y}\hat{Y}^T\} \quad \text{and} \quad h := \mathbf{E}\{\hat{Y}^T y\}$$

can be estimated in real-time by

$$\hat{F}(t) = \frac{1}{t}\sum_{\tau=0}^{t} \hat{Y}^T(\tau)\hat{Y}(\tau) \tag{4}$$

and

$$\hat{h}(t) = \frac{1}{t}\sum_{\tau=0}^{t} \hat{Y}^T(\tau)\hat{y}(\tau). \tag{5}$$

Then at the moment of the time t, the estimate is

$$\hat{p}(t) = \hat{F}^{-1}(t)\hat{h}(t) \tag{6}$$

In practice, the computation of $\hat{F}(t)$ and $\hat{h}(t)$ is done recursively by

$$\hat{F}(t) = \frac{1}{t-1}\big((t-2)\hat{F}(t-1) + \hat{Y}^T(t)\hat{Y}(t)\big)$$

and

$$\hat{h}(t) = \frac{1}{t-1}\big((t-2)\hat{h}(t-1) + \hat{Y}^T(t)y(t)\big).$$

A refinement of the algorithm is to estimate directly $\hat{F}^{-1}(t)$, so that solving the system (6) on every time instance is avoided. Furthermore, one can consider square root type algorithms, where the Cholesky factor $\hat{F}^{-1/2}(t)$ is estimated. We leave these improvements for a further research work and concentrate in this paper on the questions of the performance and the applications of the algorithm.

Next we show a simulation example. In the experiment, the parameter p is constant and the multi-model system is described by four models, *i.e.*, $n_p = 4$. Let $\mathbf{1}_i$ denotes the i-th unit vector,

$$\mathbf{1}_i := [0 \cdots 0 \underset{i}{1} 0 \cdots 0]^T.$$

The parameter vector is $p = \mathbf{1}_1$, which means that the dynamical behavior of the multi-model system coincides with this of the first model.

The simulation result is shown on the plot of Fig. 2. In red is the estimate of the first parameter. In 20 time samples $\hat{p}_1$ is sufficiently close to 1 and the estimates of the other parameters ($\hat{p}_2$ is in blue) are close to 0. The algorithm "recognizes" the correct "mode" of operation from the noisy input/output data.

Next, we modify the algorithm to allow estimation of a time varying parameter vector. In this case the estimation algorithm becomes *adaptive*. We will assume that an a priori knowledge for the rate of variation is known.

If the parameter vector is a function of time, the multi-model system is time-varying and the output is not an ergodic stochastic process. Then the expectation operations in (3) does not make sense. Nevertheless, for a short time intervals ("short" depending on the rate of change of the parameter vector) $\hat{Y}$ and y are nearly stationary.

We account for the time variation by making the averaging over a *moving window* of T past samples and weighting the data by an *exponential forgetting factor* $\lambda \in [0,1]$. With this modifications equations (4) and (5) become

$$\hat{F}(t) = \frac{1}{T}\sum_{\tau=t-T}^{t} \lambda^{t-\tau}\hat{Y}^T(\tau)\hat{Y}(\tau)$$

and

$$\hat{h}(t) = \frac{1}{T}\sum_{\tau=t-T}^{t} \lambda^{t-\tau}\hat{Y}^T(\tau)\hat{y}(\tau).$$

To demonstrate the performance of the modified estimation algorithm, we alter the simulation example described above. At time instance $t = 133$ we simulate a switching of the parameter vector from the initial value $\mathbf{1}_1$ to $\mathbf{1}_2$, *i.e.*, at time $t = 133$ the multi-model system switches from the model $H_1(z)$ to the model $H_2(z)$.

The simulation result is shown on the plot of Fig. 3. Again in red is the estimate of the first parameter and in blue is the estimate of the second parameter. After time instance $t = 20$ and before time instance $t = 133$ the algorithm correctly "recognizes" the first model as the "active" one. The switching causes a jump in the estimates and in about 100 time samples a new steady state is reached. The estimates are again correct according to the new value of the parameter vector.

IV. Adding prior knowledge in the estimation

The experiment with the switching from one model to another results in a big jump of the es-

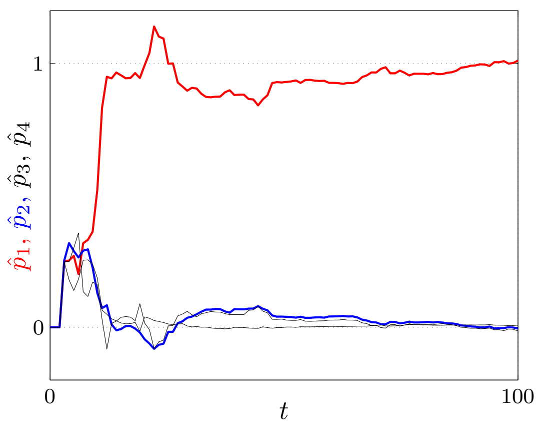

Fig. 2. Simulation result with unknown constant parameter.

