INT. J. CONTROL, 1997, VOL. 67, NO. 5, 699–711

Detection and estimation of abrupt changes in input or state

P. F. WESTON†‡ and J. P. NORTON†

Disturbances which may be represented as step changes in the state of a linear, discrete-time, dynamical system are considered. A test for detection of such disturbances is presented. It employs variables computed in fixed-interval optimal smoothing. Several interpretations of the test are offered. The optimal estimate of the state change is shown to be one of the quantities forming the test statistic. Examples show the simplicity and effectiveness of the technique.

1. Introduction

The problem considered is detection of one or more abrupt changes in the state of a linear, dynamical system described by a discrete-time, state-space model. Detection of impulses in the input can be handled as an augmented-state version of the problem, integrating the output then moving the integration to the model input. Further state augmentation allows many other simple variations in input to be treated in the same way. Abruptly changing parameters described by a state-space model (random-walk models or their generalizations, Norton 1975, 1976, Young 1984) are also covered. The approach presented here, based on fixed-interval optimal smoothing (Bierman 1977, Maybeck 1982), scans a set of input–output records for evidence of change at any point.

The control and signal-processing literature (Willsky 1976, Isermann 1984, Basseville 1988, Basseville and Benveniste 1986, Gertler 1988, Frank 1990) has mainly considered online change detection, with speed of detection a prime concern. Offline location of changes is also important, e.g. finding the origin of a plant or feed disturbance in a process plant, or a pollutant discharge into the environment, or reconstructing a sharp aircraft or missile manoeuvre. Offline processing allows the observation record to be fully exploited by fixed-interval optimal smoothing when looking for changes; substitution of fixed-lag smoothing would allow online use of the technique, at some cost in efficiency.

Detection of changes in level, slope, parameters or variance has a long history in the time-series literature. Statistical detection tests have sometimes been used (Tsay 1988) but not with optimal smoothing, where instead the more informal practice of variance intervention, increasing the variance of increments at a suspected change, is employed; Young (1984, 1994) and Young and Ng (1989) review the development of the approach. The technique presented here may be regarded as generalizing and formalizing that practice, but distinctively does not assume prior knowledge of the timing, size, direction or probability of change as often required elsewhere (e.g.

Received 25 July 1995. Revised 30 July 1996.

† School of Electronic & Electrical Engineering, University of Birmingham, Edgbaston, Birmingham B15 2TT, U.K. Tel. +44 (0)121 414 4301/4292; Fax +44 (0)121 414 4291; e-mail: J.P.Norton@bham.ac.uk.

‡ Present address: Department of Engineering, University of Leicester, Leicester LE1 7RH, U.K.

0020-7179/97 $12.00 © 1997 Taylor & Francis Ltd.

Isermann 1984, Basseville 1988, Jun 1989). It relies only on detecting behaviour inconsistent with the probabilistic specification of the forcing and observation noise (zero mean, white, of known covariance). The discrepancy between the forward state estimate from earlier observations and the backward estimate from later observations is tested, rather than the innovations sequence as is more usual (Willsky and Jones 1976). The technique has some similarity to that presented by Niedźwiecki (1994) for estimating abruptly time-varying parameters in a linear model, discussed in § 5.

Section 2 gives a test for the presence of a step in state, derived from analysis of the difference between the forward and backward state estimates. It is shown how the test may be implemented using quantities available from the modified Bryson–Frazier smoothing algorithm. (Appendix 1 picks out the necessary items of fixed-interval optimal smoothing.) Section 2 also develops alternative interpretations of the test. Section 3 discusses how the step size and direction may be estimated. Section 4 gives some numerical examples and considers practical issues arising from them. The scope of the approach is examined further in § 5, which also draws conclusions.

