REDUCTION OF RANDOMLY-SWITCHING MEASUREMENT EQUATIONS TO THE DETERMINISTIC CASE

E. W. Kamen
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0250
E-mail: kamen@ee.gatech.edu

Abstract

It is shown that any discrete-valued randomly-switching parameter vector in the measurement equation of a system state model can be completely eliminated without requiring any information on the statistics of the parameter vector. Applications to state estimation are considered, including the problem of estimating the system state when there are missed detections or sensor failures.

1. Introduction

Consider the m-input p-output n-dimensional continuous-time system given by the state model

$$\dot{x}(t) = f(x(t),u(t),w(t),t) \tag{1}$$

$$y(t) = h(x(t),w(t),t) \tag{2}$$

where f and h are linear or nonlinear functions, t is the continuous-time variable, x(t) is the n-dimensional state vector, u(t) is the m-dimensional input vector, y(t) is the p-dimensional output vector, and for each value of t, w(t) is a parameter vector of jointly distributed random variables whose values range over a finite set. In the discrete-time case, the state model is given by

$$x(k+1) = f(x(k),u(k),w(k),k) \tag{3}$$

$$y(k) = h(x(k),w(k),k) \tag{4}$$

where k is the integer-valued time index.

We are also interested in the stochastic versions of (1)-(2) or (3)-(4) which include the addition of process noise and measurement noise. This will be considered later.

A system given by (1)-(2) or (3)-(4) can switch from one mode to another as a consequence of a change in value of the parameter vector w. Due to the presence of the random parameter vector w, systems given by (1)-(2) or (3)-(4) are referred to as switching systems, jump-parameter systems, or multi-modal systems. Such systems arise in many applications, including systems with multiple operating modes and systems with multiple fault states. In general, the random vector w may be in both the state dynamics and the measurement equation as is the case above, or it may be in only one or the other. In this paper we focus on the case when w is present in only the measurement equation, and thus, random switching occurs in only the measurement equation.

In general, the probability density functions of the random variables comprising the parameter vector w are not known a priori and may be difficult or impossible to estimate on-line. In addition, the density functions of the elements of w may vary as a function t (or k); that is, we may have nonstationarity. However, in many cases the values of the parameters in w are known a priori.

The great majority of existing approaches to control or state estimation in switching systems are based on some a priori assumptions on the statistics of the elements of w (see the survey [1]). This often includes the need to know the transition probabilities in Markov models that describe the evolution of the probabilities of the elements of w as a function of time. In this paper we present an approach that does not require any information on the statistics of w. The basic idea of the approach presented here is to eliminate the presence of w in the system model by deriving an algebraic characterization of the switching. A preliminary version of the key idea was first presented by the author at the 1994 Workshop on Fuzzy Intelligent Systems, held at Arlington, TX.

2. Algebraic Characterization of Randomly-Varying Parameters

Let $\mathbb{R}$ equal the set of real numbers. For each value of t $\in \mathbb{R}$, let w(t) be a discrete random variable that takes on either of two values:

$$w(t) = w_1 \text{ or } w_2$$

where w_1 and w_2 may be known or unknown real numbers. Nothing is assumed to be known regarding the statistics of w(t); that is, the probabilities

$$P[w(t) = w_1] \text{ and } P[w(t) = w_2]$$

are not known for any value of t.

The key observation is that w(t) satisfies the quadratic equation

$$[w(t) - w_1][w(t) - w_2] = 0$$

or

$$w^2(t) - (w_1 + w_2)w(t) + w_1 w_2 = 0 \tag{5}$$

From (5), we see that w(t) can be viewed as the zero of a second-degree polynomial. Hence, the randomness of w(t) is embedded in an algebraic framework; namely, a polynomial with two possible zeros. Note that if w_1 and w_2 are known, the coefficients of the polynomial representation of w(t) are known.

Now suppose that w(t) can take on the N possible values w_i, i = 1,2,...,N. Then w(t) satisfies the algebraic equation

$$[w(t) - w_1][w(t) - w_2] \cdots [w(t) - w_N] = 0 \tag{6}$$

The left-hand side of (6) can be written as a N-degree polynomial in w(t), and thus w(t) can be viewed as the zero of a N-degree polynomial. Again, note the coefficients of the polynomial are known if the w_i are known. Hence we have the following result.

