Part II. Fault Isolation Using Analytic Redundancy

5. Kalman Filtering for Fault Isolation
6. Fault Isolation Using Analytic Redundancy
7. Multiple Model Adaptive Estimation (MMAE)

We provide three successive generalizations of the Bayesian EWMA of section 3 for fault isolation. Section 5 describes a Kalman filter providing an on-going recalibration of metrology, distinguishing and estimating drift in the intercept from that of the slope. Section 6 generalizes this for fault isolation in a nonlinear physical system (an automotive air intake system). When trying to isolate too many faults, this methodology is found to encounter problems of numerical stability, due perhaps to lack of regression leverage for estimation of the components of the state vector. (Kalman filtering can be described as regression analysis performed one observation at a time, gracefully discounting the past.) These problems can be overcome in at least some cases by running different Kalman filters in parallel, as discussed in section 7. Concluding remarks for this report appear in part III.

5. KALMAN FILTERING FOR FAULT ISOLATION

In this section, we consider Bayesian sequential updating for fault isolation. The development follows the procedure outlined in Figure 3.1, generalizing the EWMA of section 3 to estimate multiple parameters from a univariate response. This is important for monitoring because with complex systems such as an automobile, a manufacturing facility, or a human body, many things can go wrong. It is therefore desirable not merely to detect a fault but to isolate it as well: What disease do we need to treat? Which component of the automobile should be replaced? Which piece of equipment in the production line should be adjusted and how?

In some cases, fault isolation is more important than fault detection. Suppose, for example, that an on-board diagnostic (OBD) tells a car owner to fix the emission controls. Suppose further that the owner spends $200 to fix a problem that s/he can neither smell, feel, nor see, apart from an OBD light on the instrument panel. In this context, what happens if the OBD light comes back on shortly after a repair? Obviously, it depends on the customer, but some customers may simply ignore the light, refusing to attempt another repair when a substantive expense failed to fix an apparent problem that did not affect drivability. In such a context, fault detection with improper fault isolation may be *worse than useless*, because it encourages non-compliance and contempt for the law. Since OBDs for emission controls are legally required on all new automobiles sold today in the US, Canada and Europe, this issue is a part of daily life in much of the developed world today.

Something similar can happen in manufacturing. In fabrication of integrated circuits, it is a waste of time to tell engineers and operators that something in the process

is responsible for a 1 percent yield loss, unless you can isolate, for example, a particular piece of equipment that needs maintenance of a certain type. However, a 1 percent improvement in yield can be worth millions of dollars per year in many wafer fabs today. The money involved in these applications can justify substantial effort to design and implement sophisticated monitoring systems. In some cases, fault isolation can be performed effectively by secondary diagnostic procedures after a fault has been detected; we discuss here incorporating fault isolation with OBDs.

In this section, we assume that the condition of the plant evolves as a multidimensional random walk as follows:

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \mathbf{w}_t, \text{ where } \mathbf{w}_t \sim N_p(0, \mathbf{W}). \tag{5.1}$$

The information on this state is obtained from a noisy univariate measurement as follows:

$$y_t = \mathbf{H}_t \mathbf{x}_t + v_t, \text{ where } v_t \sim N(0, \sigma_v^2), \tag{5.2}$$

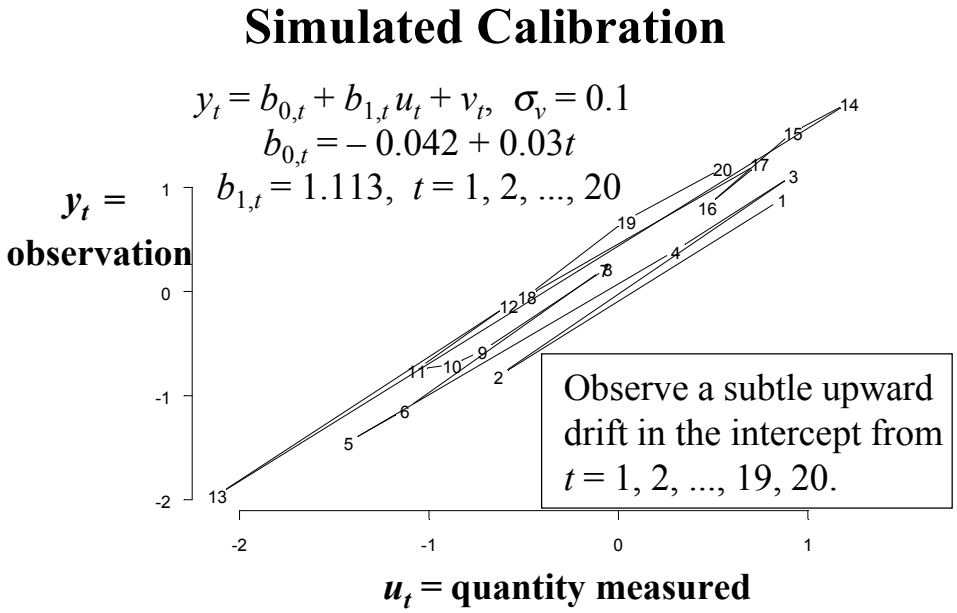
where $\mathbf{H}_t$ can in general be any p-dimensional row vector.

To fix ideas, consider the problem of continually reestimating calibration or sensitivity parameters in a measurement process. Figure 5.1 depicts a sequence of pairs of observations (u_t, y_t), where $u_1, u_2, ..., u_{20}$ are simulated samples whose value was established by concurrent measurement by a more costly reference test, and $y_1, y_2, ..., y_{20}$ were obtained using a production test. The figure suggests a linear relationship between u_t and y_t, as

$$y_t = b_{0,t} + b_{1,t}\, u_t + v_t, \text{ where } v_t \sim N(0, \sigma_v^2). \tag{5.3}$$

However, this linear relationship appears not to be constant but rather seems to be drifting gradually upwards over time.

Figure 5.1. Estimating Sensitivity Parameters in a Measurement Process

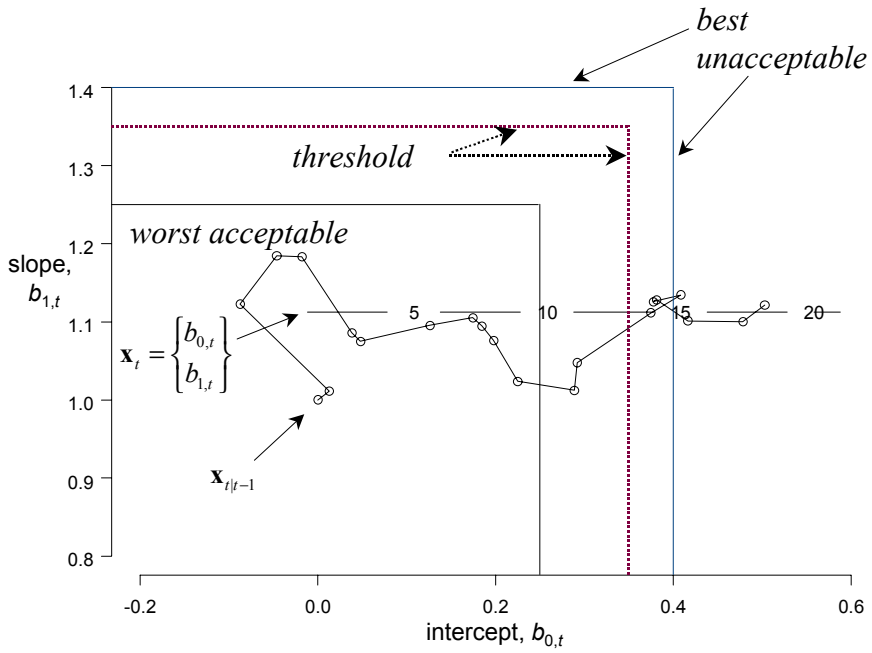


The state space for the sensitivity parameters $\mathbf{x}_t = (b_{0,t}, b_{1,t})'$ is depicted in Figure 5.2, with the simulated true but unknown condition of the plant marked as a straight horizontal line labeled $\mathbf{x}_t$ drifting to the right. The Bayesian sequential updating / Kalman filtering procedure described later in this section was applied to the observations in Figure 5.1 to determine the Bayesian prior mean at each point in time, $\mathbf{x}_{t|t-1}$ plotted in Figure 5.2; we see that this prior mean wanders around but ultimately follows the simulated true condition $\mathbf{x}_t$ on its journey to the right of Figure 5.2.

We suppose that previous experience with this hypothetical production test suggests erratic performance with occasional unacceptable nonlinearity when $\mathbf{x}_t$ had either $|b_{0,t}| \geq 0.4$ or $|b_{1,t} - 1| \geq 0.4$, but that the system has performed acceptably when $\mathbf{x}_t$ has been in the region $|b_{0,t}| \leq 0.25$ and $|b_{1,t} - 1| \leq 0.25$. In the "undefined zone" with $0.25 < |b_{0,t}| < 0.4$ or $0.25 < |b_{1,t} - 1| < 0.4$, the plant performance has not always been acceptable

but also was rarely bad enough to justify the expense of an immediate repair. In this context, we call the lines $|b_{0,t}| = 0.25$ and $|b_{1,t} - 1| = 0.25$ "worst acceptable", while $|b_{0,t}| = 0.4$ and $|b_{1,t} - 1| = 0.4$ are called "best unacceptable". To balance the risk of false alarm with that of excessive delay, we have selected the lines $|b_{0,t}| = 0.35$ and $|b_{1,t} - 1| = 0.35$ as detection thresholds: When $\mathbf{x}_{t|t-1}$ crosses either of these lines, a repair is ordered.

Figure 5.2. Detecting and Isolating a Measurement Sensitivity Fault



This procedure offers several advantages over traditional calibration. Many instruments and sensors experience gradual calibration drift, and a continual recalibration as described here can improve their performance. Indeed, this procedure can out-perform traditional calibration procedures that adjust the sensitivity parameters using only a few test samples collected at one time. This procedure can operate on, say, one test sample per week, combining the results of all previous test samples in an appropriate way, discounting older history as calibration drift makes it less relevant.

Moreover, the methodology described here can be applied in other, similar contexts and can be generalized to plants with analytical redundancy where the laws of physics can be used to check each of several sensors against the others. Without using analytic redundancy, engineers protect against faulty sensors by installing multiple copies of the same sensor. This drives up the per-unit cost and the total weight. The latter can be a substantial concern in aircraft design and may be unacceptable in military aircraft where a bullet that disconnects a sensor in a wing may simultaneously disconnect back-up sensors.

In Figure 5.2, we see that the Bayesian prior estimate of the condition of the plant $\mathbf{x}_{t|t-1}$ for t = 1, 2, ..., 21, leaves the "acceptable" region after observation $t-1$ = 12, just before observation t = 13. It enters the "unacceptable" region before observation 16 only to return to the "undefined" zone before observation 17; it becomes "unacceptable" again just before observation 19. This follows approximately the simulated unknown condition $\mathbf{x}_t$ that leaves the "acceptable" region with observation 10 and enters the "unacceptable" region for observation 15. When $\mathbf{x}_{t|t-1}$ crosses the threshold (dotted lines) on Figure 5.2 before observation 15, a malfunction is declared and the production test equipment is removed for maintenance. By examining Figure 5.2, the standards department would know the following: (*a*) The equipment suffers from an offset problem of magnitude approximately 0.47, and (*b*) it is neither hypersensitive nor dull. In some applications, this information can reduce the cost of the repair.

In many applications, this kind of continual reestimation of sensitivity parameters can provide higher quality measurements from cheaper sensors than would otherwise be

possible, and the fault-isolation information provided by the estimated $\mathbf{x}_{t|t-1}$ can substantially reduce the cost of maintenance.

As explained below, these estimates assume that the plant evolves following a simple random walk, per (5.1). A minor generalization of the present theory could track a deterministic drift much better, but would not be able to follow as well a random walk.

5.1. Bayesian Updating for Fault Isolation

Bayesian sequential updating begins by assuming that our knowledge of the state of the plant at time t can be summarized in a normal prior:

$$(\mathbf{x}_t \mid D_{t-1}) \sim N(\mathbf{x}_{t|t-1}, \mathbf{\Sigma}_{t|t-1}), \tag{5.4}$$

where $D_{t-1} = \{ y_{t-1}, y_{t-2}, \ldots y_1, \mathbf{x}_{1|0}, \mathbf{\Sigma}_{1|0}\}$. For $t = 1$, the distribution of $(x_t \mid D_{t-1})$ is just $N(\mathbf{x}_{1|0}, \mathbf{\Sigma}_{1|0})$, the distribution of the condition of the plant at first use; in many applications, this can be estimated from an appropriate external reference population. For $t > 1$, this is the output of Step 2 in the iteration of Figure 3.3.

This is processed as described in step 1 of Figure 3.1 into the posterior $N(\mathbf{x}_{t|t}, \mathbf{\Sigma}_{t|t})$ using substeps 1.1. preparing and 1.2. updating, as we now explain.

Step 1.1. Preparing. This is further divided into three substeps: (1.1a) predictive distribution, (1.1b) posterior variance, and (1.1c) Kalman gain.

Step 1.1a. Predictive Distribution. We combine the prior with the observation process (5.2) and integrate out the unknowable $\mathbf{x}_t$ to get the predictive distribution as follows:

$$(y_t \mid D_{t-1}) \sim N(f_t, \sigma^2_{y|t-1}),$$

where

$$f_t = \mathbf{H}_t\, \mathbf{x}_{t|t-1}, \tag{5.5}$$

since the expected value of the sum in (5.2) is the sum of the expectations, and

$$\sigma^2_{y|t-1} = \mathbf{H}_t \boldsymbol{\Sigma}_{t|t-1} \mathbf{H}_t' + \sigma_v^2,$$

using

$$\begin{aligned}
\text{var}(\mathbf{H}_t\, \mathbf{x}_{t|t-1} \mid D_{t-1}) &= E\left\{ [\mathbf{H}_t \mathbf{x}_t - E(\mathbf{H}_t \mathbf{x}_t \mid D_{t-1})][\mathbf{H}_t \mathbf{x}_t - E(\mathbf{H}_t \mathbf{x}_t \mid D_{t-1})]' \,\middle|\, D_{t-1} \right\} \\
&= \mathbf{H}_t E\left\{ [\mathbf{x}_t - E(\mathbf{x}_t \mid D_{t-1})][\mathbf{x}_t - E(\mathbf{x}_t \mid D_{t-1})]' \,\middle|\, D_{t-1} \right\} \mathbf{H}_t' \\
&= \mathbf{H}_t E\left\{ [\mathbf{x}_t - \mathbf{x}_{t|t-1}][\mathbf{x}_t - \mathbf{x}_{t|t-1}]' \,\middle|\, D_{t-1} \right\} \mathbf{H}_t' = \mathbf{H}_t \boldsymbol{\Sigma}_{t|t-1} \mathbf{H}_t'.
\end{aligned}$$

Step 1.1b. Posterior Variance. To derive the posterior distribution, we work with the probability density functions. Basic concepts of conditional probabilities give us

$$\begin{aligned}
p(y_t, \mathbf{x}_t \mid D_{t-1}) &= p(y_t \mid D_{t-1}) p(\mathbf{x}_t \mid y_t, D_{t-1}) = p(y_t \mid \mathbf{x}_t)\ p(\mathbf{x}_t \mid D_{t-1}) \\
(\text{joint}) &= (\text{predictive}) \times (\text{posterior}) = (\text{observation}) \times (\text{prior}),
\end{aligned}$$

where $p(\,.\mid .\,)$ = probability density. But $D_t = \{y_t, D_{t-1}\}$, so $p(\mathbf{x}_t \mid y_t, D_{t-1}) = p(\mathbf{x}_t \mid D_t)$.