2. Hypothesis test for change in state

The system is modelled by

$$\left.\begin{aligned} x_{k+1} &= \Phi_k x_k + \Gamma_k w_k \\ y_k &= H_k x_k + v_k \quad k = 1, 2, \ldots, N \end{aligned}\right\} \tag{1}$$

with $x \in \mathbb{R}^n$, $w \in \mathbb{R}^r$, $y \in \mathbb{R}^m$, $E[w_k] = 0$, $E[v_k] = 0$ and $\operatorname{cov}(w_k) = Q_k$, $\operatorname{cov}(v_k) = R_k$. An initial state estimate $\hat{x}_0^{\mathrm{f}}$ and its covariance P_0^{f} are also specified. For simplicity, there is assumed to be no deterministic forcing or correlation between w_k and v_k, although both can be accommodated. Robustness in the face of error in Φ_k or H_k is not considered. The forwards (filtered) estimate of x_k, based on $\hat{x}_0^{\mathrm{f}}$ and the observations $y_j, j = 1, 2, \ldots, k$, is denoted by $\hat{x}_k^{\mathrm{f}}$, its prediction from the previous estimate by $\hat{x}_k^{\mathrm{f}'}$ and their respective covariances by $P_k^{\mathrm{f}}, P_k^{\mathrm{f}'}$. Correspondingly, the estimate based on later observations y_j, $j = k, k+1, \ldots, N$ is denoted by $\hat{x}_k^{\mathrm{b}}$, its postdiction from that based on y_j, $j = k+1, k+2, \ldots, N$ by $\hat{x}_k^{\mathrm{b}'}$ and their covariances by $P_k^{\mathrm{b}}, P_k^{\mathrm{b}'}$.

The change–detection statistic will compare $\hat{x}_k^{\mathrm{f}'}$ with $\hat{x}_k^{\mathrm{b}}$; it yields the probability that the difference

$$\delta_k \equiv \hat{x}_k^{\mathrm{b}} - \hat{x}_k^{\mathrm{f}'} \tag{2}$$

is consistent with its covariance as computed from the specified Q_ks, R_ks and P_0^{f}. Under the null hypothesis (forcing and observation noise white, with zero means and specified covariances), $\hat{x}_k^{\mathrm{f}'}$ and $\hat{x}_k^{\mathrm{b}}$ are unbiased, $E[\delta_k]$ is zero and, as also $\hat{x}_k^{\mathrm{f}'}$ and $\hat{x}_k^{\mathrm{b}}$ are mutually uncorrelated,

$$\begin{aligned} \Delta_k \equiv \operatorname{cov} \delta_k = E[\delta_k \delta_k^{\mathrm{T}}] &= E[\hat{x}_k^{\mathrm{b}} \hat{x}_k^{\mathrm{bT}} - \hat{x}_k^{\mathrm{b}} \hat{x}_k^{\mathrm{f'T}} - \hat{x}_k^{\mathrm{f}'} \hat{x}_k^{\mathrm{bT}} + \hat{x}_k^{\mathrm{f}'} \hat{x}_k^{\mathrm{f'T}}] \\ &= E[\hat{x}_k^{\mathrm{b}} \hat{x}_k^{\mathrm{bT}}] - 2E[x_k]E[x_k^{\mathrm{T}}] + E[\hat{x}_k^{\mathrm{f}'} \hat{x}_k^{\mathrm{f'T}}] = P_k^{\mathrm{b}} + P_k^{\mathrm{f}'} \end{aligned} \tag{3}$$

If δ_k is assumed to be gaussian (usually reasonable except perhaps near the ends of the records), the statistic

$$d_k \equiv \delta_k^{\mathrm{T}} \Delta_k^{-1} \delta_k \tag{4}$$

is a χ^2 variate with n degrees of freedom. An upper acceptance threshold for d_k is given directly by the lowest acceptable probability that δ_k results by chance under the null hypothesis. The alternative hypothesis, concerning what type of abrupt change is possible, need not be specified. If d_k fails the test, the failure is attributed to a state change. Estimation of its direction and size is considered later.