Theorem 1: *Any random variable w(t) that takes on the values w_i, i = 1,2,...,N, satisfies a N-degree polynomial whose coefficients are known if the w_i are known.*

The algebraic relationship (6) can be used in the study of a system containing w(t) as a random parameter. In particular, it is possible to eliminate w(t) from the system representation by using (6). To illustrate this, suppose that a sensor provides a scalar measurement

$$y(t) = w(t)Cx(t) \tag{7}$$

of the state x(t) of a m-input n-dimensional linear time-invariant continuous-time system given by the state equation

$$\dot{x}(t) = Ax(t) + Bu(t)$$

Here w(t) is a random variable that takes on the value 0 or 1. The case w(t) = 0 corresponds to a missed detection or sensor failure. Nothing is known regarding the probability that w(t) = 0. Since w(t) = 0 or 1, it follows from (6) that w(t) satisfies the equation

$$w^2(t) - w(t) = 0$$

or

$$w^2(t) = w(t) \tag{8}$$

We can then use (8) to eliminate w(t) from the measurement equation (7): Squaring (7) gives

$$y^2(t) = w^2(t)[Cx(t)]^2 \tag{9}$$

but $w^2(t) = w(t)$, and thus (9) becomes

$$y^2(t) = w(t)[Cx(t)]^2$$

$$y^2(t) = [w(t)Cx(t)]Cx(t)$$

$$y^2(t) = y(t)Cx(t) \tag{10}$$

We now have a new measurement equation [namely, (10)] with w(t) completely eliminated. Note that the times at which jumps occur do not even enter into the new measurement equation.

Thus, the original switching system model has been reduced to a deterministic linear time-varying model given by the state model

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{11}$$

$$y^2(t) = y(t)Cx(t) \tag{12}$$

We can then proceed to generate controllers or observers for (11)-(12) using the standard deterministic theory.

It should be noted that for the measurement equation (7) with w(t) = 0 or 1, it is possible to estimate the value of w(t) directly from the value of y(t); namely, if y(t) = 0 we can set w(t) = 0, otherwise, we set w(t) = 1. Standard linear techniques can then be applied to the system model with w(t) evaluated in this way. Hence, the above construction is not really necessary in this case. But as seen in the next section, the approach generalizes to a large class of systems for which an obvious estimate of w(t) is not possible from the measurements y(t).

3. General Construction in the Single-Output Case

Consider a m-input single-output n-dimensional continuous-time system given by the state model

$$\dot{x}(t) = f(x(t),u(t),t) \tag{13}$$

$$y(t) = h(x(t),w(t),t) \tag{14}$$

where the random vector w(t) is a q-vector, and now w(t) appears only in the measurement equation (14). The vector w(t) takes on the q-vector values w_i , i = 1,2,...,N. Nothing is known regarding the probability that w(t) = w_i.

As a generalization of the construction given in Section 2, since y(t) is scalar and w(t) must take on one of the values w_i at any time t, we have the algebraic relationship

$$[y(t) - h(x(t),w_1,t)][y(t) - h(x(t),w_2,t)]\cdots[y(t) - h(x(t),w_N,t)] = 0$$

This can be written in the form

$$y^N(t) = g\big(y(t),x(t),w_1,w_2,\ldots,w_N,t\big) \tag{15}$$

for some function g which is nonlinear in general even if h is linear.

The measurement equation (15) is not a function of the random parameter vector w(t), and thus w(t) has been completely eliminated. The new measurement equation (15) is also independent of the jump times. We have therefore succeeded in reducing the original system model given by (13)-(14) to the deterministic case given by (13) and (15).

Summarizing the above construction, we have the following result.

Theorem 2: *Any randomly-switching system given by (13)-(14) with w(t) discrete valued can be reduced to a completely deterministic system given by (13) with the measurement equation (15).*

It is obvious that this result carries over to the discrete-time case also. More precisely, any discrete-time system given by the state model

$$x(k+1) = f(x(k),u(k),k)$$

$$y(k) = h(x(k),w(k),k)$$

where w(k) is discrete valued can be reduced to a deterministic system given by

$$x(k+1) = f(x(k),u(k),k)$$

$$y^N(k) = g\big(y(k),x(k),w_1,w_2,\ldots,w_N,k\big)$$

for some function g.

To illustrate this construction, consider the m-input single-output n-dimensional linear time-invariant continuous-time system given by the state model

$$\dot{x}(t) = Ax(t) + Bu(t)$$

$$y(t) = w(t)x(t)$$

where w(t) is a random n-element row vector that has the values

$$w(t) = w_1 \text{ or } w_2$$

where w_1 and w_2 are n-element row vectors. Nothing is known about the probability that w(t) = w_1 or w_2. Then

$$[y(t) - w_1x(t)][y(t) - w_2x(t)] = 0$$

or

$$y^2(t) = y(t)(w_1 + w_2)x(t) - w_1x(t)w_2x(t) \tag{16}$$

Equation (16) is a new measurement equation with w(t) completely eliminated, although the equation is time-varying and nonlinear in general. Note that if w_1 or w_2 = 0, then (16) reduces to the linear equation

$$y^2(t) = y(t)wx(t)$$

where w = w_1 or w_2.