Taking logarithms, letting $l(.\mid .) = \ln[p(.\mid .)]$, we get the following:

$$\begin{aligned}
l(y_t \mid D_{t-1}) + l(\mathbf{x}_t \mid D_t) &= l(y_t \mid \mathbf{x}_t) + l(\mathbf{x}_t \mid D_{t-1}) \\
(\text{predictive}) + (\text{posterior}) &= (\text{observation}) + (\text{prior}).
\end{aligned} \tag{5.6}$$

We will use first and second derivatives of this expression relative to $\mathbf{x}_t$:

$$\frac{\partial l(\mathbf{x}_t \mid D_t)}{\partial \mathbf{x}_t} = \frac{\partial l(\mathbf{y}_t \mid \mathbf{x}_t)}{\partial \mathbf{x}_t} + \frac{\partial l(\mathbf{x}_t \mid D_{t-1})}{\partial \mathbf{x}_t}, \tag{5.7}$$

and

$$\frac{\partial^2 l(\mathbf{x}_t \mid D_t)}{\partial \mathbf{x}_t \partial \mathbf{x}_t'} = \frac{\partial^2 l(\mathbf{y}_t \mid \mathbf{x}_t)}{\partial \mathbf{x}_t \partial \mathbf{x}_t'} + \frac{\partial^2 l(\mathbf{x}_t \mid D_{t-1})}{\partial \mathbf{x}_t \partial \mathbf{x}_t'}. \tag{5.8}$$

Recall that the *observed information* is defined as the negative of the second derivative of the log(likelihood), and the *Fisher information* is the expected value of the

observed information. Let $\mathbf{J}(\,.\,|\,.\,)$ denote the observed information matrices = the negative of the matrices of second partials in (5.8). Then (5.8) becomes

$$\begin{array}{ccccc} \mathbf{J}(\mathbf{x}_t \mid D_t) & = & \mathbf{J}(\mathbf{y}_t \mid \mathbf{x}_t) & + & \mathbf{J}(\mathbf{x}_t \mid D_{t-1}) \\ \begin{pmatrix}\text{posterior}\\ \text{information}\end{pmatrix} & = & \begin{pmatrix}\text{information from}\\ \text{observation(s)}\end{pmatrix} & + & \begin{pmatrix}\text{prior}\\ \text{information}\end{pmatrix}. \end{array} \tag{5.9}$$

We find this result quite useful if only to help us remember and understand Kalman filtering / Bayesian sequential updating. In the normal case, this information is also the precision parameter(s), being the inverse of the variance (or covariance matrix). For a discussion of this result beyond the immediate context, see the appendix in section 6 below.

We assumed in (5.4) that the prior was normal and in (5.2) that the observation was normal. Therefore,

$$l(\mathbf{x}_t \mid D_{t-1}) = c_1 - \frac{1}{2}\left(\mathbf{x}_t - \mathbf{x}_{t|t-1}\right)' \boldsymbol{\Sigma}_{t|t-1}^{-1}\left(\mathbf{x}_t - \mathbf{x}_{t|t-1}\right),$$

and (5.10)

$$l(y_t \mid \mathbf{x}_t) = c_y - \frac{1}{2\sigma_v^2}\left(y_t - \mathbf{H}_t\mathbf{x}_t\right)^2 .$$

where c_1 and c_y are appropriate constants. By examining (5.6), we see that the posterior $l(\mathbf{x}_t \mid D_t)$ must be a constant minus a paraboloid in $\mathbf{x}_t$, since the predictive distribution does not contain $\mathbf{x}_t$. This makes the posterior also a paraboloid in $\mathbf{x}_t$ and therefore normal as well. We therefore write

$$l(\mathbf{x}_t \mid D_t) = c - \frac{1}{2}\left(\mathbf{x}_t - \mathbf{x}_{t|t}\right)' \boldsymbol{\Sigma}_{t|t}^{-1}\left(\mathbf{x}_t - \mathbf{x}_{t|t}\right), \tag{5.11}$$

with appropriate choices for c, $\mathbf{x}_{t|t}$ and $\boldsymbol{\Sigma}_{t|t}^{-1}$; we will use (5.7) and (5.8) to determine $\mathbf{x}_{t|t}$ and $\boldsymbol{\Sigma}_{t|t}^{-1}$. From (5.10) and (5.11), we get the following:

$$\frac{\partial l(\mathbf{x}_t \mid D_{t-1})}{\partial \mathbf{x}_t} = \left[-\boldsymbol{\Sigma}_{t|t-1}^{-1}(\mathbf{x}_t - \mathbf{x}_{t|t-1})\right],$$

$$\frac{\partial l(\mathbf{y}_t \mid \mathbf{x}_t)}{\partial \mathbf{x}_t} = \frac{1}{\sigma_v^2}\mathbf{H}_t'(y_t - \mathbf{H}_t\mathbf{x}_t), \tag{5.12}$$

$$\frac{\partial l(\mathbf{x}_t \mid D_t)}{\partial \mathbf{x}_t} = \left[-\boldsymbol{\Sigma}_{t|t}^{-1}(\mathbf{x}_t - \mathbf{x}_{t|t})\right],$$

and

$$\mathbf{J}(\mathbf{x}_t \mid D_{t-1}) = \left[-\frac{\partial^2 l(\mathbf{x}_t \mid D_{t-1})}{\partial \mathbf{x}_t \partial \mathbf{x}_t'}\right] = \boldsymbol{\Sigma}_{t|t-1}^{-1},$$

$$\mathbf{J}(\mathbf{y}_t \mid \mathbf{x}_t) = \sigma_v^{-2}\mathbf{H}_t'\mathbf{H}_t, \tag{5.13}$$

$$\mathbf{J}(\mathbf{x}_t \mid D_t) = \left[-\frac{\partial^2 l(\mathbf{x}_t \mid D_t)}{\partial \mathbf{x}_t \partial \mathbf{x}_t'}\right] = \boldsymbol{\Sigma}_{t|t}^{-1}.$$

Substituting (5.13) into (5.9), we get

$$\boldsymbol{\Sigma}_{t|t}^{-1} = \boldsymbol{\Sigma}_{t|t-1}^{-1} + \sigma_v^{-2}\mathbf{H}_t'\mathbf{H}_t. \tag{5.14}$$

In words, this says that the posterior (Fisher) information is the prior information plus the information about $\mathbf{x}_t$ contained in the data.

For further discussion of this, see West and Harrison (1999, pp. 639-640), Press (1972, p. 77), the appendix in section 6 below, or in a univariate context, DeGroot (1970, p. 167).

If we similarly substitute (5.12) into (5.7), we get

$$\left[-\boldsymbol{\Sigma}_{t|t}^{-1}(\mathbf{x}_t - \mathbf{x}_{t|t})\right] = \sigma_v^{-2}\mathbf{H}_t'(y_t - \mathbf{H}_t\mathbf{x}_t) - \boldsymbol{\Sigma}_{t|t-1}^{-1}(\mathbf{x}_t - \mathbf{x}_{t|t-1}).$$

This holds for all $\mathbf{x}_t$ and in particular for $\mathbf{x}_t = 0$:

$$\boldsymbol{\Sigma}_{t|t}^{-1}\mathbf{x}_{t|t} = \sigma_v^{-2}\mathbf{H}_t'y_t + \boldsymbol{\Sigma}_{t|t-1}^{-1}\mathbf{x}_{t|t-1}.$$

Assuming $\boldsymbol{\Sigma}_{t|t}^{-1}$ is nonsingular, we get from this

$$\mathbf{x}_{t|t} = \boldsymbol{\Sigma}_{t|t}\left\{\boldsymbol{\Sigma}_{t|t-1}^{-1}\mathbf{x}_{t|t-1} + \mathbf{H}_t'\sigma_v^{-2}y_t\right\}, \tag{5.15}$$

Step 1.1c. Kalman Gain. The weight on the last observation y_t in (5.15) is called the Kalman gain and will be denoted as follows:

$$\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t}\mathbf{H}_t'\sigma_v^{-2}. \tag{5.16}$$

From (5.14), we see that

$$\boldsymbol{\Sigma}_{t|t-1}^{-1} = \boldsymbol{\Sigma}_{t|t}^{-1} - \mathbf{H}_t'\sigma_v^{-2}\mathbf{H}_t'.$$

We substitute these last two expressions into (5.15) to get

$$\begin{aligned}\mathbf{x}_{t|t} &= \boldsymbol{\Sigma}_{t|t}\left\{\left(\boldsymbol{\Sigma}_{t|t}^{-1} - \mathbf{H}_t'\sigma_v^{-2}\mathbf{H}_t\right)\mathbf{x}_{t|t-1} + \mathbf{H}_t'\sigma_v^{-2}y_t\right\} \\ &= \mathbf{x}_{t|t-1} + \mathbf{K}_t\left(y_t - \mathbf{H}_t x_{t|t-1}\right) \\ &= \mathbf{x}_{t|t-1} + \mathbf{K}_t\left(y_t - f_t\right),\end{aligned} \tag{5.17}$$

using (5.5).

For plants with constant observation and transition variances, all the computations of substep 1.1 can be done offline except for the mean of the predictive distribution. With or without those offline computations, if these preparations are done prior to the arrival of the latest observation, y_t, it can shorten slightly the time required to update our knowledge of the state of the plant. This may be useful with certain real-time controllers.

Step 1.2. Updating. In "updating", we compute the prediction error and use that to update the "posterior mean", our point estimate of the state of the plant.

Step 1.2a. Observed Residual. When the observation y_t arrives, we compute the observed residual as,

$$e_t = y_t - f_t. \tag{5.18}$$

Step 1.2b. Posterior Mean. With observed residual in hand, we multiply it by the Kalman gain and add the product to the prior mean to obtain the posterior mean per (5.17), as follows:

$$\mathbf{x}_{t|t} = \mathbf{x}_{t|t-1} + \mathbf{K}_t e_t . \tag{5.19}$$

This completes step 1, observation, in Bayesian sequential updating as outlined in Figure 3.3. Next, we permit the plant to transition in preparation for the next observation, per step 2.

Step 2. Prior for the Next Observation. Given the posterior mean and variance from step 1, we can easily compute using (5.1) the prior mean and variance for the next observation, as follows:

Step 2.1. Prior Mean.

$$\mathbf{x}_{t+1|t} = \mathbf{x}_{t|t}, \tag{5.20}$$

and

Step 2.2. Prior Variance.

$$\mathbf{\Sigma}_{t+1|t} = \mathbf{\Sigma}_{t|t} + \mathbf{W}. \tag{5.21}$$

This completes step 2. The resulting prior distribution at one point in time $N_p(\mathbf{x}_{t+1|t},\ \mathbf{\Sigma}_{t+1|t})$ becomes, when t is incremented to $t + 1$, an input for step 1.1 as $N_p(\mathbf{x}_{t|t-1},\ \mathbf{\Sigma}_{t|t-1})$. In this way, observations are processed sequentially as they arrive. If the model (5.1) - (5.2) is correct, then the prior $N_p(\mathbf{x}_{t+1|t},\ \mathbf{\Sigma}_{t+1|t})$ summarizes all the information in $D_t = \{y_t, y_{t-1}, ..., y_1, \mathbf{x}_{1|0},\ \mathbf{\Sigma}_{1|0}\}$ about the state of the plant at time $t+1$.

We now apply this theory to the example shown in Figure 5.1.

5.2. Example: Drift in Metrology Sensitivity Parameters

The theory of section 5.1 was applied to a simulated metrology problem. The input data are presented in Figure 5.1. A few of these observations and corresponding computations appear in Table 5.1, and the Bayesian prior means are plotted in Figure 5.2.

Table 5.1. Detecting and Isolating a Measurement Sensitivity Problem

Table 5.1.1. Scenario Simulated

Distribution at First Use	mean	variance, $\boldsymbol{\Sigma}_{1\|0}$		standard deviation		precision $\boldsymbol{\Sigma}_{1\|0}^{-1}$	
	$\mathbf{x}_{1\|0}$	$b_{0,1}$	$b_{1,1}$	$b_{0,1}$	$b_{1,1}$	$b_{0,1}$	$b_{1,1}$
$b_{0,1}$	0	0.1	0	0.316		10	0
$b_{1,1}$	1	0	0.1		0.316	0	10

Observation error	variance σ_v^2	standard deviation	precision σ_v^{-2}
	0.01	0.1	100

Migration	mean	variance, $\mathbf{W}$		standard deviation		precision $\mathbf{W}^{-1}$	
	$\mathbf{x}_{t+1}$	$b_{0,t+1}$	$b_{1,t+1}$	$b_{0,t+1}$	$b_{1,t+1}$	$b_{0,t+1}$	$b_{1,t+1}$
simulated $b_{0,t+1}$	0.03	0	0	0		∞	0
$b_{1,t+1}$	0	0	0		0	0	∞
assumed $b_{0,t+1}$	0	0.001	0	0.0316		1000	0
$b_{1,t+1}$	0	0	0.001		0.0316	0	1000

Table 5.1.2. Illustrative Calculations

		Simulated		Prior from previous step 2					Intermediate computations					
		true	Observation		variance		precision		precision		variance		Kalman	prediction
		state	$y_t = b_{0,t} +$	mean	$\boldsymbol{\Sigma}_{t\|t-1}$		$\boldsymbol{\Sigma}^{-1}_{t\|t-1}$		$\boldsymbol{\Sigma}^{-1}_{t\|t}$		$\boldsymbol{\Sigma}_{t\|t}$		Gain	error
time		$\mathbf{x}_t$	$b_{1,t}u_t + v_t$	$\mathbf{x}_{t\|t-1}$	$b_{0,t}$	$b_{1,t}$	$b_{0,t}$	$b_{1,t}$	$b_{0,t}$	$b_{1,t}$	$b_{0,t}$	$b_{1,t}$	$\mathbf{K}_t$	$\hat{e}_t$
eq'n→		(5.1)	(5.2)	(5.19) + (5.20)	(5.21)				(5.14)				(5.16)	(5.18) + (5.5)
1	$b_{0,t}$	– 0.012		0	0.1	0	10.0	0	110.0	86.3	0.0458	– 0.0468	0.542	
	$b_{1,t}$	1.113		1	0	0.1	0	10.0	86.3	84.5	– 0.0468	0.0596	0.468	
			$u_t =$ 0.863											0.024
			$y_t =$ 0.887											
2	$b_{0,t}$	0.018		0.013	0.0468	– 0.0468	93.5	72.2	193.5	9.0	0.0052	– 0.0004	0.545	
	$b_{1,t}$	1.113		1.011	– 0.0468	0.0606	72.2	72.2	9.0	112.1	– 0.0004	0.0090	– 0.607	
			$u_t =$ – 0.631											– 0.183
			$y_t =$ – 0.809											
3	$b_{0,t}$	0.048		– 0.087	0.0062	– 0.0004	162.1	6.8	262.1	99.2	0.0048	– 0.0025	0.243	
	$b_{1,t}$	1.113		1.122	– 0.0004	0.0099	6.8	100.7	99.2	186.0	– 0.0025	0.0067	0.367	
			$u_t =$ 0.924											0.169
			$y_t =$ 1.119											
		…		…					…					
18	$b_{0,t}$	0.498		0.381	0.0043	– 0.0014	256.8	71.3	356.8	23.1	0.0028	– 0.0003	0.295	
	$b_{1,t}$	1.113		1.128	– 0.0014	0.0050	71.3	219.5	23.1	242.7	– 0.0003	0.0041	– 0.227	
			$u_t =$ – 0.483											0.120
			$y_t =$ – 0.043											
19	$b_{0,t}$	0.528		0.417	0.0038	– 0.0003	262.7	13.7	362.7	17.8	0.0028	– 0.0003	0.276	
	$b_{1,t}$	1.113		1.101	– 0.0003	0.0051	13.7	195.1	17.8	195.2	– 0.0003	0.0051	– 0.004	
			$u_t =$ 0.041											0.224
			$y_t =$ 0.686											
20	$b_{0,t}$	0.558		0.478	0.0038	– 0.0003	266.0	10.9	366.0	66.0	0.0028	– 0.0010	0.236	
	$b_{1,t}$	1.113		1.100	– 0.0003	0.0061	10.9	163.2	66.0	193.5	– 0.0010	0.0055	0.204	
			$u_t =$ 0.551											0.103
			$y_t =$ 1.188											
21	$b_{0,t}$	0.588		0.503	0.0039	– 0.0010	266.0	40.6						
	$b_{1,t}$	1.113		1.121	– 0.0010	0.0065	40.6	159.9						

The defining parameters of the Kalman filter are given in Table 5.1.1. The distribution at first use is assumed to be a bivariate normal with mean $\mathbf{x}_{1|0} = (0, 1)'$ and covariance $\mathbf{\Sigma}_{1|0} = 0.1\ \mathbf{I}_2$. The noise and migration parameters are similarly given in Table 5.1.1. In many applications, these numbers are all obtainable from external sources: For manufactured products, the distribution at first use can often be obtained from data collected at final test. The noise variance can be obtained from a study of gage repeatability and reproducibility (NIST 2001). The migration parameters can be modeled parsimoniously with reference to reliability data, as suggested in section 3 above. In biostatistics, the "distribution at first use" of the condition of human subjects and the migration parameters can be estimated from previous similar clinical studies.

