Model-based Monitoring of Piecewise Continuous Behaviors using Dynamic Uncertainty Space Partitioning

Bernhard Rinner and Ulrich Weiss

Abstract. Monitoring gains importance for many technical systems such as robots, production lines or anti lock brakes. A monitoring system for technical systems must be able to deal with incomplete knowledge of the supervised system, to process noisy observations and to react within predefined time windows. This paper presents a new approach to monitoring technical systems based on imprecise models. Our approach repeatedly partitions the uncertainty space of an imprecise model and checks the derived model’s state for consistency with the measurements. Inconsistent partitions are then refuted resulting in a smaller uncertainty space and a faster failure detection. This paper further focuses on the extension of our basic approach to monitoring systems that exhibit both continuous and discrete behaviors. Our monitoring system has been implemented using COTS components and has been demonstrated in online monitoring of a non-trivial heating system.

Keywords: fault detection; hybrid systems; imprecise models; residual generation

1 INTRODUCTION

The primary objective of a monitoring system is to detect abnormal behaviors of a supervised system as soon as possible to avoid shutdown or damage. Technical systems such as robots, production lines or anti lock brakes provide a vast number of challenges for a monitoring system, i.e., it must be able to deal with incomplete knowledge about the supervised system, to process noisy observations and to react within predefined time windows.

A particularly important and widely-applied approach is model-based monitoring [6, 5] which relies on a comparison of the predicted behavior of a model with the observed behavior of the supervised system. Our approach using dynamic uncertainty space partitioning [12] is based on imprecise models where the structure of the models is known and the parameters may be imprecisely given as numeric intervals. These parameter intervals span the uncertainty space of the model. From an imprecise model based on intervals only bounds on the trajectory (envelopes) can be derived. Dynamic uncertainty space partitioning keeps the envelopes small by exploiting the measurements from the supervised system as soon as possible. Whenever new measurements arrive residuals are generated at the “corner points” of the uncertainty space and checked for consistency by comparing their signs. This results in a fast fault detection [12].

The fundamental assumption of dynamic uncertainty space partitioning is that the model’s state values are monotonic within the range of the uncertainty space. Discontinuous transitions in the system’s model may introduce non-monotonic behaviors in the state values and, therefore, violate our assumption for the consistency check. In order to preserve a conservative monitoring approach for hybrid systems, we have to extend our consistency check by a monotonicity check. Whenever the monotonicity of the state values is given the consistency check can be performed potentially resulting in a refutation of the imprecise model. If the monotonicity is not known the consistency check is simply ignored and no model is refuted.

The remainder of this paper is organized as follows. Section 2 describes the technical details of uncertainty space partitioning and the consistency check. Section 3 discusses the necessary extensions of our approach to monitoring systems which exhibit both continuous and discrete behaviors. Section 4 presents experimental results of our monitoring approach in a real-world system with several changes of a input value. A discussion and a summary of related work conclude this paper.

2 MONITORING BASED ON UNCERTAINTY SPACE PARTITIONING

2.1 Overview

Monitoring methods based on imprecise models can reason with incomplete knowledge in the model as well as with noisy measurements. A main drawback of this approach, however, is that the envelopes may diverge very rapidly which delays or even inhibits a fault recognition. We have revised this interval approach to model-based monitoring with the primary goal to keep the resulting envelopes as small as possible.

In our approach, we exploit the measurements from the supervised system as soon as possible to refine the uncertainty in the model and the derived envelopes. The key step in our approach is to partition the uncertainty space of the model into several subspaces. The trajectories derived from each subspace are then checked for consistency with the measurements. Each inconsistent subspace is refuted and excluded from further investigations. Partitioning and consistency checking are continued resulting in a smaller uncertainty space of the model. When all subspace are refuted, a discrepancy between model prediction and observation has been recognized and a fault has been detected.

2.2 Subspace Partitioning and Consistency Checking

In general, a technical system can be modeled as

\[ x_t = f(x_{t-1}, u_{t-1}, p_{t-1}) \]

\[ y_t = g(x_t, p_t) \] (1)
where \( x_t \) is the state vector at discrete time \( t \), \( u_t \) is the input vector at time \( t \), \( p_t \) is the parameter vector at time \( t \), \( y_t \) is the output vector at time \( t \), and \( g \) and \( f \) are vector functions. In an exact model, \( p_t \) is a vector of real numbers. However, in a model with uncertain parameters, \( p_t \) is replaced by a vector of intervals \( \tilde{p}_t = [(p_{1,t}, p_{1,t}], (p_{2,t}, p_{2,t}], \cdots, (p_{K,t}, p_{K,t})] \), where \( K \) is the number of uncertain parameters. A model with uncertain parameters, i.e., an imprecise model, can therefore be described as:

\[
\tilde{x}_{t+1} = f(\tilde{x}_t, u_t, \tilde{p}_t) \quad \tilde{y}_t = g(\tilde{x}_t, \tilde{p}_t)
\]