The system can then be studied using the state model given by

$$\dot{x}(t) = Ax(t) + Bu(t)$$

$$y^2(t) = y(t)(w_1 + w_2)x(t) - w_1x(t)w_2x(t)$$

Since the new measurement equation is nonlinear, we need nonlinear techniques to proceed. For example, by using a linearized least-squares approach, we can construct an observer of the form

$$d\hat{x}(t)/dt = A\hat{x}(t) + K(t)[y^2(t) - y(t)(w_1 + w_2)\hat{x}(t) + w_1\hat{x}(t)w_2\hat{x}(t)]$$

where the observer gain K(t) is a n-element column vector. We shall not pursue the details in this paper.

When w(t) takes on the q-vector values w_i , i = 1,2,...,N, the new measurement equation is generated from the algebraic relationship

$$[y(t) - w_1x(t)][y(t) - w_2x(t)]\cdots[y(t) - w_Nx(t)] = 0$$

This yields a measurement equation that is multi-linear in x(t). Observers can be generated using numerical techniques for minimizing nonlinear cost functionals.

4. The Multi-Output Case

Again consider the measurement equation

$$y(t) = h(x(t),w(t),t)$$

where now the output y(t) is a p-vector with ith component denoted by $y_i(t)$, i = 1,2,...,p. Then the output equation can be written as p scalar equations:

$$y_i(t) = h_i(x(t),w(t),t),\ i = 1,2,\ldots,p$$

If the random parameter vector w(t) takes on the q-vector values w_i, i = 1,2,...,N, for each value of i = 1,2,...,p, we have the algebraic relationship:

$$[y_i(t) - h_i(x(t),w_1,t)][y_i(t) - h_i(x(t),w_2,t)]\cdots[y_i(t) - h_i(x(t),w_N,t)] = 0$$

Thus we are able to eliminate the random vector w(t) from the p measurement equations. However, in the multi-output case, there is more than one method for eliminating w(t). We illustrate this below by considering the **data association problem**.

Consider the linear time-invariant $(n_1 + n_2)$-dimensional system given by the state equations

$$\dot{x}(t) = Ax(t) + Bu(t)$$

$$\dot{z}(t) = Fz(t) + Gu(t)$$

where x(t) is n_1 dimensional and z(t) is n_2 dimensional. The output consists of the two scalar measurement equations

$$y(t) = Cx(t) \text{ or } Hz(t) \qquad (17)$$

$$\gamma(t) = Hz(t) \text{ or } Cx(t) \qquad (18)$$

From (17) and (18), we see that the association between the "data" given by y(t) and γ(t), and the states, x(t) and z(t), is not known. This is the data association problem. It arises in multiple target tracking (see [2]) and in problems involving multiple linear regressions.

We can express (17)-(18) in terms of the scalar-valued random switching parameter w(t) with w(t) = 0 or 1. This gives

$$y(t) = w(t)Cx(t) + [1-w(t)]Hz(t) \qquad (19)$$

$$\gamma(t) = [1-w(t)]Cx(t) + w(t)Hz(t) \qquad (20)$$

To eliminate w(t) from (19) and (20), we can apply the procedure given above to generate two new measurement equations given in terms of $y^2(t)$ and $\gamma^2(t)$. Instead of doing this, we shall construct two new measurement equations by forming the sum and the product of y(t) and γ(t): From (17)-(18),

$$y(t) + \gamma(t) = Cx(t) + Hz(t) \qquad (21)$$

$$y(t)\gamma(t) = Cx(t)Hz(t) = x^T(t)C^T Hz(t) \qquad (22)$$

where superscript "T" denotes the transpose operation.

Hence, the randomly-switching parameter w(t) does not appear in the new measurement equations (21)-(22). Observers and controllers for the given system can then be generated using these measurement equations; however, since (22) is nonlinear in the state, nonlinear techniques must be applied.