The simulated true but unknown sensitivity parameters at time $t = 1, 2, 3, ..., 19, 20$, and 21 were determined as described in Table 5.1.1. A single pair of pseudo-random numbers following a bivariate normal with mean $(0, 1)'$ and variance $0.1\mathbf{I}_2$ were generated to obtain the simulated true value at $t = 1$ of $(-0.012, 1.113)'$. After that, the simulated condition of the plant was changed by an increment of $(0.03, 0)'$ at each point in time, progressing to $(0.018, 1.113)'$ at $t = 2$, to $(0.048, 1.113)'$ at $t = 3$, to ... $(0.588, 1.113)'$ at $t = 21$.

Simulated values for the reference samples u_t were generated to follow $N(0, 1)$ and appear in the third column of Table 5.1.2 along with simulated observations y_t computed as $\mathbf{H}_t\mathbf{x}_t + v_t = [1, u_t]\mathbf{x}_t + v_t$ with observation error $v_t \sim N(0, 0.01)$ as described in Table 5.1.1; these pairs of values are plotted in Figure 5.1.

Kalman filtering begins with the initial prior at time $t = 1$ taken from the distribution at first use given in Table 5.1.1. Step 1.1a tells us to compute the forecast per

(5.5) as $f_t = \mathbf{H}_t \mathbf{x}_{t|t-1} = (1, 0.863)\ (0, 1)' = 0.863$ for $t = 1$. We will not use the predictive standard deviation $\sigma_{y|t-1}$ in this discussion and so will not compute it. In many applications, it is wise to evaluate the prediction error e_t to determine if it is plausible or might be an outlier; in that case, we would use $\sigma_{y|t-1}$ to evaluate this. In other applications, we wish to compute the log(likelihood) for comparison with other models; in those cases as well, we would need $\sigma_{y|t-1}$.

Step 1.1b tells us to compute the posterior precision as the sum of the observed information from prior and observation per (5.14). At time $t = 1$, the information about $\mathbf{x}_t$ contained in the data is

$$\mathbf{H}_1' \sigma_v^{-2} \mathbf{H}_1 = \begin{bmatrix} 1 \\ 0.863 \end{bmatrix} (100) \begin{bmatrix} 1 & 0.863 \end{bmatrix} = \begin{bmatrix} 100 & 86.3 \\ 86.3 & 74.5 \end{bmatrix}.$$

To this, we add the prior information $\mathbf{\Sigma}_{t|t-1}^{-1} = 10\,\mathbf{I}_2$ to get the posterior information $\mathbf{\Sigma}_{t|t}^{-1}$ for $t = 1$ given in Table 5.1.2. The inverse of this posterior information is the posterior covariance matrix $\mathbf{\Sigma}_{t|t}$ given just to the right of $\mathbf{\Sigma}_{t|t}^{-1}$.

The posterior covariance matrix is then post-multiplied by $\mathbf{H}_t' = (1, 0.863)'$ and the observation precision $\sigma_v^{-2} = 100$ to get the Kalman gain $\mathbf{K}_t = (0.542, 0.468)'$ for $t = 1$, completing step 1.1c.

With these preparations done after the previous observation but before the current observation arrives, we now suppose that the latest observation arrived as $y_1 = 0.887$. The forecast $f_t = 0.863$ is subtracted per step 1.2a (5.18) to obtain the prediction error $e_t = 0.024$. This is then multiplied by the Kalman gain as $\mathbf{K}_t e_t = (0.013, 0.011)'$ and added to

the prior mean $\mathbf{x}_{t|t-1}$ per (5.19) to get the posterior mean $\mathbf{x}_{t|t}$ = (0.013, 1.011)' for t = 1, ending step 1.2b.

Step 2.1 tells us that the prior mean for the next observation is just the posterior mean from the previous observation. Thus, for t = 2, $\mathbf{x}_{t|t-1}$ is the vector just computed (0.013, 1.011)'. Similarly, per (5.21) step 2.2, the prior covariance matrix $\mathbf{\Sigma}_{t|t-1}$ is $0.001\mathbf{I}_2$ plus the posterior covariance matrix from the previous observation.

After repeating this iteration for all 20 observations, we got the rest of the numbers appearing in Table 5.1.2, plus others not shown there. The prior means are plotted in Figure 5.2.

5.3. Literature Review

Kalman (1960) developed the basic mathematics discussed herein to solve problems of smoothing and forecasting. He derived the minimum mean square error linear predictor, not the Bayesian posterior. When all distributions are normal, these two approaches are equivalent. In certain cases with, for example, non-normal observations, Bayesian sequential updating as discussed here can produce nonlinear estimators that outperform Kalman's minimum mean square error linear predictor, at least for some purposes.

For a more recent presentation of Kalman filtering, see, e.g., Gelb (1990). For enhancements, see, e.g., Siouris (1996), Mosca (1995), and Brown and Hwang (1997). For a discussion with a more statistical flavor, more consistent with our "Bayesian sequential updating" approach, see, e.g., Harvey (1989), Pole, West and Harrison (1994),

or West and Harrison (1999). These latter two references also include simultaneous estimation of a relative precision parameter, related to the "EWMA for variance", discussed in section 4 above.

Fault detection and isolation is an important topic in the advanced control theory literature. For recent contributions in that area, see, e.g., Chen and Patton (1999), Gertler (1998), and Natke and Cempel (1997).

The use of Bayesian sequential methods for "model monitoring" and for detecting "discontinuous changes in time series" is discussed by Gordon and Smith (1988), West (1986), Pole, West and Harrison (1994) and West and Harrison (1999). Gordon and Smith (1988) recommend "model selection" based on the posterior probability of a particular model operating. West (1986), Pole, West and Harrison (1994), Harrison and Lai (1999) and West and Harrison (1999) recommend the use of "cumulative Bayes' factors" to determine if an alternative model better describes the data; the logarithm of "cumulative Bayes' factors" is, in essence, a one-sided Cusum of log(likelihood ratio). As discussed in section 2 above, this is virtually equivalent to the increase in the posterior log odds of an abrupt jump to the alternative model, relative to the log hazard odds.

5.4. Discussion

A Kalman filter monitor similar to that discussed here will generally be fairly responsive to a real malfunction, especially if the failure process is gradual, roughly consistent with the assumed migration rate. In that case, the posterior $\mathbf{x}_{t|t}$ and the subsequent prior $\mathbf{x}_{t+1|t}$ will tend to track the true but unknown $\mathbf{x}_t$ fairly closely. An abrupt

failure will still be identified by this monitor, though perhaps not quite as quickly as by a monitor designed specifically to detect abrupt jumps. Such applications could be tracked with a Cusum (e.g., Bayes-adjusted as discussed in section 2 above) of log(likelihood ratio) of the predictive distribution assuming non-zero migration variance **W** in (5.1) compared to one that assumes **W** = **0**. As for other Cusums, detection thresholds can be selected using Monte Carlo simulation to estimate and balance selected characteristics of the run length distributions for good and bad systems.

REFERENCES

DeGroot, M. H. (1970) *Optimal Statistical Decisions* (NY: McGraw-Hill).

Gelb, A. (1999) *Optimal Applied Estimation* (Cambridge, MA: MIT Press).

Gordon, K., and Smith, A. F. M. (1988) "Modeling and Monitoring Discontinuous Changes in Time Series" in *Bayesian Analysis of Time Series and Dynamic Models*, ed. J. Spall, N.Y.: Marcel Dekker, 359-391.

Harrison, P. J., and Lai, I. C. H. (1999) "Statistical Process Control and Model Monitoring", *Journal of Applied Statistics*, 26: 273-292.

Harvey, A. C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter* (NY: Cambridge University Press).

Kalman, R. E. (1960) "A New Approach to Linear Filtering and Prediction Problems", *Journal of Basic Engineering*, 340-345.

NIST (2001) *Engineering Statistics Handbook* (Washington, DC: National Institute of Standards and Technology web-based handbook: http://www.itl.nist.gov/div898/handbook)

Pole, A., West, M., and Harrison, H. (1994) *Applied Bayesian Forecasting and Time Series Analysis* (NY: Chapman & Hall)

Press, S. J. (1972) *Applied Multivariate Analysis* (NY: Holt, Rinehart and Winston).

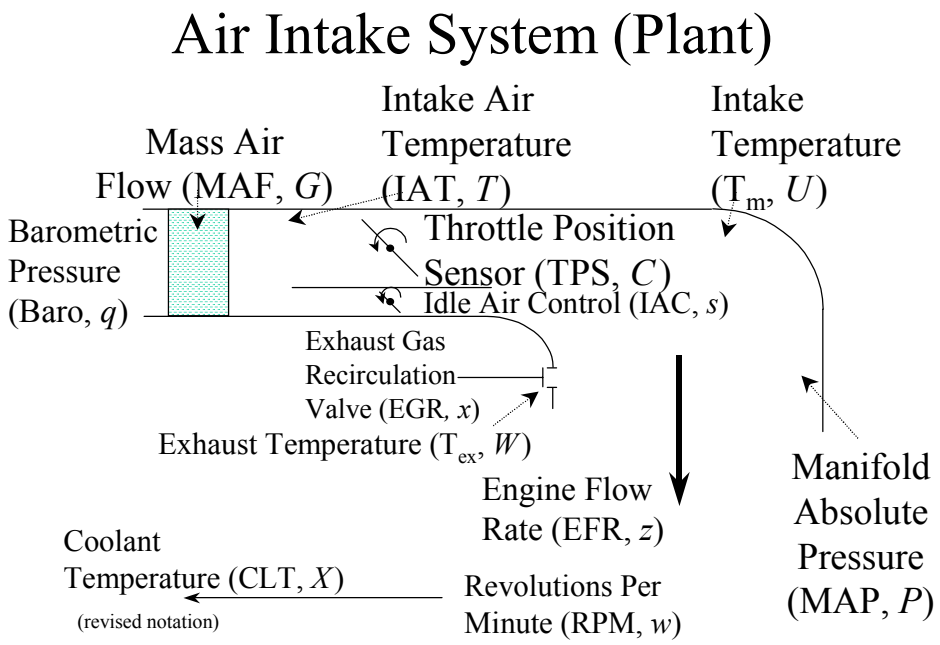
West, M. (1986) "Bayesian Model Monitoring", *Journal of the Royal Statistical Association* B, 48: 70-78.

West, M. and Harrison, P. J. (1999) *Bayesian Forecasting and Dynamic Models*, 2nd ed., corrected printing (NY: Springer).

6. FAULT ISOLATION USING ANALYTICAL REDUNDANCY

Fault detection and isolation has traditionally been accomplished by adding a separate sensor with companion hardware and software for each failure mode of interest. This increases the both the per unit cost and the physical mass of the product. The cost per unit is always a concern; in some applications (e.g., aviation), the increased weight can also be an issue. An alternative that is sometimes available is to use *analytical redundancy*, identifying and isolating faults from inconsistencies between readings from multiple sensors. To fix ideas, consider the air intake system of an automobile engine portrayed in Figure 6.1. In this case, for example, equations of fluid flow assert that a certain relationship should hold between the readings for mass air flow (MAF), the manifold absolute pressure (MAP), and the throttle position sensor (TPS), apart from model inadequacies and measurement noise whose magnitudes can be estimated. In this section, we discuss the use of Kalman filtering to continually reestimate sensitivity parameters, assuming they may drift as discussed in the previous section. In this way, we hope to identify and isolate faults with any of these primary sensors. We illustrate the method in this section and discuss an extension in the next section.

Figure 6.1. Estimating Sensitivity Parameters in a Measurement Process



In particular, section 6.1 considers a simplified version of this problem, attempting to diagnose sensitivity faults with the mass air flow (MAF) sensor and the throttle position sensor (TPS), considering MAF as both a univariate output and as a component of the state vector. To avoid confusion, we will attempt to carefully distinguish at each point in this discussion whether we are referring to the true but unknown value of MAF, a model prediction, or a noisy measurement. The consequences of misspecifying noise and migration variance parameters is discussed in section 6.2 with further general discussion provided in section 6.3. Mathematical details are described in appendix section 6.4. The example of section 6.1 is extended in section 7 to use bivariate outputs, MAF and manifold absolute pressure (MAP), to isolate problems that might arise with either MAF, MAP or TPS sensors. This theory isolated faults for us with one or two sensors but failed when considering potential problems with three sensors. Possible

reasons for this are discussed in section 6.4. In many applications, these problems can be overcome using multiple model adaptive estimation (MMAE), discussed in section 7.

6.1. Bayesian Updating for Fault Isolation

We consider a possibly nonlinear, dynamic plant described as follows:

$$\mathbf{x}_{t+1} = \mathbf{f}_t + \mathbf{w}_t, \text{ where } \mathbf{f}_t = \mathbf{f}_t(\mathbf{x}_t, \mathbf{u}_t),\ \mathbf{w}_t \sim N_p(0, \mathbf{W}_t), \tag{6.1}$$

$\mathbf{x}_t$ = state vector, and $\mathbf{u}_t$ = vector of controls and other inputs; we assume that $\mathbf{f}_t(\mathbf{x}_t, \mathbf{u}_t)$ is continuously differentiable with respect to $\mathbf{x}_t$ throughout the operating region. In this section, we focus on a special case with

$$\mathbf{x}_t = \begin{Bmatrix} G_t \\ a_{0,t} \\ a_{1,t} \\ b_{0,t} \\ b_{1,t} \end{Bmatrix}, \text{ and } \mathbf{f}_t(\mathbf{x}_t, \mathbf{u}_t) = \begin{Bmatrix} \widetilde{G}_t(\mathbf{x}_t, \mathbf{u}_t) \\ a_{0,t} \\ a_{1,t} \\ b_{0,t} \\ b_{1,t} \end{Bmatrix}, \tag{6.2}$$

where

G_t = true but unknown MAF at time t (see Figure 6.1),

$\widetilde{G}(\mathbf{x}_t, \mathbf{u}_t)$ = predicted MAF at time $(t + 1)$ given state and inputs $(\mathbf{x}_t, \mathbf{u}_t)$,

$a_{0,t}$ and $a_{1,t}$ are the true but unknown MAF sensitivity parameters,

$b_{0,t}$ and $b_{1,t}$ are the true but unknown TPS sensitivity parameters, and

$\mathbf{u}_t$ = a vector of inputs including measured TPS_t, which is related to the true but unknown throttle position at time t, C_t, as $\text{TPS}_t = c_{0,t} + c_{1,t}C_t$.

In section 7, expression (6.2) will be generalized to include in the state vector $\mathbf{x}_t$ the manifold absolute pressure, MAP, P_t, and different combinations of sensitivity

parameters for these and other sensors. In the present discussion, we shall assume that the functions $\widetilde{G}(\mathbf{x}_t, \mathbf{u}_t)$ and $\mathbf{f}_t(\mathbf{x}_t, \mathbf{u}_t)$ are known.

The first of the six lines in Figure 6.2 presents 2,000 observations at 100 millisecond intervals on MAF through the air intake system of a 1998 production vehicle with a 4.6 liter dual overhead cam (DOHC) V-8 engine modified for test purposes; thus, this image spans 200 seconds. The second line presents the TPS reading for the same period. This figure matches our intuition that the air flow tends to be high when the throttle is open, and conversely the air flow is low when the throttle is closed. Figure 6.3 presents a simulation of a dual fault, where the minimum mass air flow is 20, even when the throttle is closed, and where the TPS readings are 40 percent larger than they should be. The bottom four lines in Figures 6.2 and 6.3 show the estimated trajectories of sensitivity parameters for MAF and TPS for the corresponding MAF and TPS readings.