Equation 2 is the starting point of our approach. It defines an imprecise model of the supervised system with \( K \) uncertain parameters. Thus, this model has a \( K \)-dimensional uncertainty space. In order to divide this uncertainty space we have to define a partition \( \tilde{q}_t = [(\tilde{q}_{1,t}, \tilde{q}_{1,t}], (\tilde{q}_{2,t}, \tilde{q}_{2,t}], \cdots, (\tilde{q}_{K,t}, \tilde{q}_{K,t})] \) with \( \tilde{q}_t \subseteq \tilde{p}_t \).

A complete partitioning of the uncertainty space at any time \( t \) into \( M \) partitions must satisfy the following condition \( \bigcup_{m=1}^{M} \tilde{q}_m = \tilde{p}_t \) where \( m = 1, \ldots, M \). A model based on a partition of the uncertainty space is referred to as subspace model. From the definition of a partition, we can finally define the state of a subspace model:

\[
\begin{align*}
\tilde{x}^{(m)}_t &= f(\tilde{x}^{(m)}_{t-1}, u_{t-1}, \tilde{q}^{(m)}_{t-1}) \\
\tilde{y}^{(m)}_t &= g(\tilde{x}^{(m)}_t, \tilde{q}^{(m)}_t).
\end{align*}
\]

With the monotonicity assumption of \( f \) and \( g \) with regard to the parameters \( p_t \) over the range of the intervals, the (uncertain) state of a subspace model can be represented by (the exact) state of the corner points of a subspace. The corner points of a subspace are defined as all combinations of upper and lower bounds of a partition \( \tilde{q}_t \) and can be represented as set \( \tilde{Q}^{(m)}_t = \{q^{(m)}_t \} \) with \( i = 1, \ldots, 2^K \). Thus, an uncertainty space of dimension \( K \) results in \( 2^K \) corner points. The states at the corner points can be represented as set

\[
\begin{align*}
X^{(m)}_t &= \{x^{(m)}_t : x^{(m)}_t = f(x^{(m)}_{t-1}, u_{t-1}, q^{(m)}_{t-1}) \} \\
Y^{(m)}_t &= \{y^{(m)}_t : y^{(m)}_t = g(x^{(m)}_t, q^{(m)}_t) \}
\end{align*}
\]

where \( q^{(m)}_{t,ij} \) is an exact parameter vector at time \( t \) from the subspace \( m \) and at corner \( i = 1, \ldots, 2^K \) of this subspace. Note, that \( x^{(m)}_{t,ij} \) are state vectors, and also \( y^{(m)}_{t,ij} \) are output vectors with exact values. Note that this approach assumes that the parameters of the system are constant, and are not varying in time. This assumption will be discussed later.

This representation of an uncertain state is directly exploited by our consistency check for a given subspace \( m \). First, a residual is calculated for each state at a corner point using the measurements at time \( t \), i.e., \( r^{(m)}_{t,ij} = y^{(m)}_{t,measured} - y^{(m)}_{t,ij} \), where \( y^{(m)}_{t,measured} \) has the same dimension \( J \) as \( y^{(m)}_{t,ij} \). Then, the minimum and maximum values of the residual are determined as

\[
\begin{align*}
&\min_t \{r^{(m)}_{t,ij,j} \} = r^{(m)}_{t,\min,ij} \\
&\max_t \{r^{(m)}_{t,ij,j} \} = r^{(m)}_{t,\max,ij}
\end{align*}
\]

with \( i = 1, \ldots, 2^K \), and \( j = 1, \ldots, J \). Finally, subspace model \( m \) is checked for consistency simply by comparing the signs of \( r^{(m)}_{t,\min,ij} \) and \( r^{(m)}_{t,\max,ij} \). The subspace model \( m \) is consistent with the measurements, i.e.,

\[
\text{sgn}(r^{(m)}_{t,\min,ij}) \neq \text{sgn}(r^{(m)}_{t,\max,ij})
\]

holds for all elements \( j = 1, \ldots, J \).

Informally, Equation 7 checks whether the zero vector lies within the "residual subspace" (see Figure 1). If this equation is violated, the subspace model \( m \) is refuted. This simple consistency check holds also if not all elements of \( y \) are included in the measurements. In this case, a comparison with the missing elements is simply ignored. Since this technique is based on the calculation of an exact state (at corner points), we can use standard numerical methods for computing the solution of differential equations. Note that subspaces are only refuted when they are genuinely inconsistent with the measurements.