The choice of taking the sum and product of (17) and (18) in forming the new measurement equations appears to be a much better one in this case than forming new measurement equations in terms of $y^2(t)$ and $\gamma^2(t)$. Part of the reason for this is that one of the two new measurement equations, namely (21), is linear, whereas forming $y^2(t)$ and $\gamma^2(t)$ results in two nonlinear equations. It would be nice if w(t) could be eliminated by forming two new linear measurement equations from (17) and (18), but this does not appear to be possible. In other words, there is a "cost" in eliminating the randomness due to w(t), and that cost is having to deal with nonlinear equations.

In the general multi-output case, it is an interesting question as to whether or not there is an "optimal way" to eliminate w(t) from the measurement equation. It appears that there is no general procedure; rather, the best method for eliminating w(t) depends on how w(t) enters into the measurement equations. It also appears that vector or tensor operations may be appropriate for eliminating w(t) in some situations. For example, suppose that the system under consideration has two outputs:

$$y(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix}$$

and

$$y(t) = h(x(t),w(t),t)$$

with $w(t) = 0$ or 1. Then we can eliminate $w(t)$ by forming the relationship

$$[y(t)-h(x(t),0,t)][y^T(t) - h(x(t),1,t)^T] = 0 \qquad (23)$$

However, there are four measurement equations contained in (23), and thus the use of a vector operation has increased the number of measurement equations from two to four. It's not clear if the increase in the number of measurement equations leads to a better result.

5. The Stochastic Case

Again consider the single-output case, but now suppose that there is additive measurement noise $v(t)$ so that the measurement equation becomes

$$y(t) = h(x(t),w(t),t) + v(t)$$

where $v(t)$ is zero-mean white noise with $E[v^2(t)] = \sigma^2$ and $w(t)$ takes on the q-vector values w_i, $i = 1,2,...,N$. To eliminate $w(t)$, we form the relationship

$$[y(t)-h(x(t),w_1,t)-v(t)][y(t)-h(x(t),w_2,t)-v(t)]\cdots [y(t) - h(x(t),w_N,t)-v(t)] = 0$$

which can be written in the form

$$y^N(t) = g\big(y(t),x(t),w_1,w_2,...,w_N,t\big) + \gamma(t)$$

where $\gamma(t)$ is a noise term that is a function of $v(t)$, $x(t)$, $y(t)$, and w_i, $i = 1,2,...,N$. Thus we have a new measurement equation with $w(t)$ completely eliminated. We illustrate this construction below in the discrete-time case.

Consider the m-input single-output linear time-invariant n-dimensional discrete-time system given by the state model

$$x(k+1) = Ax(k) + Bu(k) \qquad (24)$$

$$y(k) = w(k)Cx(k) + v(k) \qquad (25)$$

where $w(k) = 0$ or 1 and $v(k)$ is zero-mean white noise with $E[v^2(k)] = \sigma^2$. The case $w(k) = 0$ corresponds to a missed detection or sensor failure.

To eliminate $w(k)$ from (25), we form the relationship

$$[y(k) - v(k)][y(k) - Cx(k) - v(k)] = 0$$

which can be written in the form

$$y^2(k) = y(k)Cx(k) + \gamma(k) \qquad (26)$$

where

$$\gamma(k) = [2y(k) - Cx(k)]v(k) - v^2(k) \qquad (27)$$

Inserting (25) into (27) gives

$$\gamma(k) = [2w(k)-1]Cx(k)v(k) + v^2(k)$$

$$\gamma(k) = \pm Cx(k)v(k) + v^2(k) \qquad (28)$$

Note that $\gamma(k)$ contains a signal × noise term. Taking the expectation of both sides of (28), we see that

$$E[\gamma(k)] = E[v^2(k)] = \sigma^2$$

and thus the noise term $\gamma(k)$ is not zero mean. To have zero-mean noise, we need to subtract σ^2 from both sides of the measurement equation (26), which gives

$$y^2(k) - \sigma^2 = y(k)Cx(k) + \gamma(k) - \sigma^2$$

Defining the noise term $\bar{\gamma}(k) = \gamma(k) - \sigma^2$, we have the measurement equation

$$y^2(k) - \sigma^2 = y(k)Cx(k) + \bar{\gamma}(k) \qquad (29)$$

Since $v(k)$ is white noise and is independent of $x(k)$, it follows that $\bar{\gamma}(k)$ is white, and thus the "measurement noise" in (29) is white. The variance of $\bar{\gamma}(k)$ can be approximated by σ^4.

To estimate the state $x(k)$, we can then apply the standard Kalman filter approach to the state equation (24) with the measurement equation (29). Note that observability of the system given by (24) and (29) depends on the measurements $y(k)$, and thus the performance of the filter depends on having "sufficiently rich" observations $y(k)$ (i.e., giving observability).