With the good plant in Figure 6.2, we see that the estimated ($a_{0,t}$, $a_{1,t}$, $b_{0,t}$, $b_{1,t}$) stay relatively close to (0, 1, 0, 1), as they should. However, when the simulated fault appears in the middle of Figure 6.3, the estimated values for ($a_{0,t}$, $a_{1,t}$, $b_{0,t}$, $b_{1,t}$) move fairly quickly towards the simulated values of (20, 1, 0, 1.4). Roughly 30 seconds (300 observations) after the onset of the simulated fault, the estimated sensitivities have largely adjusted to the change, with $a_{0,t}$ and $b_{0,t}$, close to the simulated values of 20 and 0, respectively, while the slope $b_{1,t}$ increased to 1.3, stopping short of the simulated 1.4 and the slope $a_{1,t}$ drifted away from the simulated 1 to around 0.8.

Figure 6.2. Mass Air Flow Measured, Throttle Position, and Filtered Sensitivities

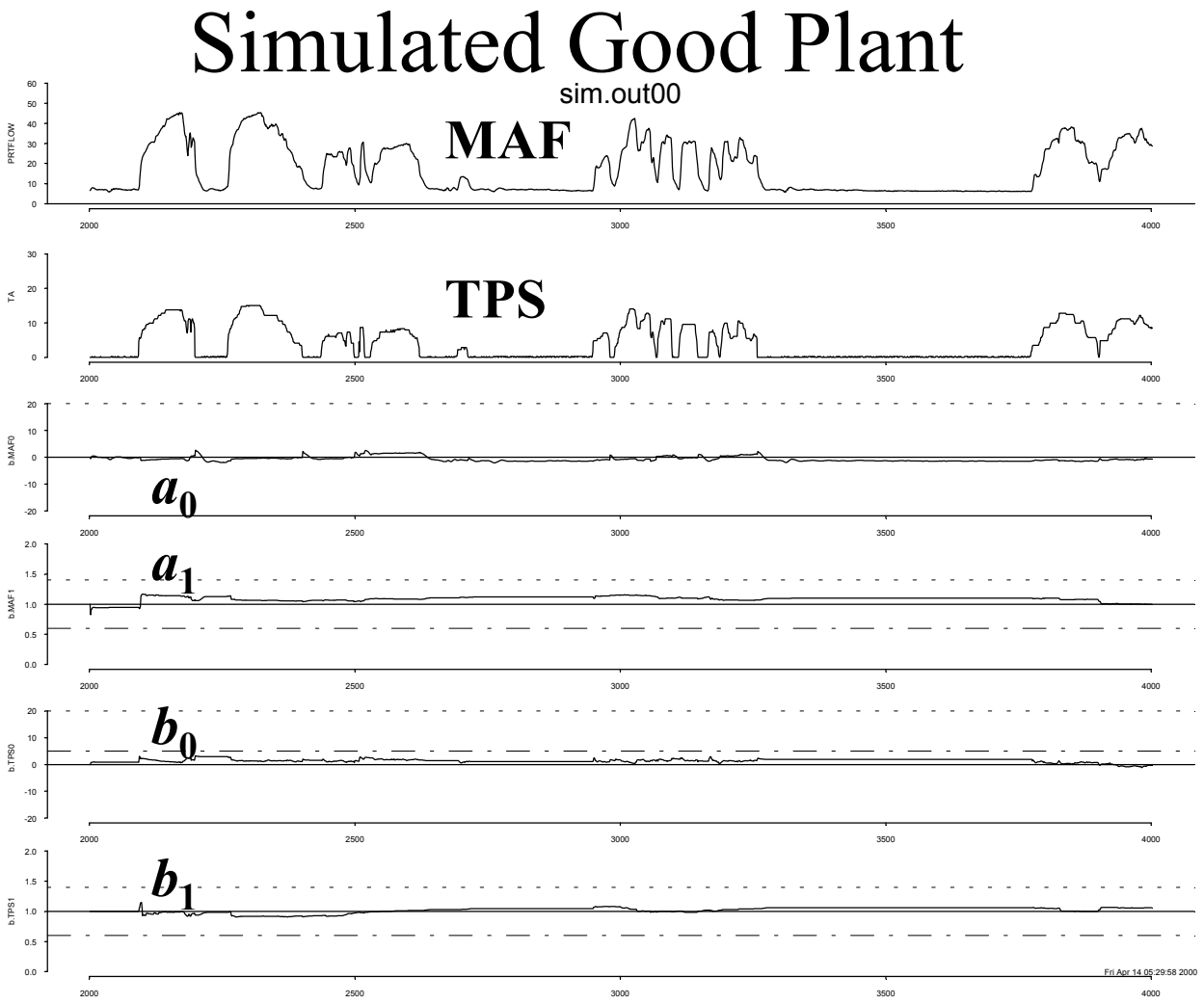
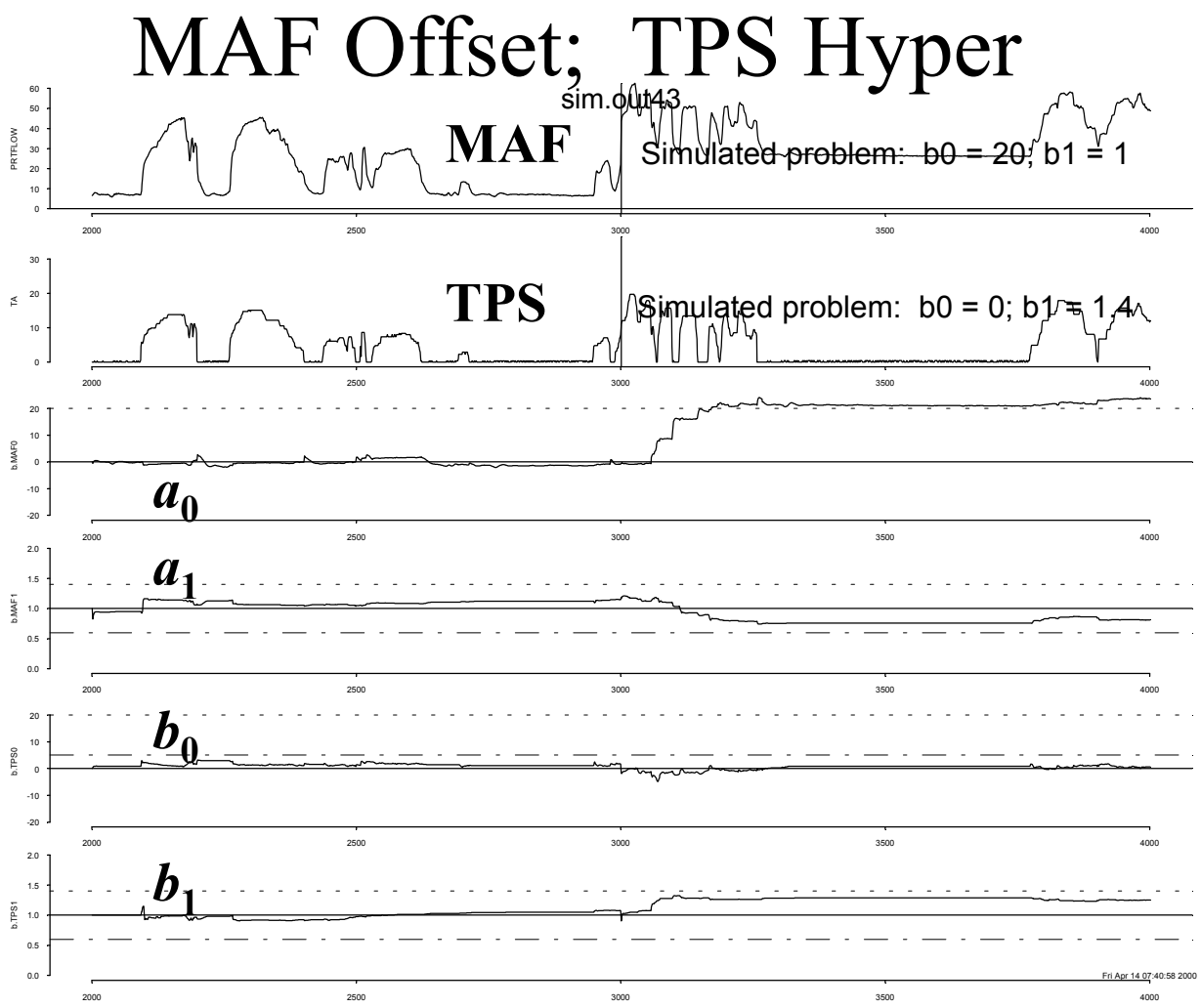


Figure 6.3. Filtered Sensitivities with Simulated Faults



The difference between $(a_{1,t}, b_{1,t}) \approx (0.8, 1.3)$ and the simulated (1, 1.4) push us to ask "Why?" Unfortunately, since this is a nonlinear plant, it is difficult to evaluate the exact source of this bias. However, we believe it is related either to the nonlinear nature of $\widetilde{G}(\mathbf{x}_t, \mathbf{u}_t)$ or to controls that place a premium on vehicle performance in ways that provide poor leverage for estimating sensitivity parameters; this issue is considered further in the discussion section 6.3 below.

Filtered estimates in situations such as this are obtained from noisy measurements as follows:

$$\mathbf{y}_t = \mathbf{h}_t + \mathbf{v}_t, \text{ where } \mathbf{h}_t = \mathbf{h}_t(\mathbf{x}_t, \mathbf{u}_t), \mathbf{v}_t \sim N_k(0, \mathbf{V}_t), \tag{6.3}$$

and $\mathbf{x}_t$ and $\mathbf{u}_t$ are vectors of state and inputs, as with (6.1). In the example of this section, $p = 1$, and $\mathbf{h}_t(\mathbf{x}_t, \mathbf{u}_t), = a_{0,t} + a_{1,t}G_t$, where G_t = the true but unknown MAF at time t. In the examples of section 7, $p = 2$, so $\mathbf{h}_t$ is multivariate in addition to being a nonlinear function of $\mathbf{x}_t$.

To complete the specification of this portion of the model per (6.1) and (6.3), we need values for the transition and observation covariance matrices, $\mathbf{W}_t$ and $\mathbf{V}_t$. For Figures 6.2 and 6.3, we took $\mathbf{V}_t = 0.2$ and $\mathbf{W}_t$ = a diagonal matrix with elements $\mathbf{W}_t(G_t) = 1$, $\mathbf{W}_t(a_{0,t}) = \mathbf{W}_t(b_{0,t}) = 0.1$, and $\mathbf{W}_t(a_{1,t}) = \mathbf{W}_t(b_{1,t}) = 10^{-5}$.

As in sections 3-5, we initiate Bayesian sequential updating by assuming that $\mathbf{x}_1 \sim N_p(\mathbf{x}_{1|0}, \mathbf{\Sigma}_{1|0})$, with $\mathbf{x}_{1|0} = (20, 0, 1, 0, 1)'$ and $\mathbf{\Sigma}_{1|0} = \text{diag}(100, 4, 0.1, 4, 0.1)$. As discussed in section 6.4, these assumptions with (6.1) and (6.3) drive the Kalman iteration outlined in Figure 3.3 above to produce the images in Figures 6.2 and 6.3.

These values for $\mathbf{V}_t$, $\mathbf{W}_t$, $\mathbf{x}_{1|0}$, and $\mathbf{\Sigma}_{1|0}$ were chosen after conducting several designed experiments varying appropriate combinations of elements systematically and selecting combinations that seemed to provide the greatest discrimination between good and bad. Brown and Hwang (1997, sec. 9.3) and West and Harrison (1999, sec. 12.2) suggest running multiple Kalman filters in parallel and recursively computing posterior probabilities of the different combinations of parameters. This should eventually settle on the best combination of parameter values. While this would ultimately be a wise thing to do, it should be done in conjunction with simulating a suite of faults on a variety of different vehicles. Otherwise, we might select a model and parameters that predict better the behavior of a particular good plant but is unable to detect faults or perform well under untested operating conditions or with a different copy of ostensibly the same plant. Moreover, we want a monitor that performs well even with modest misspecification of model parameters. If it requires too much tuning to make it work, it may not perform properly with the inevitable variability between ostensibly equivalent applications, e.g., different copies of the same design vehicle or different patients in the same clinical trial or different production lines for similar products. Also, before we expend the resources to tune a monitor, we want to have some confidence that our monitor is probably otherwise close to being acceptable. In this case, we concluded that to diagnose faults of interest with the air intake system of Figure 6.1, we needed multiple model adaptive estimation, discussed in section 7.

6.2. Consequences of Misspecifying Noise and Migration Variances

To evaluate the consequences of misspecifying $\mathbf{W}_t$ and $\mathbf{V}_t$, we first note that a single common "relative precision" parameter can be factored out of both $\mathbf{W}_t$ and $\mathbf{V}_t$ without affecting the estimated state of the plant. This can be seen in the Kalman Gain computations described below. In the special case where observation $\mathbf{y}_t$ and state $\mathbf{x}_t$ are both univariate and where $\mathbf{f}_t(\mathbf{x}_t, \mathbf{u}_t) = \mathbf{h}_t(\mathbf{x}_t, \mathbf{u}_t) = \mathbf{x}_t$, this has already been discussed the previous section on a "Bayesian EWMA for mean and variance". There, we learned that the relative precision affects the width of confidence intervals but not the prior and posterior means for $\mathbf{x}_t$. This principle applies also to the present development: If $\mathbf{W}_t$ and $\mathbf{V}_t$ are both multiplied by a factor of 100, the width of confidence intervals is multiplied by a factor of $10 = \sqrt{100}$.

This common factor in $\mathbf{W}_t$ and $\mathbf{V}_t$ can be estimated sequentially by replacing $\mathbf{W}_t$ and $\mathbf{V}_t$ with $(\mathbf{W}_t/\phi_t)$ and $(\mathbf{V}_t/\phi_t)$, respectively, and developing the theory following the steps described with the previous section on a "Bayesian EWMA for mean and variance"; for linear Kalman theory, this has already been discussed by West and Harrison (1999) and Pole, West and Harrison (1994). If we do this and then multiply $\mathbf{W}_t$ and $\mathbf{V}_t$ by 100, we will find that ϕ_t would be replaced by $100\phi_t$, the EWMA for variance would be divided by 100, and the resulting confidence intervals for prior and posterior state and prediction intervals for future observations would be unaffected. We shall not pursue this generalization here.

While the overall level of $\mathbf{W}_t$ and $\mathbf{V}_t$ directly affect the width of confidence intervals, the relative magnitudes of different elements of $\mathbf{W}_t$ and $\mathbf{V}_t$ affect the filtered estimates in more subtle ways.

In general, we favor setting $\mathbf{V}_t$ via studies of gage repeatability and reproducibility and $\mathbf{W}_t$ considering reliability data, as discussed in the previous section on "designing Bayesian EWMA monitors using gage R & R and reliability data". Failing that (or in addition), we prefer to estimate an EWMA for variance in conjunction with the means and conditional variance parameters of prior and posterior distributions for the state vector and for the predictive distributions, generalizing the present section and the previous section on a "Bayesian EWMA for mean and variance".

While developing prototype monitors using the theory described herein, we conducted several designed experiments, varying different elements of $\mathbf{W}_t$ and $\mathbf{V}_t$ systematically to evaluate their impact on monitor performance. In a typical experiment, the migration parameters for intercept and slope were varied by factors of 10 and 100 in a $2\times 2+(\text{center point})$ design, outlined in 5 of the 13 rows of Table 7.1 in the next section. Clearly, the effectiveness of an otherwise acceptable monitor could be destroyed by sufficiently bad choices for $\mathbf{W}_t$ and $\mathbf{V}_t$. However, in our limited experiments of this nature, we saw no cases where an acceptable monitor ceased to function properly, nor where an unacceptable monitor could be made to function dramatically better. Our conclusion from this and other efforts is that the primary point of leverage in monitoring is in better models, improving $\mathbf{f}_t(\mathbf{x}_t,\ \mathbf{u}_t)$ and $\mathbf{h}_t(\mathbf{x}_t,\ \mathbf{u}_t)$, and thereby also reducing the magnitude of $\mathbf{W}_t$ and $\mathbf{V}_t$; better choices for $\mathbf{W}_t$ and $\mathbf{V}_t$ can improve monitor performance, but monitor performance seemed reasonably robust over a fairly wide range of plausible values for $\mathbf{W}_t$ and $\mathbf{V}_t$.