Due to the uncertainty in the parameters this method may result in diverging envelopes. This deviation of the predicted value to the ‘correct’ value over time is referred to as accumulation uncertainty. In order to keep this deviation small we have also introduced a dynamic partitioning of the subspace models. During monitoring consistent subspaces are further partitioned resulting in smaller subspace models that potentially describe the supervised system more precisely [12].

3 MONITORING PIECEWISE CONTINUOUS BEHAVIORS

3.1 Monotonicity at Transitions

In order to extend our approach to monitoring piecewise continuous behaviors and discrete transitions, we must have a closer look at our monotonicity assumption. Remember that the result of our consistency check is only valid if the state values within the subspace are monotonic.

In general the monotonicity of the state variables with regard to the parameters is not guaranteed by the monotonicity of the system equations \( f \) and \( g \). The monotonicity is only given when the following assumptions also hold:

1. the system input \( u \) does not change, and
2. the initial values of a subspace model are the same over its complete uncertainty space.

Both assumptions are important for monitoring discrete and continuous behaviors. The first assumption is especially relevant for transitions because they are often triggered by stepwise changes of the

![Figure 1. Consistency check with one uncertain parameter \( p \) and three subspaces \( q_1, q_2, \) and \( q_3 \). The residuals at the corner points of subspace \( q_1 \) are both negative, therefore, the model with the subspace \( q_1 \) is inconsistent with the measurement. In subspace \( q_2 \), the residuals at the corner points have different signs. Thus, \( q_2 \) is consistent. For the parameter range of subspace \( q_3 \) the monotonicity assumption is violated. In this case, checking the residuals’ signs at the corner points is not feasible.](image)
system input (e.g., caused by operator actions). Such transitions violate, therefore, the first assumption. The second assumption is a simple consequence of the integration of the given differential equation:

\[ x(t) = x_0 + \int_{t_0}^{t} \dot{x}(\tau) d\tau \]  

If the initial states \( x_0 \) are different at some corners in the subspace model, the state values \( x_i \) may not be monotonic (even if \( \dot{x} \) is monotonic). However, monotonicity is guaranteed after some time.

As discussed above discontinuous transitions may result in a non-monotonicity of the state values with regard to the parameters (for a limited period of time), which in turn leads to an incorrect consistency check. Thus, to maintain a correct (and conservative) monitoring technique we must extend the consistency check by a check for monotonicity. If the monotonicity is not guaranteed the consistency check is simply ignored and this subspace can not be refuted. At some time after the transition the subspace may become monotonic again and the consistency check can be applied again.

### 3.2 Checking for Monotonicity

The monotonicity of the state values for an individual subspace is checked by the following method.

We define a matrix \( B(t, x, p) \) with the elements

\[ b_{ij}(t, x, p) = \frac{\partial x_i(t, x, p)}{\partial p_j}, \]

where \( t \) is the time, \( x \) the state vector, and \( p \) the parameter vector with its elements \( p_j \). We also define the matrix \( C(t, x, p) \) with the elements

\[ c_{ij}(t, x, p) = \frac{\partial x_i(t, x, p)}{\partial p_j}. \]

The matrix \( C(t, x, p) \) is calculated by

\[ \frac{dC(t, x, p)}{dt} = A(t, x, p)C(t, x, p) + B(t, x, p), \]

where \( C(0, x_0, p) = 0 \) (the empty matrix), and the matrix \( A(t, x, p) \) is defined as

\[ a_{ij}(t, x, p) = \frac{\partial x_i(t, x, p)}{\partial x_j}. \]

The elements \( c_{ij}(t, x, p) \) give us the trend of the state value \( x_i(t, x, p) \) with regard to the parameter \( p_j \). This is exploited by our monotonicity check: The state values of a subspace model are monotonic, iff

\[ \text{sgn}(c_{ij,\text{min}}) = \text{sgn}(c_{ij,\text{max}}) \]

holds for all state values \( i = 1, \ldots, I \) and all directions of the uncertainty space \( j = 1, \ldots, K \). \( c_{ij,\text{min}} \) are the appropriate values of \( c_{ij}(t, x, p) \) at the corner \( \text{min} \), and \( c_{ij,\text{max}} \) are the values of \( c_{ij}(t, x, p) \) at the corner \( \text{max} \) of that subspace model (as described with Equations 5 and 6).