If the Fraser–Potter implementation is used, d_k can be evaluated directly from the forward and backward estimates and their estimated covariances. However, the modified Bryson–Frazier algorithm will ultimately provide broader insight; d_k may be expressed conveniently in terms of quantities produced by this algorithm. First write Δ_k from (3) as

$$\Delta_k = P_k^{\mathrm{b}} + P_k^{\mathrm{f}'} = P_k^{\mathrm{b}}(P_k^{\mathrm{f}'-1} + P_k^{b-1})P_k^{\mathrm{f}'} = P_k^{\mathrm{b}} P_k^{-1} P_k^{\mathrm{f}'} \tag{5}$$

As P_k^{b} does not appear in the algorithm, an alternative expression for Δ_k will be found shortly, but (5) is useful in deriving δ_k and thence d_k. From (A1.4)

$$\hat{x}_k^{\mathrm{b}} = P_k^{\mathrm{b}}(P_k^{-1}\hat{x}_k - P_k^{\mathrm{f}'-1}\hat{x}_k^{\mathrm{f}'}) \tag{6}$$

so

$$\begin{aligned}\delta_k &= P_k^{\mathrm{b}}(P_k^{-1}\hat{x}_k - P_k^{\mathrm{f}'-1}\hat{x}_k^{\mathrm{f}'} - P_k^{\mathrm{b}-1}\hat{x}_k^{\mathrm{f}'}) = P_k^{\mathrm{b}}P_k^{-1}(\hat{x}_k - \hat{x}_k^{\mathrm{f}'})\\ &= \Delta_k P_k^{\mathrm{f}'-1}(\hat{x}_k - \hat{x}_k^{\mathrm{f}'}) \equiv -\Delta_k\lambda_k \end{aligned}\tag{7}$$

where λ_k is defined by

$$\lambda_k \equiv -P_k^{\mathrm{f}'-1}(\hat{x}_k - \hat{x}_k^{\mathrm{f}'}) \tag{8}$$

The equation for the smoothed state estimate in the modified Bryson–Frazier algorithm gives (8) directly, and

$$\begin{aligned}\Delta_k^{-1} &= (P_k^{\mathrm{f}'} + P_k^{\mathrm{b}})^{-1} = P_k^{\mathrm{f}'-1} - P_k^{\mathrm{f}'-1}(P_k^{\mathrm{f}'-1} + P_k^{\mathrm{b}-1})^{-1}P_k^{\mathrm{f}'-1}\\ &= P_k^{\mathrm{f}'-1} - P_k^{\mathrm{f}'-1}P_kP_k^{\mathrm{f}'-1}\end{aligned}\tag{9}$$

which is the expression for Λ_k given by the smoothed-covariance equation. More directly, using (7) and noting that cov $\lambda_k \equiv \Lambda_k$ and under the null hypothesis, $E[\delta_k]$ is zero, so $E[\lambda_k]$ is zero

$$\Delta_k \equiv E[\delta_k\delta_k^{\mathrm{T}}] = E[\Delta_k\lambda_k\lambda_k^{\mathrm{T}}\Delta_k] = \Delta_k(\mathrm{cov}\ \lambda_k + E[\lambda_k]E[\lambda_k^{\mathrm{T}}])\Delta_k = \Delta_k\Lambda_k\Delta_k \tag{10}$$

and so

$$\Delta_k^{-1} \equiv \Lambda_k \tag{11}$$

Alternative interpretations of d_k exist. Using (4), (7), (8) and (5) in succession

$$\begin{aligned}d_k &= \lambda_k^{\mathrm{T}}\Delta_k\lambda_k = (\hat{x}_k - \hat{x}_k^{\mathrm{f}'})^{\mathrm{T}}P_k^{\mathrm{f}'-1}\cdot P_k^{\mathrm{b}}P_k^{-1}P_k^{\mathrm{f}'}\cdot P_k^{\mathrm{f}'-1}(\hat{x}_k - \hat{x}_k^{\mathrm{f}'})\\ &= (\hat{x}_k - \hat{x}_k^{\mathrm{f}'})^{\mathrm{T}}P_k^{\mathrm{f}'-1}(P_k^{-1} - P_k^{\mathrm{f}'-1})^{-1}P_k^{-1}(\hat{x}_k - x_k^{\mathrm{f}'})\\ &= (\hat{x}_k - \hat{x}_k^{\mathrm{f}'})^{\mathrm{T}}(P_k^{\mathrm{f}'} - P_k)^{-1}(\hat{x}_k - \hat{x}_k^{\mathrm{f}'})\end{aligned}\tag{12}$$