6.3. Discussion

With simulations such as those summarized in Figures 6.2 and 6.3, we have demonstrated a potential for fault isolation using Kalman filters that include ongoing reestimation of potentially drifting sensitivity parameters for various sensors. However, when we generalized the model of (6.2) to include three pairs of sensitivity parameters, for MAF, MAP and TPS (as described in Figure 6.1), the Kalman filter just got lost: In our simulations, random variability in the observations translated into excessive movement in the 6 sensitivity parameters that made it extremely difficult to diagnose a fault from the estimated sensitivity parameters and in some cases even hampered the ability of the filter to predict future observations. We were able to reduce the severity of some of the instabilities we encountered by careful choice of noise and migration variances. In addition, some improvements could doubtless be obtained by first transforming all variables so they are approximately centered at 0 with typical variation (e.g., standard deviation) of 1; this will improve the numerical stability of the algorithm. However, even if rescaling produced acceptable fault isolation with this 8-dimensional Kalman filter, we would likely still encounter the problem if we wanted also to isolate faults with the exhaust gas recirculation valve or the idle air control or something else. Moreover, a Kalman filter that will work fine with the double precision (64-bit) arithmetic now standard with modern computers may fail miserable with the 8- or 16-bit microprocessors used for control and fault detection plus isolation in many products sold in the thousands and millions of units. An algorithm that requires 64- or 128-bit arithmetic will not likely become a production standard in this environment.

Part of this problem is related to the concept of leverage in regression (Belsley, Kuh, and Welsch 1980): Every student of algebra learns that two points determine straight line, with some restrictions. Given 100 observations on y with $x = 2$ in all 100 observations, we cannot estimate both b_0 and b_1 in $y = b_0 + b_1 x$. Similarly, given 100 observations on y with $x_1 = x_2$ in all 100 observations, we cannot estimate both b_1 and b_2 in $y = b_0 + b_1 x_1 + b_2 x_2$. By extension, it is easy to imagine a situation where the laws of physics combined with automatic controls might constrain MAF, MAP and TPS to move together in ways that would not support simultaneous estimation of sensitivity parameters for three different sensors. In some cases, the leverage required for estimation might be obtained by dithering the automatic controls. However, this might have an emissions impact and might therefore displease governmental regulators and others concerned about air quality, even if it were only invoked by an automotive technician for off-line fault isolation. It might also cause the vehicle to cough and shudder; few auto makers would want to tell to a customer, "That's a feature: We designed the car to do that!"

Fortunately, Menke and Maybeck (1995) and Eide and Maybeck (1996) have provided a vision of a principle for fault detection and isolation that will not require such extreme measures in many cases: They suggest multiple model adaptive estimation (MMAE), which involves simultaneously running multiple Kalman filters, each designed to detect a different fault or combination of faults. They were concerned with determining when to eject a pilot from a high performance fighter aircraft, where a delay of more than half a second from fault onset may make it impossible to eject the pilot safely, thereby sacrificing the pilot as well as the aircraft. In section 7, we discuss MMAE.

6.4. Appendix: Extended Kalman Filtering

In this section, we derive "extended Kalman filtering" from Bayesian sequential updating, as outlined in Figure 3.3. West and Harrison (1999, p. 496) cite earlier literature that derive this from the principle of minimum mean squared error prediction. In the process, we also derive this for non-normal observations assuming approximate normality of the state vectors. West and Harrison (1999) discuss a variety of approaches to nonlinear and non-normal dynamic modeling including the following:

- Linearization (sec. 13.2)
- Various techniques for numerical integration including Gaussian quadrature and a variety of Monte Carlo techniques (sec. 13.4 - 13.6 and ch. 15),
- Multi-process models that may switch between alternative dynamic linear models at each observation (ch. 12), and
- Exponential family dynamic models (ch. 14).

In this section, we derive Bayesian sequential updating with nonlinear transitions and possibly non-normal observations, assuming the distribution of the state is always adequately approximated by normal distributions. Our approach is to linearize the log(probability density). Compared to West and Harrison's (1999, ch. 14) exponential family dynamic models, our approach does not require observations following an exponential family and is furthermore, we believe, simpler when the approximations involved are adequate.

Following West and Harrison (1999, p. 495), we are more concerned with forecast accuracy than with how well expression (6.1) actually models the transition process. First order Taylor approximations do not have to be highly accurate for this system to work well. We need only that (*a*) with moderately high probability, the error is moderate relative to the stochastic component, and (*b*) when the error in the linear approximation is substantial, the plant and model are still sufficiently stable that these deficiencies will be corrected with subsequent observations.

The utility of the normal distribution as an approximation to the distribution of the state vector is supported by the central limit theorem: Under suitable conditions, the distribution of a weighted sum of random variables is generally more nearly normal than the distributions of the component random variables. The "suitable conditions" may not be satisfied if the transition or observation functions, $\mathbf{f}_t(\mathbf{x}_t, \mathbf{u}_t)$ and $\mathbf{h}_t(\mathbf{x}_t, \mathbf{u}_t)$ in (6.1) and (6.3), are highly nonlinear and violently discontinuous in regions occupied with moderately high probability. However, the applications of interest here will generally be sufficiently well behaved that any non-normalities introduced by nonlinearities in $\mathbf{f}_t(\mathbf{x}_t, \mathbf{u}_t)$ and $\mathbf{h}_t(\mathbf{x}_t, \mathbf{u}_t)$ can be ignored. For further discussion of these issues, see, e.g., Gnedenko and Kolmogorov (1968), Rao (1973), or Skovgaard (1986).

As in previous sections, we organize this development around the two steps of Bayesian sequential updating: 1. Observation, and 2. Transition. We begin with some implications of Bayes' theorem, which seem to us to be new and which provide a more direct path to the required computations than the less general approach discussed in previous sections.

Step 1. Observation Using Bayes' Theorem. We begin with the following relationship based on the definition of conditional probabilities:

$$\begin{array}{c} p(\mathbf{y}_t, \mathbf{x}_t \mid D_{t-1}) = p(\mathbf{y}_t \mid D_{t-1}) p(\mathbf{x}_t \mid \mathbf{y}_t, D_{t-1}) = \; p(\mathbf{y}_t \mid \mathbf{x}_t) \;\; p(\mathbf{x}_t \mid D_{t-1}) \\ (\text{joint}) \qquad = (\text{predictive}) \times (\text{posterior}) = (\text{observation}) \times (\text{prior}) \end{array},$$

where $p(\,.\mid .\,)$ = probability density. But $D_t = \{\mathbf{y}_t, D_{t-1}\}$, so $p(\mathbf{x}_t \mid \mathbf{y}_t, D_{t-1}) = p(\mathbf{x}_t \mid D_t)$. Now taking logarithms, letting $l(.\mid .) = \ln[p(.\mid .)]$, we get the following:

$$\begin{array}{c} l(\mathbf{y}_t \mid D_{t-1}) + l(\mathbf{x}_t \mid D_t) \;\; = \;\; l(\mathbf{y}_t \mid \mathbf{x}_t) \;\; + l(\mathbf{x}_t \mid D_{t-1}) \\ (\text{predictive}) + (\text{posterior}) = (\text{observation}) + (\text{prior}) \end{array}.$$

We will use first and second derivatives of this expression relative to $\mathbf{x}_t$:

$$\frac{\partial l(\mathbf{x}_t \mid D_t)}{\partial \mathbf{x}_t} = \frac{\partial l(\mathbf{y}_t \mid \mathbf{x}_t)}{\partial \mathbf{x}_t} + \frac{\partial l(\mathbf{x}_t \mid D_{t-1})}{\partial \mathbf{x}_t}, \tag{6.4}$$

and

$$\frac{\partial^2 l(\mathbf{x}_t \mid D_t)}{\partial \mathbf{x}_t \partial \mathbf{x}_t'} = \frac{\partial^2 l(\mathbf{y}_t \mid \mathbf{x}_t)}{\partial \mathbf{x}_t \partial \mathbf{x}_t'} + \frac{\partial^2 l(\mathbf{x}_t \mid D_{t-1})}{\partial \mathbf{x}_t \partial \mathbf{x}_t'}. \tag{6.5}$$

Now recall that the *observed information* is defined as the negative of the second derivative of the log(likelihood), and the *Fisher information* is the expected value of the observed information. Let $\mathbf{J}(\,.\mid .\,)$ denote the observed information matrices = the negative of the matrices of second partials in (6.5). Then (6.5) becomes

$$\begin{array}{c} \mathbf{J}(\mathbf{x}_t \mid D_t) \;\; = \qquad \mathbf{J}(\mathbf{y}_t \mid \mathbf{x}_t) \qquad + \mathbf{J}(\mathbf{x}_t \mid D_{t-1}) \\ \begin{pmatrix} \text{posterior} \\ \text{information} \end{pmatrix} = \begin{pmatrix} \text{information from} \\ \text{observation(s)} \end{pmatrix} + \begin{pmatrix} \text{prior} \\ \text{information} \end{pmatrix} \end{array}. \tag{6.6}$$

We find this result quite useful if only to help us remember and understand Kalman filtering / Bayesian sequential updating. In the normal case, this information is also the precision parameter(s), being the inverse of the variance (or covariance matrix). More generally, with non-normal observations, the information from observations could

fail to be nonnegative (or nonnegative definite). For example, consider the simple mixture of two normals, $y_t \sim 0.5\,N(\mu + 2,1) + 0.5\,N(\mu - 2,1)$. The likelihood is less for $\mu = y_t$ than for $\mu = y_t \pm 2$. By this fact and by the symmetry of this case, the second derivative of the log(likelihood) must be positive at $\mu = y_t$, which makes the observed information negative there. West and Harrison (1999, ch. 12) deal with this by running a mixture of dynamic models. For the present, we will assume that the information from observations is always nonnegative (or nonnegative definite).

As mentioned above, we assume that the prior and posterior are both adequately approximated by normal distributions, $N_p\left(\mathbf{x}_{t|t-1}, \mathbf{\Sigma}_{t|t-1}\right)$ and $N_p\left(\mathbf{x}_{t|t}, \mathbf{\Sigma}_{t|t}\right)$, so

$$l\left(\mathbf{x}_t \mid D_{t-1}\right) = c_1 - \frac{1}{2}\left(\mathbf{x}_t - \mathbf{x}_{t|t-1}\right)' \mathbf{\Sigma}_{t|t-1}^{-1}\left(\mathbf{x}_t - \mathbf{x}_{t|t-1}\right),$$

and

$$l\left(\mathbf{x}_t \mid D_t\right) = c - \frac{1}{2}\left(\mathbf{x}_t - \mathbf{x}_{t|t}\right)' \mathbf{\Sigma}_{t|t}^{-1}\left(\mathbf{x}_t - \mathbf{x}_{t|t}\right),$$

where c and c_1 are appropriate constants (relative to $\mathbf{x}_t$). From this, we get the following:

$$\begin{aligned} \frac{\partial l\left(\mathbf{x}_t \mid D_{t-1}\right)}{\partial \mathbf{x}_t} &= \left[-\mathbf{\Sigma}_{t|t-1}^{-1}\left(\mathbf{x}_t - \mathbf{x}_{t|t-1}\right)\right], \\ \frac{\partial l\left(\mathbf{x}_t \mid D_t\right)}{\partial \mathbf{x}_t} &= \left[-\mathbf{\Sigma}_{t|t}^{-1}\left(\mathbf{x}_t - \mathbf{x}_{t|t}\right)\right], \end{aligned} \tag{6.7}$$

$$\mathbf{J}\left(\mathbf{x}_t \mid D_{t-1}\right) = \left[-\frac{\partial^2 l\left(\mathbf{x}_t \mid D_{t-1}\right)}{\partial \mathbf{x}_t \partial \mathbf{x}_t'}\right] = \mathbf{\Sigma}_{t|t-1}^{-1},$$

and

$$\mathbf{J}\left(\mathbf{x}_t \mid D_t\right) = \left[-\frac{\partial^2 l\left(\mathbf{x}_t \mid D_t\right)}{\partial \mathbf{x}_t \partial \mathbf{x}_t'}\right] = \mathbf{\Sigma}_{t|t}^{-1}. \tag{6.8}$$

Substituting (6.8) into (6.6), we get

$$\mathbf{\Sigma}_{t|t}^{-1} = \mathbf{J}\left(\mathbf{y}_t \mid \mathbf{x}_t\right) + \mathbf{\Sigma}_{t|t-1}^{-1}. \tag{6.9}$$

Note, however, that this immediately exposes a problem: We assume that the prior is normal, which means that $\boldsymbol{\Sigma}_{t|t-1}^{-1}$ is assumed to be constant. However, unless the observation $\mathbf{y}_t$ is a linear function of the state $\mathbf{x}_t$ plus normal noise, the information $\mathbf{J}(\mathbf{y}_t \mid \mathbf{x}_t)$ might not be a constant. Therefore, if (6.9) is a strict equality, the posterior information $\boldsymbol{\Sigma}_{t|t}^{-1}$ might not be constant. Fortunately, $\mathbf{J}(\mathbf{y}_t \mid \mathbf{x}_t)$ tends not to vary much over the range of plausible values for $\mathbf{x}_t$. If $\mathbf{J}(\mathbf{y}_t \mid \mathbf{x}_t)$ is sufficiently well behaved, we approximately the log(posterior density) by a parabola, replacing $\mathbf{J}(\mathbf{y}_t \mid \mathbf{x}_t)$ by a constant. In these cases, the posterior mode occurs at $\mathbf{x}_{t|t}$. This suggests that we replace $\mathbf{J}(\mathbf{y}_t \mid \mathbf{x}_t)$ in (6.9) by $\mathbf{J}(\mathbf{y}_t \mid \mathbf{x}_{t|t})$. In many applications, the difference between $\mathbf{J}(\mathbf{y}_t \mid \mathbf{x}_{t|t-1})$ and $\mathbf{J}(\mathbf{y}_t \mid \mathbf{x}_{t|t})$ will be small, and we will often use $\mathbf{J}(\mathbf{y}_t \mid \mathbf{x}_{t|t-1})$ in place of $\mathbf{J}(\mathbf{y}_t \mid \mathbf{x}_t)$ in (6.9).

Expression (6.9) relates to a more general result, namely that the sampling distribution of maximum likelihood estimators is, under very general regularity conditions, approximately normal with covariance matrix being the inverse of the information (e.g., Rao 1973, or Skovgaard 1981). Thus, even with non-normal observations, $\mathbf{J}(\mathbf{y}_t \mid \mathbf{x}_t)$ generally acts like precision parameter(s), being the inverse of variance / covariance matrices. When the dimensionality of the state $\mathbf{x}_t$ exceeds that of the observations $\mathbf{y}_t$, the information $\mathbf{J}(\mathbf{y}_t \mid \mathbf{x}_t)$ will generally be singular but still nonnegative definite. With certain non-normal cases, the information may be negative for certain outcomes $\mathbf{y}_t$. However, these are generally rare, and when several observations are combined, this condition almost never persists. For the present discussion, we will ignore such cases.

Before proceeding, note that (6.6) and (6.9) expose a duality between Bayes' theorem and the addition / convolution of independent random variables: Recall that the variance of a sum of independent random variables is the sum of the variances. By contrast, when combining observations with prior information, the *information matrices*, not the variances, add. This is analogous to how resistances combine in series and parallel circuits in electronics: In series, the resistances add; in parallel, the conductances (being the reciprocal resistances) add.