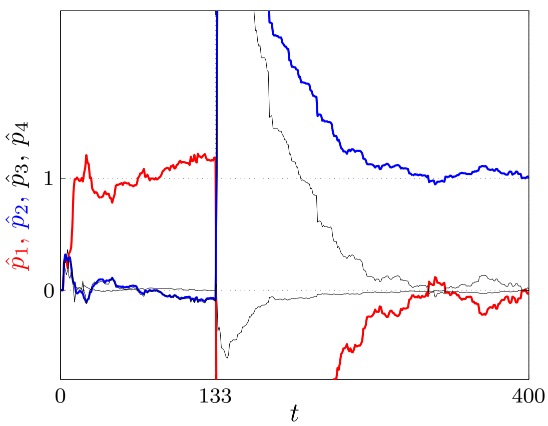

Fig. 3. Simulation result with switching.

timates. Also the convergence to the new steady state is rather slow. The performance can be improved by adding prior knowledge for the possible (or allowed) parameter values. An example of prior knowledge are upper and lower bounds for the parameters.

We use as a prior knowledge the constraint

$$\sum_{i=1}^{n_p} p_i = 1, \quad p \geq 0. \tag{7}$$

The subset in the parameter space defined by (7) is called the *probability simplex*. It allows to interpret the elements of the parameter vector as probabilities; $\hat{p}_i(t)$ is the probability, that at time t, $H_i(z)$ is active. By "$H_i(z)$ active" we mean that it governs the dynamical behavior of the multi-model system.

To account for the constraint (7), we have to solve on each iteration step the quadratic programming problem

$$\min_p \|\hat{F}p - \hat{h}\|^2 \quad \text{s.t.} \quad (7).$$

The equality constraint in (7) can be eliminated leading to another quadratic programming problem with $n_p - 1$ variables.

$$\begin{aligned} &\min_q q^T N^T \hat{F}^T \hat{F} N q - 2(\hat{h} - \hat{F}\hat{p}_0)^T \hat{F} N q \\ &\text{s.t.} \quad Nq \geq -\hat{p}_0, \end{aligned}$$

where $N := \text{Null}(\mathbf{1}^T)$ and $\hat{p}_0$ is a particular solution of $\mathbf{1}^T p = 1$, $\mathbf{1} := [1 \cdots 1]^T$.

Next, we apply the algorithm with the probability simplex constraint to the above simulation example. The plot of Fig. 4 shows a result with a constant parameter $p = \mathbf{1}_1$. For all t the estimates are confined to the interval $[0, 1]$. The convergence to the true parameter values is smoother and faster. The plot of Fig. 5 shows a result for the experiment with switching from $p = \mathbf{1}_1$ to $p = \mathbf{1}_2$. Compared with the plot on Fig. 3, the big jump is avoided and the convergence is faster.

V. Applications

In this section we discuss two areas of application of the multi-model parameter estimation problem.

The first one is *fault detection*. Assume that the real-life system, we model, has a finite number of *modes* in which it can operate and one or more of them are *faulty*. We consider the problem of designing a device that monitors the behavior of the real-life system and issues warning when it enters one of the faulty modes.

In this setting, the fault detection problem is a direct application of the algorithm in the paper. Let the modes be modeled by discrete-time linear time-invariant systems $\{H_i(z)\}_{i=1}^{n_p}$ (there are n_p modes in total) and consider the multi-model system $H(z, p) = \sum_{i=1}^{n_p} p_i H_i(z)$. Only one mode is active (*i.e.*, in use) in every moment of the time, so that for any time instance t, the parameter vector $p(t) = \mathbf{1}_{i(t)}$ for some $i(t) \in \{1, \ldots, n_p\}$. The index of the nonzero element of $p(t)$ corresponds to the index of the currently active mode. Assuming $\hat{p}(t) \approx p(t)$, the index of the largest entry of the parameter estimate $\hat{p}$ also indicates the currently active mode. The estimated current mode can be checked against membership to the set of the faulty modes.

The second application is for observer for the discrete state of a *hybrid system*. A hybrid system is a multi-model system which parameter is the state of

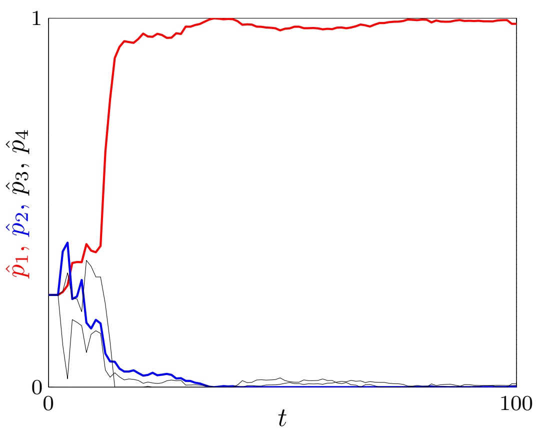

Fig. 4. Simulation result with unknown constant parameter and probability simplex constraint.