and

$$\mathrm{cov}\ (\hat{x}_k - \hat{x}_k^{\mathrm{f}'}) = E[\hat{x}_k\hat{x}_k^{\mathrm{T}} - \hat{x}_k\hat{x}_k^{\mathrm{f}'\mathrm{T}} - \hat{x}_k^{\mathrm{f}'}\hat{x}_k^{\mathrm{T}} + \hat{x}_k^{\mathrm{f}'}\hat{x}_k^{\mathrm{f}'\mathrm{T}}] \tag{13}$$

where

$$\begin{aligned}E[\hat{x}_k\hat{x}_k^{\mathrm{f}'\mathrm{T}}] &= E[P_k(P_k^{\mathrm{f}'-1}\hat{x}_k^{\mathrm{f}'} + P_k^{\mathrm{b}-1}x_k^{\mathrm{b}})\hat{x}_k^{\mathrm{f}'\mathrm{T}}]\\ &= P_k(P_k^{\mathrm{f}'-1}(P_k^{\mathrm{f}'} + E[x_k]E[x_k^{\mathrm{T}}]) + P_k^{\mathrm{b}-1}E[x_k]E[x_k^{\mathrm{T}}])\\ &= P_k(I + (P_k^{\mathrm{f}'-1} + P_k^{b-1})E[x_k]E[x_k^{\mathrm{T}}]) = P_k + E[x_k]E[x_k^{\mathrm{T}}]\end{aligned}\tag{14}$$

so

$$\operatorname{cov}(\hat{x}_k - \hat{x}_k^{f'}) = P_k + E[x_k]E[x_k^T] - 2(P_k + E[x_k]E[x_k^T]) + P_k^{f'} + E[x_k]E[x_k^T]$$
$$= P_k^{f'} - P_k \tag{15}$$

Thus, (12) interprets d_k as the χ^2 test statistic for the adjustment $\hat{x}_k - \hat{x}_k^{f'}$ which brings the information from later observations into the estimate of x_k.

The change–detection test also has an enlightening interpretation in terms of λ_k. From (4), (7) and (11)

$$d_k = \delta_k^T \Delta_k^{-1} \delta_k = \lambda_k^T \Delta_k \lambda_k = \lambda_k^T \Lambda_k^{-1} \lambda_k \tag{16}$$

so the text is a χ^2 significance test for λ_k. This links it with Gauss–Markov or generalized least-squares (g.l.s.) estimation. Optimal smoothing minimizes the g.l.s. cost function

$$J_N = \frac{1}{2}\left(\sum_{k=1}^{N}\{(y_k - H_k x_k)^T R_k^{-1}(y_k - H_k x_k)\} + \sum_{k=0}^{N-1}\{w_k^T Q_k^{-1} w_k\} + (x_0 - \hat{x}_0^f)^T P_0^{f-1}(x_0 - \hat{x}_0^f)\right) \tag{17}$$

with respect to the sequence $\{x_k\}$, subject to the constraints

$$x_k = \Phi_{k-1} x_{k-1} + \Gamma_{k-1} w_{k-1} \quad k = 1, 2, \ldots, N \tag{18}$$

It is minimized by setting to zero the gradients of the augmented cost function

$$L_N \equiv J_N - \sum_{k=1}^{N} \lambda_k^T \{x_k - \Phi_{k-1} x_{k-1} - \Gamma_{k-1} w_{k-1}\} \tag{19}$$

with respect to each x_k (including x_0), w_{k-1} and λ_k. The Lagrange multiplier λ_k in (19) may be identified with λ_k defined by (8), as shown in Appendix 1. It is the gradient of J_N with respect to the value of the right-hand side of (18). In other words, λ_k shows the effect on J_N of inserting an extra step between the smoothed estimates $\Phi_{k-1}\hat{x}_{k-1} + \Gamma_{k-1}\hat{w}_{k-1}$ and $\hat{x}_k$ yielded by the model (1) and its noise specifications. The change–detection statistc, interpreted as in (16), indicates how large this gradient is, allowing for its normal statistical variation.