To obtain the posterior mode, we let $\mathbf{x}_t = \mathbf{x}_{t|t}$ in (6.7) and substitute the result into (6.4) as follows:

$$0 = \left[\frac{\partial l\left(\mathbf{y}_t \mid \mathbf{x}_t = \mathbf{x}_{t|t}\right)}{\partial \mathbf{x}_t}\right] - \boldsymbol{\Sigma}_{t|t-1}^{-1}\left(\mathbf{x}_{t|t} - \mathbf{x}_{t|t-1}\right). \tag{6.10}$$

To solve this for $\mathbf{x}_{t|t}$, we expand $\partial l\left(\mathbf{y}_t, \mathbf{x}_t\right)/\partial \mathbf{x}_t$ about a point $\boldsymbol{\xi}$ using Taylor's theorem, as follows:

$$\left[\frac{\partial l\left(\mathbf{y}_t \mid \mathbf{x}_t = \mathbf{x}_{t|t}\right)}{\partial \mathbf{x}_t}\right] = \left[\frac{\partial l\left(\mathbf{y}_t \mid \mathbf{x}_t = \boldsymbol{\xi}\right)}{\partial \mathbf{x}_t}\right] - \mathbf{J}\left(\mathbf{y}_t, \mathbf{x}_t = \boldsymbol{\xi}\right)\left(\mathbf{x}_{t|t} - \boldsymbol{\xi}\right).$$

We substitute this into (6.10) to get the following:

$$0 = \left[\frac{\partial l\left(\mathbf{y}_t \mid \mathbf{x}_t = \boldsymbol{\xi}\right)}{\partial \mathbf{x}_t}\right] - \mathbf{J}\left(\mathbf{y}_t, \mathbf{x}_t = \boldsymbol{\xi}\right)\left(\mathbf{x}_{t|t} - \boldsymbol{\xi}\right) - \boldsymbol{\Sigma}_{t|t-1}^{-1}\left(\mathbf{x}_{t|t} - \mathbf{x}_{t|t-1}\right).$$

The coefficient of $\mathbf{x}_{t|t}$ in this expression can be written using (6.9) as follows:

$$\boldsymbol{\Sigma}_{t|t,\xi}^{-1} = \mathbf{J}\left(\mathbf{y}_t \mid \mathbf{x}_t = \boldsymbol{\xi}\right) + \boldsymbol{\Sigma}_{t|t-1}^{-1}. \tag{6.11}$$

We use this to rearrange the previous expression as follows:

$$\boldsymbol{\Sigma}_{t|t,\xi}^{-1}\mathbf{x}_{t|t} = \left[\frac{\partial l\left(\mathbf{y}_t \mid \mathbf{x}_t = \boldsymbol{\xi}\right)}{\partial \mathbf{x}_t}\right] + \mathbf{J}\left(\mathbf{y}_t \mid \mathbf{x}_t = \boldsymbol{\xi}\right)\boldsymbol{\xi} + \boldsymbol{\Sigma}_{t|t-1}^{-1}\mathbf{x}_{t|t-1}. \tag{6.12}$$

We now rewrite (6.11) as $\boldsymbol{\Sigma}_{t|t-1}^{-1} = \boldsymbol{\Sigma}_{t|t,\xi}^{-1} - \mathbf{J}(\mathbf{y}_t \mid \mathbf{x}_t = \xi)$ and substitute into (6.12) to get

$$\boldsymbol{\Sigma}_{t|t,\xi}^{-1}(\mathbf{x}_{t|t} - \mathbf{x}_{t|t-1}) = \left[\frac{\partial l(\mathbf{y}_t \mid \mathbf{x}_t = \xi)}{\partial \mathbf{x}_t}\right] + \mathbf{J}(\mathbf{y}_t \mid \mathbf{x}_t = \xi)(\xi - \mathbf{x}_{t|t-1}).$$

Next, we solve this for $\mathbf{x}_{t|t}$ to obtain the following:

$$\mathbf{x}_{t|t} = \mathbf{x}_{t|t-1} + \boldsymbol{\Sigma}_{t|t,\xi}\left\{\left[\frac{\partial l(\mathbf{y}_t \mid \mathbf{x}_t = \xi)}{\partial \mathbf{x}_t}\right] + \mathbf{J}(\mathbf{y}_t \mid \mathbf{x}_t = \xi)(\xi - \mathbf{x}_{t|t-1})\right\}. \tag{6.13}$$

This expression can now be used to define an iteration. For the first cycle, we let $\xi = \mathbf{x}_{t|t-1}$; with this choice for ξ, (6.13) simplifies to the following:

$$\mathbf{x}_{t|t} = \mathbf{x}_{t|t-1} + \boldsymbol{\Sigma}_{t|t}\left[\frac{\partial l(\mathbf{y}_t \mid \mathbf{x}_t = \mathbf{x}_{t-1})}{\partial \mathbf{x}_t}\right]. \tag{6.14}$$

For most applications, we will stop after one iteration, relying on future observations to correct any suboptimality created by stopping too early.

We now specialize this to the particular observation process invoked in (6.3), for which

$$l(\mathbf{y}_t \mid \mathbf{x}_t) = c_y - \frac{1}{2}(\mathbf{y}_t - \mathbf{h}_t)' \mathbf{V}_t^{-1}(\mathbf{y}_t - \mathbf{h}_t),$$

where $\mathbf{h}_t = \mathbf{h}_t(\mathbf{x}_t, \mathbf{u}_t)$. Then

$$\frac{\partial l(\mathbf{y}_t \mid \mathbf{x}_t)}{\partial \mathbf{x}_t} = \mathbf{h}_{t,1}' \mathbf{V}_t^{-1}(\mathbf{y}_t - \mathbf{h}_t),$$

where (6.15)

$$\mathbf{h}_{t,1} = \frac{\partial \mathbf{h}_t(\mathbf{x}_t, \mathbf{u}_t)}{\partial \mathbf{x}_t'}$$

is the indicated $k \times p$ matrix of partial derivatives; unless otherwise specified, $\mathbf{h}_{t,1}$ will be evaluated at $\mathbf{x}_t = \mathbf{x}_{t|t-1}$. Taking second derivatives while assuming that the curvature of $\mathbf{h}_{t,1}$ is negligible gives us

$$\mathbf{J}(\mathbf{y}_t \mid \mathbf{x}_t) = \mathbf{h}'_{t,1}\mathbf{V}_t^{-1}\mathbf{h}_{t,1}. \tag{6.16}$$

In the work reported in this section, $\mathbf{h}_{t,1}$ and $\mathbf{J}(\mathbf{y}_t \mid \mathbf{x}_t)$ were always evaluated at $\mathbf{x}_t = \mathbf{x}_{t|t-1}$.

This completes all the pieces required for "step 1. Observation" in Bayesian sequential updating. Organizing this material into substeps as outlined in Figure 3.3 above will be left as an exercise for the reader.

Step 2. Transition. Given the posterior $(\mathbf{x}_t \mid D_t) \sim N_p(\mathbf{x}_{t|t}, \mathbf{\Sigma}_{t|t})$, after the transition (6.1), the prior at time $(t + 1)$ is given by $(\mathbf{x}_{t+1} \mid D_t) \sim N_p(\mathbf{x}_{t+1|t}, \mathbf{\Sigma}_{t+1|t})$, where using Taylor's theorem we get the following:

$$\mathbf{x}_{t+1|t} = \mathbf{f}_t(\mathbf{x}_{t|t}, \mathbf{u}_t)$$

and

$$\mathbf{\Sigma}_{t+1|t} = \mathbf{f}_{t,1}\mathbf{\Sigma}_{t+1|t}\mathbf{f}'_{t,1} + \mathbf{W}_t$$

where

$$\mathbf{f}_{t,1} = \frac{\partial \mathbf{f}_t(\mathbf{x}_t = \mathbf{x}_{t|t}, \mathbf{u}_t)}{\partial \mathbf{x}_t}.$$

We now have all the pieces required for Bayesian sequential updating, as outlined in Figure 3.3 above. We note also that this work can be further generalized to estimate an EWMA for variance, analogous to section 4 above, through the introduction of a scalar relative precision parameter in the obvious places.

REFERENCES

Belsley, D. A., Kuh, E. and Welsch, R. E. (1980) *Regression Diagnostics*, (NY: Wiley).

Brown, R. G, and Hwang, P. Y. C. (1997) *Introduction to Random Signals and Applied Kalman Filtering* (NY: Wiley).

Eide, P., and Maybeck, P. (1996) "An MMAE Failure Detection System for the F-16", *IEEE Transactions on Aerospace and Electronic Systems*, 32: 1125-1136.

Gnedenko, B. V., and Kolmogorov, A. N. (1968) *Limit theorems for Sums of Independent Random Variables*, revised (Reading, MA: Addison-Wesley).

Menke, T. E., and Maybeck, P. S. (1996) "Sensor / Actuator Failure Detection in the Vista F-16 by Multiple Model Adaptive Estimation", *IEEE Transactions on Aerospace and Electronic Systems*, 31: 1218-1229.

Pole, A., West, M., and Harrison, H. (1994) *Applied Bayesian Forecasting and Time Series Analysis* (NY: Chapman & Hall)

Rao, C. R. (1973) *Linear Statistical Inference and Its Applications*, 2nd ed. (NY: Wiley).

Skovgaard, I. M. (1981) "Edgeworth Expansions of the Distribution of the Maximum Lieklihood Estimators in the General (non i.i.d.) Case", *Scandinavian Journal of Statistics*, 8, 227-236.

_____ (1986) "On Multivariate Edgeworth Expansions", *International Statistical Review*, 54: 29-32.

West, M. and Harrison, P. J. (1999) *Bayesian Forecasting and Dynamic Models*, 2nd ed., corrected 2nd printing (NY: Springer)

7. MULTIPLE MODEL ADAPTIVE ESTIMATION (MMAE)

In this section, we discuss the use of multiple Kalman filters, each designed to detect a different set of failure modes in a common plant. We consider in particular the air intake system sketched in Figure 6.1. In that particular application, we found that Kalman filters designed to estimate sensitivity parameters for one or two sensors could adapt appropriately and could therefore detect and isolate certain simulated faults.

However, a Kalman filter designed to estimate sensitivity parameters for three sensors simultaneously just got lost. Two possible sources for this problem were identified. First, Kalman filtering is in essence a least-squares regression, performed one observation at a time, gracefully discounting older data that is less relevant because the regression parameters (state of the plant) have presumably migrated over time. In this context, the concept of leverage in regression is important: Physical laws of fluid flow combined with automatic controls might make it practically impossible to estimate this many parameters. Some of this lack of leverage for regression might be overcome by dithering the controls. However, this might make the vehicle cough and shudder. This would clearly be unacceptable in routine use. It might be allowable in a special diagnostic mode induced by an automotive repair technician. However, since that might have an emissions impact, governmental regulators and others concerned with air quality might object.

A second possible source of computational difficulties might be numerical precision: The components of the Kalman state vector were not rescaled to roughly mean zero and standard deviation of 1 under routine operations, nor did we necessarily take great care to ensure that all computations used numerically stable techniques such as

Cholesky decompositions. If we had done so, we doubtless would have gotten better monitors for the air intake system of Figure 6.1, although we do not know if the filter estimating three pairs of sensitivity parameters would have performed adequately even then. Moreover, the difficulties we encountered might still be a problem with in other, more subtle applications or when trying to implement the solutions found there on 8- or 16-bit microprocessors used for control and on-board diagnostics on modern vehicles. A more robust methodology is needed.

Fortunately, a more robust methodology is available in the form of multiple model adaptive estimation (MMAE); see, e.g., Menke and Maybeck (1995) or Eide and Maybeck (1996). This in essence consists of running multiple Kalman filters in parallel, each designed to model a particular fault, and using Bayes' theorem to select the most likely failure mode. In applications where some time may be permitted to pass between fault detection and isolation, we recommend separating detection from isolation. Under normal operations, a Kalman filter of a good plant only, with no components in the state vector for sensitivity parameters, will predict the observations with a certain amount of error. At the onset of a fault, the prediction error will increase. (If a fault does not increase prediction error, it should exhibit some other symptom detectable by an appropriate monitor.) With evidence of an increase in prediction error, the on-board diagnostics then start running a bank of monitors simultaneously, each one designed to detect a different fault or combination of faults. In section 7.1, we discuss fault detection on the basis of an increase in prediction error. In section 7.2, we discuss fault isolation using multiple Kalman filters. Section 7.3 further considers issues of designing either computer or physical experiments to tune parameters of the Kalman filter and residual

Cusum monitors. Section 7.4 briefly reviews questions related to the numerical stability of algorithms. Concluding remarks appear in section 7.5.

7.1. Fault Detection

Suppose now we have a model that predicts observations as

$$(\mathbf{y}_t \mid D_{t-1}) \sim N_k(\, \mathbf{h}_{t,0},\ \boldsymbol{\Sigma}_{y|t-1}), \tag{7.1}$$

where predictions are obtained some way, e.g., by a Kalman filter as discussed in (5.4) above or the generalization in the appendix to section 6. Suppose further that when the plant malfunctions, the observation errors increase, e.g., as

$$(\mathbf{y}_t \mid D_{t-1}) \sim N_k(\, \mathbf{h}_{t,0},\ \rho\, \boldsymbol{\Sigma}_{y|t-1}), \tag{7.2}$$

where ρ is a variance inflation factor indicating a loss of prediction ability. We can monitor for an increase in ρ in two ways: One is to via an EWMA for a univariate variance factor as described in section 4 above; the theory of section 4 can be combined naturally with the theory of sections 5 and 6 to produce Kalman filters that are accompanied by a simultaneous EWMA for variance, as discussed by West and Harrison (1999) and Pole, West and Harrison (1994). Another is via a Cusum of log(likelihood ratio) for an increase in ρ from 1 per (7.1) to something larger in (7.2), or to from a worst acceptable $\rho_0 > 1$ to a best unacceptable $\rho_1 > \rho_0$. This Cusum could be a Bayes-adjusted Cusum, as described in section 2 above. In that case, it could potentially incorporate reliability information, as discussed there, modeling the increased failure rate anticipated from older plants. With good data on an increasing hazard rate, that information could be used to improve reliability and availability at lower total cost, as

suggested by the discussion in section 2 above. Automobile manufacturers may be reluctant to use reliability information in this way to avoid charges of manipulating on-board diagnostics to increase sales of replacement parts, unless its use were encouraged by regulators.

Whether using an EWMA for variance or a Cusum of log(likelihood ratio), we would need to estimate ρ in (7.2) for both worst acceptable and best unacceptable plants. For this purpose, we use the maximum likelihood estimate for ρ, which can be derived as follows. The probability density corresponding to (7.2) is

$$f(\mathbf{y}_t \mid D_{t-1}) = \frac{1}{(2\pi)^{k/2}\left|\rho\,\boldsymbol{\Sigma}_{y|t-1}\right|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{y}_t - \mathbf{h}_{t,0})'(\rho\,\boldsymbol{\Sigma}_{y|t-1})^{-1}(\mathbf{y}_t - \mathbf{h}_{t,0})\right\}, \quad (7.3)$$

so the log(likelihood) is

$$l(\mathbf{y}_t \mid D_{t-1}) = \left(-\frac{1}{2}\right)\left\{k\ln(2\pi\rho) + \ln\left|\boldsymbol{\Sigma}_{y|t-1}\right| + \rho^{-1}(\mathbf{y}_t - \mathbf{h}_{t,0})'\boldsymbol{\Sigma}_{y|t-1}^{-1}(\mathbf{y}_t - \mathbf{h}_{t,0})\right\}, \quad (7.4)$$

With N observations, we get

$$\sum_{t=1}^{N} l(\mathbf{y}_t \mid D_{t-1}) = Nc - \frac{1}{2}\sum_{t=1}^{N}\left\{k\ln(\rho) + \rho^{-1}(\mathbf{y}_t - \mathbf{h}_{t,0})'\boldsymbol{\Sigma}_{y|t-1}^{-1}(\mathbf{y}_t - \mathbf{h}_{t,0})\right\},$$

with $c = (-0.5)\left\{k\ln(2\pi) + \ln\left|\boldsymbol{\Sigma}_{y|t-1}\right|\right\}$. By differentiating this with respect to ρ^{-1} and setting the result to 0, we find that this likelihood is maximized over ρ at

$$\hat{\rho} = \frac{1}{kN}\sum_{t=1}^{N}(\mathbf{y}_t - \mathbf{h}_{t,0})'\boldsymbol{\Sigma}_{y|t-1}^{-1}(\mathbf{y}_t - \mathbf{h}_{t,0}). \quad (7.5)$$

This relative variance parameter $\hat{\rho}$ was estimated for 13 different Kalman filters for the air intake system of Figure 6.1 under 9 different good and malfunctioning scenarios. The 13 different filters are described in Table 7.1, and the 9 scenarios are

outlined in Table 7.2. The relative variances $\hat{\rho}$ for these 13 × 9 filter and scenario combinations are presented in Table 7.3. In this table, we see that $\hat{\rho}$ performed largely as expected except for the six 8-d full adaptive filters: If the filter was able to adapt to the particular simulated fault, it did so, and $\hat{\rho}$ was relatively small. Otherwise, it was larger.