Figure 2 depicts the monotonicity check. In general, the information at the corner points is not sufficient to decide on monotonicity. However, assuming the monotonicity of the functions \( f \) and \( g \) with regard to the parameter, the monotonicity check becomes sufficient.

The calculation of the monotonicity check implies a numerical solution of the differential equation (Equation 12). However, since we use also a differential description of the system (\( f = \dot{x} \)), the monotonicity check does not significantly increase the computational load. Note that matrix \( A \) is constant for linear systems.

### 4 THE MONOTONICITY CHECK IN A REAL-WORLD SYSTEM

We now examine the monotonicity behavior on a “real” technical system which is comprised of three heating/cooling components mounted on a thermal conductive plate. A process control computer (B&R 2003) controls the three heating/cooling components. The measured samples as well as the control actions issued are transferred to the monitoring system via a RS 232 interface.

Our model which includes the three components with heating elements is given as

\[ \dot{T}_1 = \frac{1}{C_1} (q_{11} - L_1(T_1 - T_0) - L_{12}(T_1 - T_2)) \]

\[ \dot{T}_2 = \frac{1}{C_2} (q_{12} + L_{12}(T_1 - T_2) - L_2(T_2 - T_3) - L_{23}(T_2 - T_3)) \]

\[ \dot{T}_3 = \frac{1}{C_3} (q_{13} + L_{23}(T_2 - T_3) - L_3(T_3 - T_0)) \]

where \( T_i \) is the temperature of the three components, \( C_i \) is the mass of the components, \( q_i \) is the heat flow into the components, \( L_i \) the thermal conductivity between the component \( i \) and the environment, \( L_{ij} \) the thermal conductivity between the component \( i \) and \( j \), and \( T_0 \) the temperature of the environment. We can reduce the complexity of this model by exploiting the symmetric construction of the heating system \( (L_3 = L_1, L_{23} = L_{12}, C_3 = C_1) \) resulting in a total of five uncertain parameters.

The state vector is given as \( x = (T_1, T_2, T_3)^T \), the input vector as \( \mathbf{u} = (q_{11}, q_{12}, q_{13}, T_0)^T \), and the output vector as \( \mathbf{y} = (T_1 + n_1, T_2 + n_2, T_3)^T \), where \( n_i \) is the noise of each temperature sensor. The noise parameters are also included in the uncertainty space resulting in a total of eight uncertain parameters. Note that noise parameters are not dynamically partitioned into smaller intervals and they are not considered by the monotonicity check.

We have measured the input values with \( q_{u1} = 1.24 \) W and \( q_{u3} = 34.8 \) W (heating element is either turned off or turned on). With an initial refinement step, we get the parameter intervals as \( L_1 = [0.12, 0.13], L_2 = [0.15, 0.18], L_{12} = [0.62, 0.73], C_1 = [51, 54], C_3 = [61, 65]. \) The refinement step is performed in a single continuous behavior segment [12].

To observe the non-monotonic behavior in the system, we observe the system after a transition, and count the subspace models,
which are marked as non-monotonic. Over time, this gives us a picture, how the transition produce non-monotonicity in the state values. We choose the following scenario:

**Control state 1:** Heat $T_2$, if $T_2 < 70$. Then go to state 2.

**Control state 2:** Heat $T_2$, if $T_2 < 70$. If $t_{state.2} \geq 100 sec$, go to state 3.

**Control state 3:** Heat $T_2$, if $T_2 < 90$. If $t_{state.3} t_{state.4} \geq 100 sec \wedge T_2 \geq 90$, go to state 4.

**Control state 4:** Do not heat. If $T_2 \leq 50$, go to state 1.

Figure 3 plots the resulting measurements for this scenario. The heating flag $H_2$ (generated by the PCC) is used, to get a discrete change of an input. To implement the heating element characteristic, we assume an additional mass $C_h$ and a thermal conductivity $L_{th}$ between component 2 and the heating mass:

\[
    \dot{T}_{H2} = \frac{1}{C_h} \left(32.82 H_2 - L_{th}(T_{H2} - T_2)\right) \quad (15)
\]

\[
    q_{H2} = 1.24 + L_{th}(T_{H2} - T_2) \quad (16)
\]

To demonstrate the non-monotonic effect after a stepwise change of an input, we check the monotonicity of all subspace models, and count non-monotonic subspace models, i.e., which violate Equation 13. Figure 4 shows a part of the scenario, where the temperature of component 2 is hold at 90 degree (control state 3). For this plot, we have started with 128 subspace models, and no dynamic partitioning is introduced. Due to the discrete controller the heating is turned on and off several times. At each transition about 40 subspace models are non-monotonic. An interesting observation in this figure is, that the non-monotonic subspaces disappear quickly, if the heating flag is turned off only for a short time.