3. Estimation of direction and size of step in state

The optimal estimate of a step change in state between $\Phi_{k-1}x_{k-1}$ and x_k will be shown to be δ_k given by (7). Complete freedom in the step $x_k - \Phi_{k-1}x_{k-1}$ is allowed by specifying

$$\Gamma_{k-1} = I, \quad Q_{k-1} = \lim_{\sigma^2 \to \infty} \sigma^2 I \tag{20}$$

for the forcing $\Gamma_{k-1}w_{k-1}$. The inverse $P_k^{f'-1}$ of the predicted state-error covariance is then zero, so from (A 1.3) P_k is P_k^b. Hence from (A 1.4) $\hat{x}_k$ is $\hat{x}_k^b$: observations before the step have no influence on $\hat{x}_k$, because $\Gamma_{k-1}w_{k-1}$ is utterly uncertain.

Similarly, the observations after the step have no effect on the smoothed estimate of $\hat{x}_{k-1}$. It can be verified through the Fraser–Potter equations that $\hat{x}_{k-1}$ is $\hat{x}_{k-1}^f$ and P_{k-1} is P_{k-1}^f. The optimally smoothed estimate of an unconstrained change is therefore

$$\hat{x}_k - \Phi_{k-1}\hat{x}_{k-1} = \hat{x}_k^{\mathrm{b}} - \Phi_{k-1}\hat{x}_{k-1}^{\mathrm{f}} = \hat{x}_k^{\mathrm{b}} - \hat{x}_k^{\mathrm{f}'} \equiv \delta_k \tag{21}$$

Although Γ_{k-1} and Q_{k-1} were replaced by I and $\lim_{\sigma^2\to\infty}\sigma^2 I$ to derive (21), $\hat{x}_k^{\mathrm{b}}$ and $\hat{x}_k^{\mathrm{f}'}$ are independent of Γ_{k-1} and Q_{k-1}, so δ_k in (21) is as computed with the original Γ_{k-1} and Q_{k-1}. Thus, the optimal estimates of possible state changes are obtained, for every point in the record, by a single optimal smoothing process using the 'no extra steps' process-noise specification. (Optimization of Q and R away from the steps is an important but separate issue; iteration may be necessary.)

With δ_k the optimal estimate of the state change and λ_k the gradient of cost J_N with respect to that change, yet another interpretation of the change-test statistic d_k suggests itself. From (4) and (7)

$$d_k = -\lambda_k^{\mathrm{T}}\delta_k \tag{22}$$

so d_k gives the reduction in J_N due to δ_k. The test statistic thus indicates how much the Gauss–Markov cost would be reduced (and the corresponding gaussian log-likelihood function increased) by allowing, through (20), a free step δ_k in state at time k.

The view of λ_k as the marginal benefit of a constraint relaxation also clarifies the rôle of Λ_k in the change–test statistic (16). The mean-square change in cost J_N due to a step δ in state at time k is, under the null hypothesis,

$$E[(\lambda_k^{\mathrm{T}}\delta)^2] = \delta^{\mathrm{T}}E[\lambda_k\lambda_k^{\mathrm{T}}]\delta = \delta^{\mathrm{T}}(\mathrm{cov}\,(\lambda_k) + E[\lambda_k]E[\lambda_k^{\mathrm{T}}])\delta \tag{23}$$