Table 7.1. Thirteen Alternative Kalman Filter Models for the Air Intake System

Code	Dimension of State Vector	Components of State Vector	Adaptive Migration Variances	
			intercept	slope
Good	2	(MAF, MAP)	NA	NA
4-d MAF	4	(MAF, MAP) + (intercept, slope) for MAF	0.001	10^{-7}
4-d MAP	4	(MAF, MAP) + (intercept, slope) for MAP	0.001	10^{-7}
4-d TPS	4	(MAF, MAP) + (intercept, slope) for TPS	0.001	10^{-7}
6-d w/o MAF	6	(MAF, MAP) + (intercept, slope) for MAP, TPS	0.001	10^{-7}
6-d w/o MAP	6	(MAF, MAP) + (intercept, slope) for MAF, TPS	0.001	10^{-7}
6-d w/0 TPS	6	(MAF, MAP) + (intercept, slope) for MAF, MAP	0.001	10^{-7}
8-d 00	8	(MAF, MAP) + (intercept, slope) for MAF,	5×10^{-6}	10^{-8}
8-d 01	8	MAP, TPS	5×10^{-6}	10^{-6}
8-d 10	8	"	5×10^{-4}	10^{-8}
8-d 11	8	"	5×10^{-4}	10^{-6}
8-d cp	8	"	5×10^{-5}	10^{-7}
8-d 0b	8	"	5×10^{-6}	10^{-5}

Table 7.2. Faulty Plants Simulated

Table 7.2.1. Four Problems to Simulate: Best Unacceptable Levels(+)

	Intercept	Slope
1. Constant Plausible	~ mean(*)	0
2. Dull	0	0.6
3. Hyper	0	1.4
4. Offset	20	1

(+) Worst Acceptable was selected as half of the difference from the nominal as the best unacceptable, i.e., 10 for the intercept and 0.8 and 1.2 for the slope.

(*) The intercepts for constant plausible and offset were roughly equal to the mean of the observations, rounded off grossly; for some sensors, the intercept was not the same for "constant plausible" and "offset".

Table 7.2.2. Eight Scenarios Simulated

000 = no fault
100 = MAF constant
400 = MAF offset
020 = 0.6(MAP) dull
030 = 1.4(MAP) hyper
003 = 1.4(TPS) hyper
402 = (MAF offset) + 0.8(TPS) dull
341 = 1.4(MAF) hyper + (MAP offset) + (TPS constant)

Codes = XYZ for (MAF, MAP, TPS)

where X, Y, or Z = 0 for good or 1-4 for one of the faults listed in Table 7.2.1

Table 7.3. Estimated Relative Variance(+) $\hat{\rho}$

Kalman Filter	Good (000)	MAF constant(*) (100)	MAF offset (400)	MAP dull (020)	MAP hyper (030)	TPS hyper (003)	dual fault (402)	triple fault (341)
2-d Good	*2.06*	**30.69**	**87.20**	**37.23**	**47.01**	**18.57**	**138.65**	**138.72**
4-d MAF	*1.74*	*2.14?*	*1.61*	**36.84**	**42.57**	**11.40**	**9.38**	**75.67**
4-d MAP	*1.42*	**27.05**	**81.03**	*1.10*	*2.05*	**7.12**	**125.72**	**84.11**
4-d TPS	*0.71*	**18.28**	**47.78**	**17.78**	**16.86**	*0.66*	**47.45**	**3.81**
6-d w/o MAF	*0.64*	**19.04**	**150.87**	*0.61*	*0.92*	*0.60*	**57.98**	**12.72**
6-d w/o MAP	*0.54*	*5.24?*	*0.62*	**6.53**	**10.78**	0.49	0.64	**6.58**
6-d w/o TPS	*1.14*	*3.78?*	*1.09*	*0.81*	*1.91*	**5.52**	**3.42**	**233.37**
8-d 00	*0.56*	*27.99?*		*5.04*		*1.01*	*5.69*	
8-d 01	*0.41*	*9.41?*		*0.35*		*0.57*	*1.71*	
8-d 10	*0.42*	*14.92?*	*(not*	*1.28*	*(not*	*0.48*	*0.68*	*(not*
8-d 11	*0.34*	*0.21?*	*run)*	*0.29*	*run)*	*0.41*	*0.30*	*run)*
8-d cp	*0.47*	*14.34?*		*0.77*		*0.66*	*1.90*	
8-d 0b	*0.29*	*3.16?*		*0.26*		*0.33*	*1.44*	

(+) All numbers were estimated from the last 500 observations of a 2,000 observation (200 second) segment, with any simulated faults commencing between observation 501 and 1,000.

italics = Should be small: Kalman filter estimates the discrepant sensitivities

Bold = Should be large: Kalman filter does not estimate the discrepant sensitivities

(*) A question mark (?) is appended to certain numbers in the "MAF constant" column, because none of the Kalman filters model well a dead sensor; see text.

The entries in Table 7.3 can be separated into three categories:

Filter Estimating a Constant: None of the Kalman filters considered were designed to handle well the case of a dead sensor, as indicated in the "MAF Constant" column. However, some of the sensors estimated MAF sensitivity parameters, and therefore might perform better than the others for this particular simulated fault; a question mark "?" was appended to the estimated relative variance $\hat{\rho}$ for those cases.

Should be small: The indicated Kalman filter should be able to adapt to a fault of the nature indicated, leaving a small value for the estimated relative variance $\hat{\rho}$. The largest number in this category in the table is 2.06, ignoring the "MAF constant" column.

Should be large: For many of the combinations in Table 7.3, the estimated relative variance should be large because the Kalman filter for that row of the table was not designed to be able to compensate for a fault like that simulated. The smallest number in this category in the table is 3.42.

Our initial idea in approaching this problem was that the 8-d full adaptive filter should be able to isolate all but the constant sensor faults in this table. Unfortunately, our early tests revealed a number of cases with unacceptably large prediction errors and $\hat{\rho}$, e.g., "8-d 00" for MAP dull (030). To overcome this difficulty, we performed designed experiments with Kalman filter model parameters such as the $2\times 2+(\text{center point})$ outlined in five of the last six rows of Table 7.1. The results of these tests, summarized in Table 7.3, suggested that further improvement in performance might be obtained in the general direction of the bottom row there, labeled "8-d 0b". Unfortunately, in all these cases, when the noise variance was acceptable, the values of the estimated sensitivity parameters *could not reliably isolate simulated faults*.

This led us to reconsider the work of Menke and Maybeck (1995) and Eide and Maybeck (1996), who suggested running multiple Kalman filters in parallel; we had earlier hoped we could avoid the added complexity of this approach. To explore this further, we produced the first seven rows of Table 7.3. These results suggest that we

should be able to get a working monitor by separating fault detection from isolation: We can track the noise variance ρ from the "2-d Good" filter to detect a fault, and then isolate it by tracking it from the 4- and 6-d monitors: In the presence of a fault, the relative variance ρ of the "2-d Good" filter explodes, at least in the limited cases considered in Table 7.3. After fault detection, we look among candidate 4-d filters designed to adapt to a single defective sensor. If none of the EWMAs or 4-d filters have small prediction error, we next look at the 6-d filters, which should be able to adapt to certain double faults. If one is found with small prediction error, that isolates the fault. Otherwise, the problem must be something else not considered, e.g., a triple fault.

The poorest separation between good and bad in Table 7.3 appeared with the "MAF constant" scenario. As noted above, none of these Kalman filters could necessarily be expected to model well a dead sensor. Therefore, to deal with this issue, we may need to add EWMAs to detect dead sensors from low prediction error.

Menke and Maybeck (1995) and Eide and Maybeck (1996) suggest computing a Bayesian posterior distribution for each alternative model, with two modifications: They put a floor of 0.001 under the posterior probability for any of the multiple models, and they deleted the determinant in the denominator of the density in (7.3). In our approach, by explicitly including "Step 2. Transition" in the update cycle, we automatically get a floor under the posterior probability for a particular fault. We would have to study the issue more carefully before commenting on deleting the determinant from (7.3) before using the modified density in Bayes' theorem to track the probability of alternative models.

However, rather than attempting that, we thought it might be just as simple to run a Cusum on the log(likelihood ratio) to detect a jump in the relative variance parameter ρ in (7.3). From analyses such as Table 7.3 using a different data set, we selected a "worst acceptable" $\rho_0 = 5.02$, and a "best unacceptable" $\rho_1 = 7.14$. By taking the difference between the log(likelihood) in (7.4) for $\rho = \rho_1$ and $\rho = \rho_0$, we get the log(likelihood ratio) as follows:

$$\text{log(likelihood ratio)} = \frac{1}{2}\left\{k\ln(\rho_0/\rho_1) + \left(\rho_0^{-1} - \rho_1^{-1}\right)\mathbf{e}_t'\boldsymbol{\Sigma}_{y|t-1}^{-1}\mathbf{e}_t\right\} = \frac{1}{2}\left(\frac{\rho_1 - \rho_0}{\rho_0\rho_1}\right)\tilde{\Lambda}_t,$$

where with $\mathbf{e}_t = (\mathbf{y}_t - \mathbf{h}_{t,0})$,

$$\tilde{\Lambda}_t = \left(\mathbf{e}_t'\boldsymbol{\Sigma}_{y|t-1}^{-1}\mathbf{e}_t\right) - \Lambda^*,$$

and (7.6)

$$\Lambda^* = k\left(\frac{\rho_0\rho_1}{\rho_1 - \rho_0}\right)\ln\left(\frac{\rho_1}{\rho_0}\right).$$

With $\rho_0 = 5.02$, and $\rho_1 = 7.14$, we get $\Lambda^* = 11.9$.

To help us understand (7.6), suppose that $\rho_1 = \rho_0(1+\delta)$, with δ small. Then

$$\Lambda^* = k\rho_0\left(\frac{1+\delta}{\delta}\right)\ln(1+\delta).$$

Using $\ln(1+\delta) = \delta - 0.5\delta^2 + O(\delta^3)$, we get $\Lambda^* = k\rho_0\{1 + 0.5\delta + O(\delta^2)\}$. But $\rho_0(1+0.5\delta) = (\rho_0 + \rho_1)/2 = \bar{\rho}$, say. Thus, $\Lambda^* = k\bar{\rho} + O(\delta^2)$. But $k\rho_0$ is the expected value of the quadratic form $\left(\mathbf{e}_t'\boldsymbol{\Sigma}_{y|t-1}^{-1}\mathbf{e}_t\right)$ when the plant is good, and $k\bar{\rho}$ is halfway between this expectation for good and bad plants. While Λ^* differs from this when the difference between ρ_0 and ρ_1 is large, this analysis helps sharpen our intuition about the behavior of $\tilde{\Lambda}_t$ in (7.6).

With $\Lambda^* = 11.9$, the 2-d and 4-d monitors in Table 7.1 were applied to a different segment of 2,000 observations, invoking the good and faulty scenarios listed in Table 7.2. Table 7.4 lists the maximum in these 2,000 observations achieved by

$$Q_t^\sigma = \max\left\{0,\ Q_{t-1}^\sigma + \tilde{\Lambda}_t\right\}, \tag{7.7}$$

starting with $Q_0^\sigma = 0$.

Table 7.4. Max Cusum$^{(+)}$ Q_t^σ

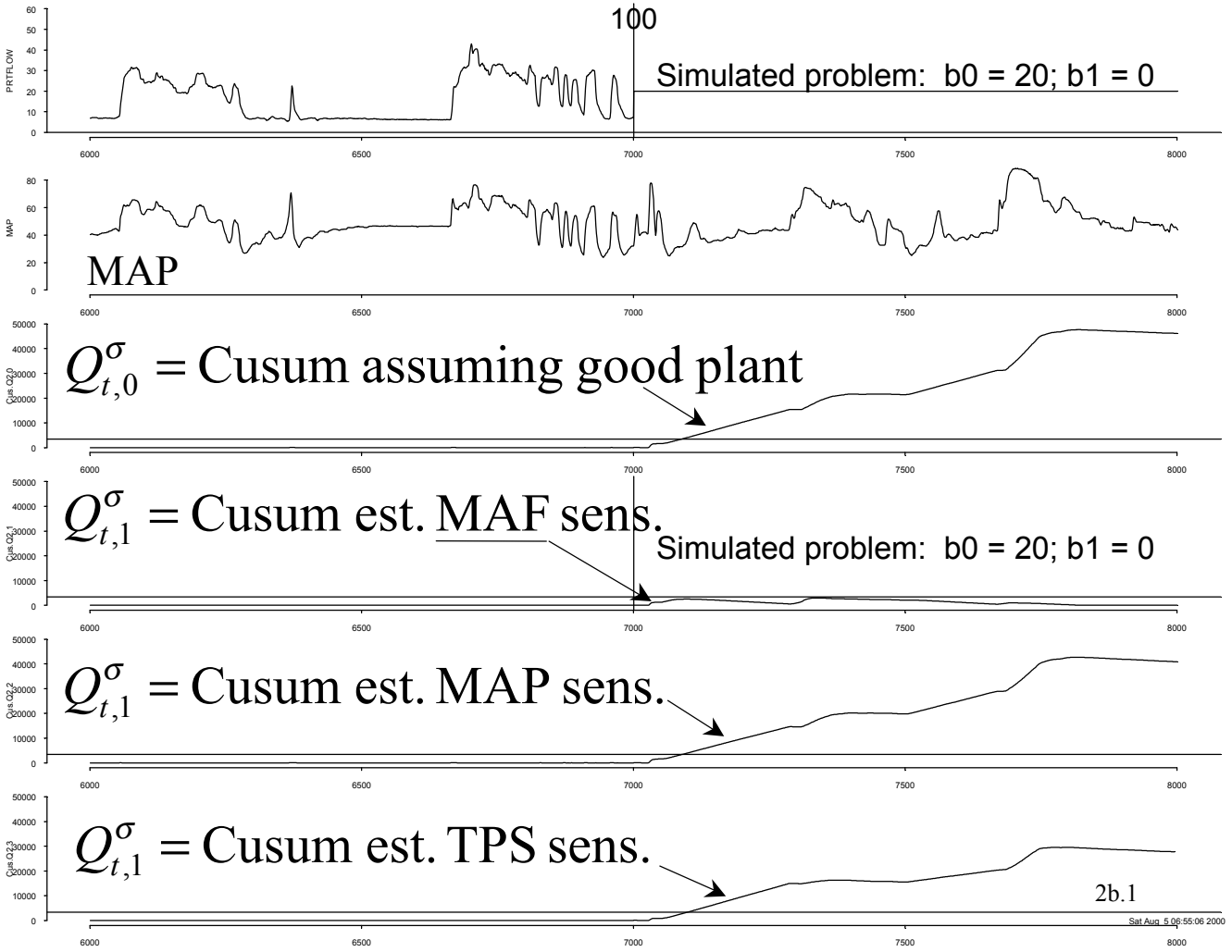
		Kalman Filter			
Scenario	(code)	2-d Good	4-d MAF	4-d MAP	4-d TPS
Good	(000)	*368*	*236*	*109*	*48*
MAF constant$^{(*)}$	(100)	**47,656**	*2,927?*	**42,715**	**29,544**
MAF offset	(400)	**157,931**	*3,277*	**149,242**	**83,814**
MAP dull	(020)	**44,279**	**43,886**	*1,423*	**17,389**
MAP hyper	(030)	**62,805**	**57,746**	*2,527*	**17,201**
TPS hyper	(003)	**26,058**	**12,302**	**3,513**	*174*
dual fault	(402)	**302,041**	**9,287**	**286,279**	**106,538**
triple fault	(341)	**221,664**	**118,924**	**132,708**	**11,778**

(+) Maximum over 2,000 observations (different from the 2,000 used for Table 7.3)
italics = Should be small: Kalman filter estimates the discrepant sensitivities
Bold = Should be large: Kalman filter does not estimate the discrepant sensitivities
(*) None of these monitors are designed to handle well a dead sensor such as "MAF constant". We have therefore appended a question mark (*?*) to each of the potentially small numbers in the "MAF constant" row.