Figure 5 shows the number of the non-monotonic subspace models after control state 3. The peak here is about 30 subspace models. It shows, that non-monotonic subspace models are also existing for a “longer” time period (here about 400 seconds) after the last discrete change of an input.

Non-monotonic subspace models are not refuted, and, therefore, do not make any contribution to decrease the uncertainty space. Although the number of non-monotonic subspace models are quite high (about 50 percent of the current subspace models) for some times, it has not a significantly influence to the refutation. The reason is, however, that such peaks does not hold for long time, so the consistency check soon becomes valid again. At this example the number of consistent subspace models at the end of the scenario is about 20.

5 DISCUSSION

In this paper, we have presented a model-based monitoring approach based on uncertainty space partitioning. The fundamental assumption of this approach is the monotonicity of the state values with regard to the range of the parameters. In systems which exhibit both discrete and continuous behaviors the monotonicity can not be guaranteed only by the monotonicity of the vector functions. Thus, in order to apply our basic approach to monitor hybrid systems, we have introduced a monotonicity check for the state values.

Note the difference of monitoring based on pre-calculated envelopes with our approach. With pre-calculated envelopes, the envelopes remain constant over the complete monitoring process. In our approach, the envelopes may become smaller than the initial ones due to the refutation of inconsistent subspaces during monitoring. This results in an earlier detection of faults. However, there is a significant increase in the computational load of subspace partitioning.

Our approach is based on computing the envelopes of differential equations. For complex models, the overall runtime of our monitoring algorithm is dominated by solving the differential equations, especially when a high-precision method such as Runge-Kutta is used. The computational complexity of our algorithm for a single time-step can be estimated as

\[
    \mathcal{O}(M^2 K (\rho + \mu)) \quad (17)
\]

where $M$ is the number of partitions, $K$ is the number of uncertainty parameters, $\rho$ is the time of the Runge-Kutta algorithm, and $\mu$ is the time of the matrix multiplication according to Equation 11. The time $\rho$ strongly depends on the dynamic properties of the system, and for high dynamic systems, the assumption $\rho > \mu$ holds.
This approach can also be seen as system identification, because refuting subspace models reduces the uncertainty space, resulting in smaller bounding intervals on the parameters. Measurement noise can also be handled by introducing additional uncertainty parameters into the model.

However, this approach is in contrast to traditional system identification where the model space is specified by a parameterized differential equation. Identification selects numerical parameter values so that simulation of the model best matches the measurements. By using refutation instead of search our method is able to derive guaranteed bounds on the trajectories.

Model-based monitoring using uncertainty space partitioning is related to the interval identification algorithm of Schaich et al. [13]. In their approach the consistency check is only performed at the qualitative level. Thus, valuable detection time is lost, as long as the fault is only manifested in a quantitative value. Petridis and Kehagias [10] have also developed an algorithm with subspace partitioning. The partitioning is only performed in advance and the consistency check is based on probabilities depending on the noise in the system. Other work in monitoring [7, 9, 3] uses multiple models for fault detection. These models represent known faults of the supervised system. From the viewpoint of system identification, our approach is closely related to semi-quantitative system identification [8]. Identification of both approaches are grounded on the refutation of subspace models that are known to be inconsistent with the measurements. Semi-quantitative system identification performs refinement at the qualitative and interval level. Semi-quantitative system identification has also been applied to model-based monitoring [11]. Bonarini and Bontempi [4] have developed a quite similar approach to our consistency check. However, they have focused on uncertainty initial state values, which are given as intervals. Also related to our work is Armengol et al. [1, 2]. The simulation is based on modal interval arithmetic, which produces overbounded and underbounded envelopes of a technical system. To minimize the rate of false and missed alarms, the uncertainty space is only partitioned at critical measurements (which are between the underbounded and overbounded envelopes). In comparison to our approach, we simulate at each corner of the uncertainty space, which leads to exact envelopes (no false and missed alarms, according to observability) for linear systems.

Directions for future work include (i) the incorporation of (unknown) discontinuous transitions in our monitoring approach, (ii) further investigations on the monotonicity properties after a discontinuous transition, especially in the context of non-linear systems, and (iii) the improvement of the dynamic uncertainty space partitioning.

REFERENCES


Figure 4. First overview of the monotonicity of the technical system. Drawn are the measured $T_2$ with its envelopes, $H_2$ is the heating flag for the second component, and $NM$ is the number of the non-monotone subspace models. The discrete change of the input makes a directly effect to the monotonicity of the state values.