As Λ_k is $\mathrm{cov}\,(\lambda_k)$ and, under the null hypothesis, δ_k and hence λ_k are zero-mean

$$E[(\lambda_k^{\mathrm{T}}\delta)^2] = \delta^{\mathrm{T}}\Lambda_k\delta \equiv d(\delta),\ \text{say} \tag{24}$$

Thus, with $\delta = \delta_k$, d_k gives the mean-square reduction in cost (under the null hypothesis) due to the optimal step. More generally, Λ_k shows the expected influence of, and hence weight attached to, each component of a prospective step. It is easy, for instance, to find the step direction with most or least effect on cost. If δ and the mutually orthogonal eigenvectors m_{ki}, $i = 1, 2, \ldots, n$ of the positive-definite Λ_k are normalized to unit euclidean length $\|\cdot\|$, the unit-length step maximizing (minimizing) the mean-square change in J_N is

$$\hat{\delta} = \underset{\|\delta\|=1}{\arg\max\,(\min)}\ \delta^{\mathrm{T}}\Lambda_k\delta = m_{kj} \tag{25}$$

where m_{kj} is the eigenvector corresponding to the largest (smallest) eigenvalue of Λ_k. The eigenstructure of Λ_k thus reflects the prior potential of the observed variables for change detection; the ellipsoidal constant-d contour

$$\delta^{\mathrm{T}}\Lambda_k\delta \equiv d^* \tag{26}$$

shows how small a state change in any direction may be detected with the confidence corresponding to d^*.

Equations (11), (16) and (22) express a duality between δ_k and λ_k. Equation (24) has the counterpart

$$E[(\lambda^{\mathrm{T}}\delta_k)^2] = \lambda^{\mathrm{T}}\Delta_k\lambda \equiv d(\lambda) \tag{27}$$

with the adjoint variables regarded as the basis for change detection. The smallest significant size of λ in any direction is given by the eigenstructure of Δ_k. The inverse relation between Δ_k and Λ_k reflects the fact that a smaller change in state is detectable if the expected sensitivity of J_N to it is larger, and vice versa.

Finally, in view of the use of optimal smoothing by some authors for *ad hoc* change detection, and the popularity of variance intervention (sharp increase in Q at suspected steps), one might ask whether a step in state can be estimated equally well by the optimally smoothed estimate

$$\hat{x}_k - \Phi_{k-1}\hat{x}_{k-1} = -\Gamma_{k-1}\hat{w}_{k-1} \tag{28}$$

It cannot because, in contrast to λ_k (or δ_k), $\Gamma_{k-1}\hat{w}_{k-1}$ is influenced by Q_{k-1} and by the rank of Γ_{k-1}. Setting the derivative of L_N in (19) with respect to w_{k-1} to zero then premultiplying by $\Gamma_{k-1}Q_{k-1}$

$$\Gamma_{k-1}\hat{w}_{k-1} = -\Gamma_{k-1}Q_{k-1}\Gamma_{k-1}^{\mathrm{T}}\lambda_k \tag{29}$$

Although Γ_{k-1} has full column rank r, $r \leq n$ so $\Gamma_{k-1}Q_{k-1}\Gamma_{k-1}^{\mathrm{T}}$ may not be of full rank. Information would then be lost if $\Gamma_{k-1}\hat{w}_{k-1}$ were used instead of λ_k (or δ_k) in change detection and estimation. Even if $\Gamma_{k-1}Q_{k-1}\Gamma_{k-1}^{\mathrm{T}}$ is of full rank, $\Gamma_{k-1}\hat{w}_{k-1}$ is restricted by the finite specified Q_{k-1}. Moreover, the restriction smears the influence of a state change between times $k-1$ and k over adjacent sample instants, making its detection and location harder. Plainly, a change cannot be detected as efficiently by inspecting smoothed state estimates as by the test statistic d. Heuristically increasing Q locally to loosen the restriction would merely increase sensitivity to observation noise; to acquire the information in $\{\delta_k\}$, variance intervention would have to be performed at every point in turn, requiring many smoothing runs.