The numbers in Table 7.4 display what we would expect: When the filter is capable of adapting to the fault, the max Cusum is relatively low; otherwise, it is larger. Some experimentation with the migration rate parameters, as for the 3-d filter in Table 7.1, might increase the separation between scenario-filter combinations for which we do and do not want a fault to be detected. Note also that the max Cusums for the "Good" scenarios are substantially below the max Cusums for all the fault scenarios. This is, again, as we would expect: We simulated a sudden fault onset. When a Kalman filter

first encounters a sudden fault onset, the prediction errors will be large. For a filter that adapts to the fault, the Cusum will tend to rise until the filter has adapted sufficiently; after that, the Cusum will drift back towards 0. For a filter that can not adapt to the fault, the Cusum will generally continue trending upwards towards the stratosphere. This is, indeed, the image we get from most of Table 7.4 and from plots of the Cusums such as Figure 7.1. The primary exception to this rule appears to be the response of the MAP adaptive filter to a hyper TPS. This exception can be partially explained by the fact that the TPS is much less important than the MAF and MAP sensors for the control of the plant of Figure 6.1. However, this case deserves more careful study to try to understand why the response is not more pronounced and what would likely happen if it had more time to respond.

Figure 7.1. Mass Air Flow, Manifold Absolute Pressure, and Cusums for Variance for Four Models



Based on the numbers of Table 7.4, we selected for illustration a threshold of 3,400, between the worst acceptable figure of 3,277 and the best unacceptable figure of 3,513. We would use this as an initial threshold, to be adjusted to balance the probabilities of a false alarm in the design life of the plant and of an excessive delay to detecting a real fault, as discussed in Box et al. (2000). Figure 7.1 plots MAF, MAP and Q_t^σ of (7.7) for the constant MAF scenario (code 100) and for the 4 Kalman filters considered in Table 7.4. In this figure, at fault onset, Q_t^σ for all 4 filters started on a general upward trend. Almost immediately, the MAF adaptive filter compensated for the problem, and Q_t^σ began to decline for that filter; meanwhile, Q_t^σ for the other three Kalman filters continued climbing. Thus, in this example, we had rapid fault detection with the 2-d filter assuming a "good" plant. This was combined with successful fault isolation by comparing levels of Q_t^σ for the filters designed to adapt to single faults in MAF, MAP and TPS, respectively.

If we need more separation between good and bad than what we see in Figure 7.1, we might note that the worst "good" $\hat{\rho}$ in Table 7.3 comes with "MAF Constant". If we got acceptable results from selecting Λ^* ignoring this case, we could add a separate diagnostic for this problem, e.g., an EWMA for mean and variance, triggering when the EWMA for variance dropped below a certain threshold.

7.3. Designing Experiments for Monitor Tuning and Evaluation

Monitors work by looking for discrepancies between observations and predictions from a model. Ideally, we would like to have models for both good and malfunctioning

plants, as explained by Box et al. (2000): If the prediction error is low for one filter and high for another, we can usually conclude that the filter with low prediction error better models plant behavior.

If a monitor is not adequate, in general the greatest opportunities for improvement lie in improving the models upon which the monitor is built. Designed experiments can be extremely valuable for organizing and reducing the amount of effort required to build better models and for improving the quality of information gained from a given effort. Basic principles for designing such experiments are discussed elsewhere (e.g., Box, Hunter, and Hunter 1978, and Box and Draper 1987) and will not be reviewed here.

Ignoring for the moment the "constant plausible" (dead sensor) fault, a standard response surface approach might be used to evaluate the response of a monitor to "dull", "hyper" and "offset" problems mentioned in Table 7.2.1. This would allow us to evaluate monitor response to "worst acceptable" levels of 0.8 and 1.2 for "dull" and "hyper" slopes by parabolic interpolation between the "best unacceptable" levels of 0.6 and 1.4 and the "good" slope of 1, without ever testing the "worst acceptable" levels directly. With slope and intercept for each of the three sensors of this example, MAF, MAP, and TPS, this would give us a total of 6 factors. As noted by Box and Draper (1986, Table 15.10), this would require a minimum of 28 runs just to estimate the intercept, the 6 main effects, the 15 two-factor interactions and the 6 pure quadratics, if we knew in advance that there was zero measurement and replication error.

Much of this work could be done by computer experiments, randomly sampling different segments of data from physical tests of different vehicles and simulating the proposed fault combination. This kind of computer experiment might process segments

of data nested within vehicles as random effects. These random effects in turn would require extra care in evaluation beyond the standard ordinary least squares regression.

If computer time were sufficiently cheap, we might recommend estimating a complete set of parameters from each segment of data, using the same number of data segments sampled more or less at random from each of several different vehicles (balancing to the extent feasible an appropriate range of operating conditions). A primary response variable in these analyses might be the logarithm of the mean square residuals $\hat{\rho}$ from the Kalman filter predictions, as in Table 7.3 above.

Models of $\log(\hat{\rho})$ might then be used to select "worst acceptable" and "best unacceptable" levels, ρ_0 and ρ_1, as discussed with Table 7.3, but based on the regression fit from the data segment - vehicle combinations processed. These could then be used to determine a Cusum centering coefficient, Λ^* of (7.6). Further experiments might explore the sensitivity of the max Cusum of Table 7.4 to variations in Λ^*. In this analysis, the logarithm of the max Cusum might be the response variable. The computations required to explore variations in Λ^* could potentially reprocess Kalman filter residuals stored from earlier tests, thereby possibly reducing the computer time required for these evaluations.

Beyond this testing, we would need to plan separate tests for "constant plausible" faults, mentioned in Table 7.2.1, some of which may need to be combined with other faults with other sensors. For the preparation of this report, the allotted time did not permit us to run the response surface experiments just suggested. We therefore selected the "typical" fault combinations considered in Tables 7.2 - 7.4. Experiments with

"constant plausible" conditions in at least one of the three sensors could then be designed to explore monitor response to these faults in plausible combination with other faults, producing a list of test conditions similar to those discussed here.

After prototype monitor design is completed including selection of Λ^*, we would still require final confirmation in prototype units, possibly employing the concepts of "accelerated testing for on-board diagnostics" discussed by Bisgaard et al. (2001).

7.4. Numerical Stability of Computations

A number of techniques were tried to improve the performance of the 8-d Kalman filter estimating the sensitivity parameters for three sensors. These included the following:

- *Program for numerical stability*: We used some of the features in S-PLUS to attempt to preserve numerical precision on the matrix inversions. Nothing we tried fixed the problem. We did not, however, rescale all components of the state vector to roughly mean zero and standard deviation of one, nor did we meticulously implement the complete numerically stable Kalman filtering algorithms of Bierman (1977).
- *Control step size*: We placed limits on the maximum step size. In some cases, with inappropriate implementation of the algorithm, especially with inputs outside the range of that considered in the nonlinear transfer functions programmed to model the air intake

system of Figure 6.1, we got very large steps. We programmed the algorithm to place a limit on the magnitude of a change between one observation and the next. We did not test this idea extensively, but our preliminary efforts in this direction did not obviously improve the performance of the monitor.

- *Check for outliers*: We checked the prediction errors for normality. We found the pattern of a contaminated normal: A normal probability plot (with data on the horizontal axis) showing a steep section in the middle (small standard deviation) and points with a common, shallower slope (higher standard deviation) in the tails (Titterington, Smith, and Makov 1985). We were able to explain roughly half of these outliers and modify the algorithm to eliminate them. This improved the performance of all the Kalman filters slightly, but did not substantially improve the performance of any of them.

There are other techniques that we did not try but that may improve the numerical precision of Kalman filtering computations:

- *Rescale*: In computations of this type, especially using 8- and 16-bit microprocessors, it is generally wise to rescale all variables so they have very roughly mean 0 and standard deviation of 1. We recently had an experience computing 100 x 100 covariance matrices from thousands of observations in which rescaling made the difference between numerical stability and instability: Without rescaling,

virtually all our covariance matrices were numerically singular. With rescaling, this problem completely disappeared.

- *Limit the use of estimated correlations in the covariance matrices for measurements*, $\mathbf{V}_t$, *and migration*, $\mathbf{W}_t$: Practical experience suggests that routine use of estimated correlations in relatively high dimensions may encounter numerical difficulties. In k dimensions, an estimated covariance matrix includes, in essence, the estimation of k standard deviations plus $k(k-1)/2$ correlations. In such situations, it may be wise to impose a variety of constraints to limit the number of independent parameters estimated in $\mathbf{V}_t$ *and* $\mathbf{W}_t$, perhaps using concepts from factor analysis and / or structural equation estimation.
- *Verify that* $\mathbf{\Sigma}_{t|t}$ *and* $\mathbf{\Sigma}_{t|t-1}$ *are always comfortably nonsingular*: If either of these matrices gets close to being numerically singular, this might generate numerical instabilities in the computations. If that happens it would be wise to check the matrix of first partial derivatives of the transition function, $\mathbf{f}_t(\mathbf{x}_t, \mathbf{u}_t)$, in (6.1) above. Recall from the appendix in section 6 above that $\mathbf{\Sigma}_{t+1|t} = \mathbf{f}_{t,1}\mathbf{\Sigma}_{t|t}\mathbf{f}_{t,1}' + \mathbf{W}_t$, where

$$\mathbf{f}_{t,1} = \frac{\partial \mathbf{f}_t\left(\mathbf{x}_t = \mathbf{x}_{t|t}, \mathbf{u}_t\right)}{\partial \mathbf{x}_t}.$$

  If this matrix poorly conditioned, it could affect other computations in the Kalman update cycle, generating numerical stability problems.
- *Ignore wild observations*: West and Harrison (1999, ch. 11) discussed simply ignoring observations with extreme prediction errors. While

we did some checks for outliers and made some effort to limit resulting step sizes, we did not try this. In some cases, large prediction errors contain essentially no information about the state of the plant, having been generated by substantially different process. In such cases, it may be best simply to skip the Kalman updating step for that observation. Of course, one must be careful in doing this, because such outliers could by themselves signal a malfunction of interest. A reasonable compromise might be to limit outliers to a certain maximum that would still permit them to have some effect on the estimated state of the plant while contributing appropriately to a Cusum of prediction errors indicating an inadequate model, as discussed in section 7.1 above. Under certain circumstances, West and Harrison (1999, ch. 12) suggested running two Kalman filters for a short period of time, one ignoring a wild observation and the other processing it. Two or three observations later, the one with the larger prediction error can be terminated and the wild observation can then be classified as either something that should or should not be considered in the model.

- *Cholesky Computations*: While we made some effort to improve the numerical precision of computations, we did not try programming Kalman filtering in terms of the Cholesky square roots of the relevant covariance matrices, as discussed by Bierman (1977). By some

reports, this can cut in half the number of significant digits required in certain applications.

7.5. Conclusions

In this section, we have described a procedure for detecting faults by an excessive elevation in a relative variance parameter. This could be accomplished either using a Cusum for variance as discussed here or an EWMA for variance, generalizing the techniques discussed in section 4 above. Fault isolation could then be achieved by scanning "multiple models", each one designed to describe a particular fault or combination of faults. Fault isolation is achieved by looking for a model with low relative variance. In the example discussed in Tables 7.3 and 7.4, we scanned first a list of monitors designed to mimick isolated faults. If the relative variances for all those models were high, the plant must be experiencing a multiple fault. A subsequent review of models for pairwise failures should then find a model with low relative variance if it is only a pairwise fault. Otherwise, we conclude that the fault has is not one of those modeled, e.g., a triple fault.

For extensions to the current work, see the literature on *distributed* Kalman filters, e.g., in Bar-Shalom (1990) and Bar-Shalom and Blair (2000).

REFERENCES

Bar-Shalom, Y., ed. (1990) *Multitarget-Multisensor Tracking: Applications and Advances*, vol. I (Norwood, MA: Artech House; reprinted 1996 by Yaakov Bar-Shalom, Stoors, CT, ybs@ee.uconn.edu).

Bar-Shalom, Y., and Blair, W. D., eds. (2000) *Multitarget-Multisensor Tracking: Applications and Advances*, vol. III (Norwood, MA: Artech House).

Bierman, G. J. (1977) *Factorization Methods for Discrete Sequential Estimation* (NY: Academic Press).

Bisgaard, S., Graves, S., Kulahci, M., Van Gilder, J., James, J., Marko, K., Ting, T., Wu, C., and Zatorski, H. (2001) "Accelerated Testing of On-Board Diagnostics", technical report.

Box, G. E. P., and Draper, N. R. (1987) *Empirical Model Building and Response Surfaces* (NY: Wiley).

Box, G., Graves, S., Bisgaard, S., Van Gilder, J., Marko, K, James, J., Seifer, M., Poublon, M., and Fodale, F. (2000) "Detecting Malfunctions in Dynamic Systems", *SAE Technical Paper Series* 2000-01-0363.

Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978) *Statistics for Experimenters* (NY: Wiley).

Pole, A., West, M., and Harrison, H. (1994) *Applied Bayesian Forecasting and Time Series Analysis* (NY: Chapman & Hall)

Titterington, D. M., Smith, A. F. M., and Makov, U. E. (1985) *Statistical Analysis of Finite Mixture Distributions* (NY: Wiley).

West, M. and Harrison, P. J. (1999) *Bayesian Forecasting and Dynamic Models*, 2nd ed., corrected 2nd printing (NY: Springer)

Part III. Concluding Remarks

This study has explored the power of Bayesian sequential updating as a basic principle for the design of procedures to detect and isolate faults in complex systems.

PART III. CONCLUDING REMARKS

In our complex, advanced society, our productivity, quality of life, and even our very survival depend on proper functioning of complex systems. Much of our time is spent trying to figure out why one system or another does not operate as we anticipate. Even worse, many people are killed each year when the vehicles in which they are riding malfunction or otherwise perform in ways they did not anticipate. People also die or suffer unnecessarily because certain problems are misdiagnosed or are diagnosed too late. The batteries or the electrodes on heart pacemakers fail without alerting patient or physician. In recent years, people have died from bacterial contamination of the drinking water in cities like Milwaukee, WI, where the water is supposedly monitored routinely to ensure potability. To combat air pollution, governments in the US, Canada and Europe now attempt to improve air quality by requiring that all new automobiles sold today in their jurisdictions contain on-board diagnostics (OBDs) designed to inform the driver of conditions that may jeopardize the emission control systems; other nations around the world are adopting similar regulations. These situations all call for effective monitors to detect and help diagnose malfunctions.

In this report, we have tried to establish Bayesian sequential updating as a conceptual foundation for monitoring. We distinguish monitoring from testing in that testing attempts to evaluate a fixed, unchanging state or condition of nature, while monitoring looks for a change. Of course, there is some overlap: In clinical trials, researchers test to determine if a new therapy is safe and effective. They also monitor for changes in each patient's condition. In some cases, monitoring has been described as a sequence of tests. In fact, monitoring is fundamentally different from testing, because in

monitoring we are more concerned with the *delay to detection* of a problem than with the probability of failing to detect it at a particular point in time.

Bayesian sequential updating is a two-step procedure:

1. *Observation and Bayes' theorem*: Collect data and update one's perception of the condition of the plant by applying Bayes' theorem.
2. *Migration*: Permit the plant to deteriorate or otherwise transition to a different, possibly deficient, condition.

In this report, we have derived common monitoring procedures from different assumptions about the observation and migration processes. In so doing, we have derived procedures essentially equivalent to the popular cumulative sum (Cusum, section 2) and exponentially weighted moving average (EWMA, sections 3 and 4) monitors. We have also derived other procedures for addressing more subtle monitoring applications, using Kalman filtering (sections 5 and 6) and multiple model adaptive estimation (MMAE, section 7) to exploit the analytic redundancy inherent in applications where the laws of physics permit the use of multiple sensors to check each other.

To apply Bayesian sequential updating, there is no need for subjective probabilities. It is difficult to imagine someone designing a monitor for something that is unrelated to anything that has occurred before. If there were no reference population of faults, there would be no motivation to design a monitor. In section 2, we made explicit use of the hazard rate to design a Bayes-adjusted Cusum. In section 3, we described how the key parameters for designing an exponentially weighted moving average monitor can

be estimated from reliability data and from studies of gage repeatability and reproducibility.

One class of common monitors was not discussed: statistical control charts. In essence, these procedures assume that every observation comes from a mixture of distributions where one component is dominant, and we want to detect the observations that come from a (presumed rare) contaminating distribution. In terms of Bayesian sequential updating, we could say that the migration step is absent, and the application of Bayes' theorem is converted into a simple assessment of whether each point comes from the dominant or contaminating distribution.

In sum, we believe we have established Bayesian sequential updating as a very powerful and flexible principle for designing monitors for specialized applications.

ACKNOWLEDGEMENT

The research described in this report was supported by the Low Emissions Research and Development Partnership of DaimlerChrysler, Ford, and General Motors.