As an example showing the filtering application, suppose that we have noisy measurements $y(k)$ of an unknown real constant c with $y(k)$ given by

$$y(k) = w(k)c + v(k)$$

where $w(k) = 0$ or 1and $v(k)$ is zero-mean Gaussain white noise with variance $\sigma^2 = 10$. For the case when $c = 10$ and $w(k)c$ is given by the plot in Figure 1, we generated the realization of the measurements shown in Figure 2. The objective is to estimate the value of c from the measurements shown in Figure 2 with **no a priori information** on the statistics of $w(k)$. To accomplish this, we first set up the system model, which in this case is given by

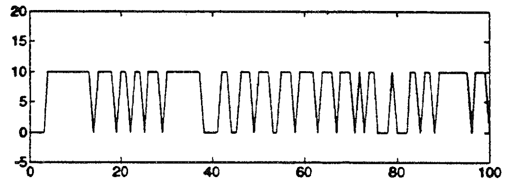

Figure 1. Plot of w(k)c versus time.

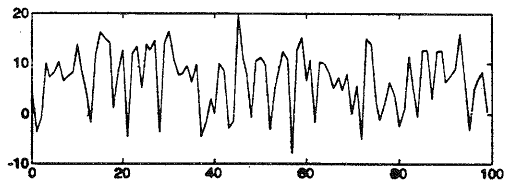

Figure 2. Plot of the measurements.

$$x(k+1) = x(k) \tag{30}$$

$$y^2(k) - 10 = y(k)x(k) + \bar{\gamma}(k) \tag{31}$$

where

$$\bar{\gamma}(k) = \pm x(k)v(k) + v^2(k) - 10$$

Applying the Kalman filter approach to the system given by (30)-(31) results in the recursive estimate

$$\hat{x}(k+1) = \hat{x}(k) + L(k)[y^2(k) - 10 - y(k)\hat{x}(k)] \tag{32}$$

where the filter gain L(k) is given by

$$L(k) = \frac{P(k)y(k)}{P(k)y^2(k) + 100} \tag{33}$$

and P(k) satisfies the scalar Riccati equation

$$P(k+1) = P(k) - \frac{P^2(k)y^2(k)}{P(k)y^2(k) + 100} \tag{34}$$

The filter given by (32)-(34) was run with initializations $\hat{x}(0) = 0$ and $P(0) = (10)I$. The resulting estimates are plotted in Figure 3. Note the convergence of the estimates to the correct value of 10.

We can study the performance of the filter by considering the estimation error e(k) defined by

$$e(k) = x(k) - \hat{x}(k)$$

Then

$$e(k+1) = x(k+1) - \hat{x}(k+1)$$

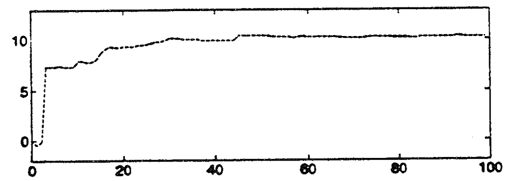

Figure 3. Plot of estimates

and using (30) and (32) gives

$$e(k+1) = [1 - L(k)y(k)]e(k) + L(k)\bar{\gamma}(k)$$

It turns out that

$$|1 - L(k)y(k)| < 1 \text{ if } y(k) \neq 0 \tag{35}$$

and thus the effect of the initial estimation error e(0) is diminished as time increases, which explains the result shown in Figure 3. The proof of (35) follows easily by solving (34) for P(k) and inserting the result into (33) and then evaluating 1 - L(k)y(k). The details are left to the reader.

6. Conclusions

As shown in this paper, a discrete-valued randomly-switching parameter vector in the measurement equation can be completely eliminated by forming new measurements from the given measurements. As a result, conventional techniques can be applied to switching systems without the need for any a priori information on the statistics of the switching parameters. Extension of the approach to switching parameters in the state dynamics is currently under investigation.

7. Acknowledgment

The author wishes to thank Louis Bellaire for generating the estimates for the constant-signal example considered in Section 5. This work was supported in part by the BMDO/IST and managed by the Office of Naval Research under Contract No. N00014-92-J-1745.

References

[1] X. R. Li, "Hybrid Estimation techniques," *Control & Dynamic Systems*, vol. 76, pp. 213-287, 1996.

[2] E. W. Kamen and C. R. Sastry, "Multiple Target Tracking Using Products of Position Measurements," *IEEE Trans. Aerospace & Electronic Systems*, vol. 29, pp. 476-493, 1993.