This distinction between estimates δ_k and $\Gamma_{k-1}\hat{w}_{k-1}$ of a state change (the former optimal for change unrestricted by Γ_{k-1} or Q_{k-1}, the latter for the nominal Γ_{k-1} and Q_{k-1}) has implications for the detection of more than one change within a record. The optimality of δ_k depends on correct process-noise specfication throughout the rest of the record. If more than one change (inconsistent with the Γs and Qs) is present, misspecified Γ and Q for any one change renders δ for all the others incorrect. If changes are close enough, this may prevent their correct detection from the sequence of ds obtained by the initial optimal smoothing run. The simplest course is to accept a change where d is highest, set Γ to identity and Q to infinity there, repeat optimal smoothing, examine the new sequence of ds for a second change, and continue until d is always below the test threshold.

4. Application of the change-detection test

The scope and practical aspects of the test are best illustrated by examples in which the actual behaviour is known.

Example 1: Figure 1 shows time-series data used by Jun (1989) Young and Ng (1989. Example 3) to test detection of a level change. The constant-Q_k optimal-smoothing estimates, also in Fig. 1, are uninformative. In Fig. 2, the probability of the test statistic d_k reaching its calculated value in the absence of a jump is plotted on a log scale, with a possible threshold of 0·005 marked. It clearly indicates the presence of a single jump at $k = 70$. If the forward and backward estimates are decoupled at that point by making Q_{70} infinite, the statistic is brought below the threshold: Fig. 3. The behaviour just before the jump, distorted in the constant-Q_k case, Fig. 1, is now undistorted, Fig. 4. □

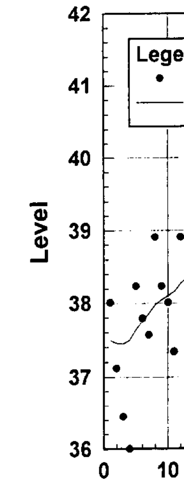

Figure 1. Example 1: ob

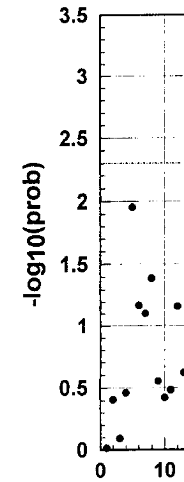

Figure 2. Example 1: –
time k,

Example 2: The prob
a system modelled by

$x_{k=1} = \begin{bmatrix} 0{\cdot}83 \\ -0{\cdot} \\ 0{\cdot}04 \end{bmatrix}$

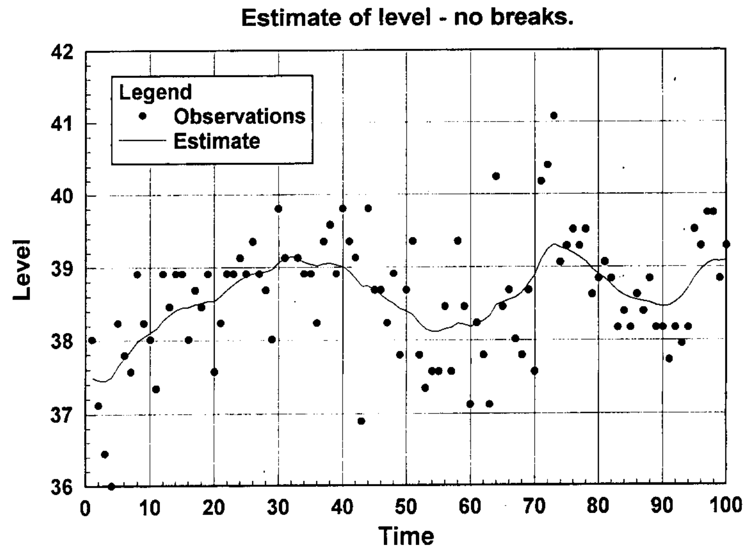

Figure 1. Example 1: observed time series and optimal-smoothing estimates (interpolated).