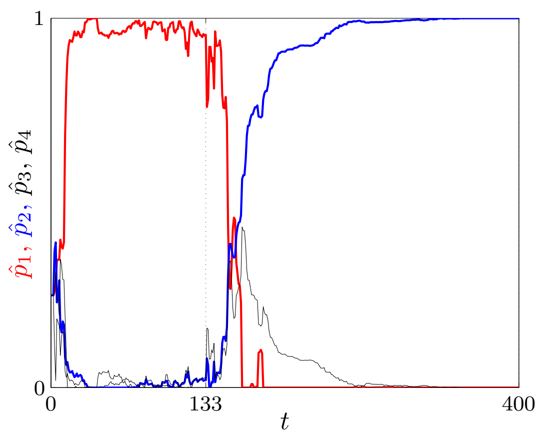

Fig. 5. Simulation result with switching and probability simplex constraint.

a discrete-event system. The estimation algorithm described in the paper can be viewed as an observer for the discrete state of a hybrid system described, by the the set of systems $\{H_i(z)\}_{i=1}^{n_p}$. While hybrid systems become popular modeling framework, the discrete state observer problem has potentially wide application domain.

VI. Conclusions

Multi-model system is a linearly parameterized system. It is a convenient tool to incorporate supervisory control or higher level discrete-event dynamics in the framework of the linear time-invariant systems. We introduced a parameter estimation problem for multi-model systems. The system is driven by a measured input and a disturbance signal and the output is measured with additive noise. The optimal solution of the multi-model parameter estimation problem is a structured total least squares problem. It is difficult to compute off-line and currently there are no recursive algorithms. We propose a simpler to implement, suboptimal solution and demonstrated by simulation examples its effectiveness. Taking into account prior knowledge improves the convergence of the estimates. We show an estimation procedure with the constraint that the parameter vector belongs to the probability simplex. This constraint makes possible to interpret the parameters as probabilities.

Acknowledgments

The first author is grateful to Professor N. Madjarov for introducing him to the topic of the multi-model systems.

I. Markovsky is a Ph.D. student, S. Van Huffel is an associate professor, and B. De Moor is a full professor at the Katholieke Universiteit Leuven, Belgium.

This paper presents research results of the Belgian Programme on Interuniversity Poles of Attraction (IUAP V-10-29), of the European Community TMR Programme, Networks, project CHRX-CT97-0160, of the Brite-Euram Programme, Thematic Network BRRT-CT97-5040 'Niconet', of the IDO/99/03 project (K.U.Leuven) "Predictive computer models for medical classification problems using patient data and expert knowledge", of the FWO projects G.0078.01, G.0200.00, and G0.0270.02.

Research Council KUL: Concerted Research Action GOA-Mefisto 666 (Mathematical Engineering), IDO (IOTA Oncology, Genetic networks), several PhD/postdoc & fellow grants;

Flemish Government: Fund for Scientific Research Flanders (several PhD/postdoc grants, projects G.0256.97 (subspace), G.0115.01 (bio-i and microarrays), G.0240.99 (multilinear algebra), G.0197.02 (power islands), G.0407.02 (support vector machines), research communities ICCoS, ANMMM), AWI (Bil. Int. Collaboration Hungary/ Poland), IWT (Soft4s (softsensors), STWW-Genprom (gene promotor prediction), GBOU-McKnow (Knowledge management algorithms), Eureka-Impact (MPC-control), Eureka-FLiTE (flutter modeling), several PhD grants);

Belgian Federal Government: DWTC (IUAP IV-02 (1996-2001) and IUAP V-10-29 (2002-2006): Dynamical Systems and Control: Computation, Identification & Modelling), Program Sustainable Development PODO-II (CP-TR-18: Sustainibility effects of Traffic Management Systems);

Direct contract research: Verhaert, Electrabel, Elia, Data4s, IPCOS;

The scientific responsibility is assumed by its authors.

References

[1] I. Markovsky, J. Suykens, and S. Van Huffel, "Linear parametric design: Approximation, estimation and control," Tech. Rep. 01–39, K.U.Leuven, Dept. EE, K.U.Leuven, December 2000, Available at: `http://www.esat.kuleuven.ac.be/~sistawww`.

[2] B. De Moor, "Structured total least squares and L_2 approximation problems," *Lin. Alg. and Its Appl.*, vol. 188–189, pp. 163–207, 1993.

[3] P. Lemmerling, *Structured total least squares: analysis, algorithms and applications*, Ph.D. thesis, ESAT/SISTA, K.U. Leuven, 1999.