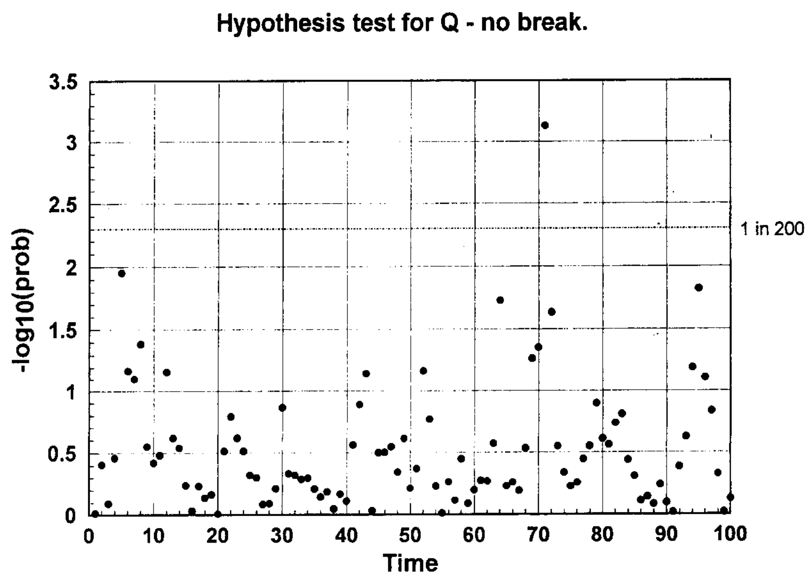

Figure 2. Example 1: $-\log_{10}$ prob (*d_k exceeds computed value with no step present*) against time k, with increment variance Q_k specified as constant.

Example 2: The problem is to detect two abrupt changes in the scalar input u of a system modelled by

$$x_{k=1} = \begin{bmatrix} 0{\cdot}833 & 0{\cdot}3 & 0 \\ -0{\cdot}3 & 0{\cdot}833 & 0 \\ 0{\cdot}04 & 0{\cdot}04 & 0{\cdot}96 \end{bmatrix} x_k + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u_k;\ y_k = [0 \quad 0 \quad 1] x_k + v_k \qquad (30)$$

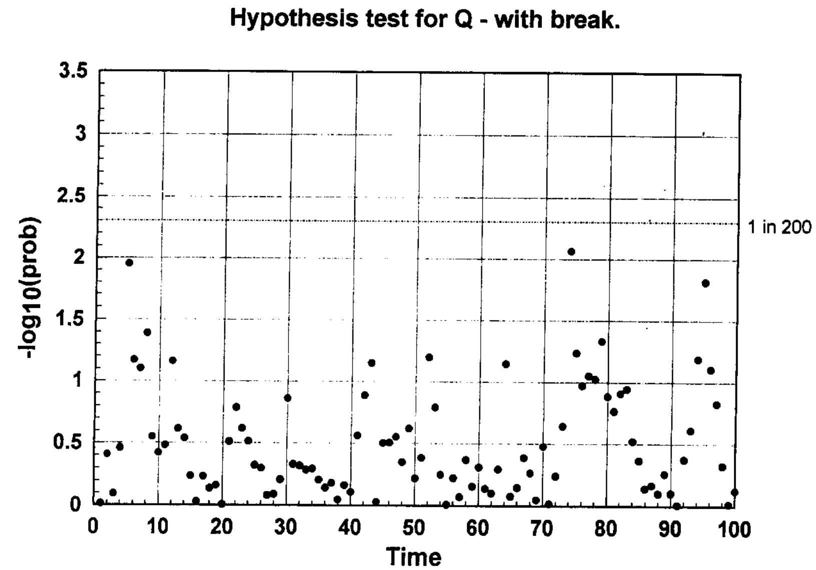

Figure 3. Example 1: $-\log_{10}$ prob (*d_k exceeds computed value with no step present*) against time k, with increment variance Q_k infinite at $k = 70$ and otherwise constant.

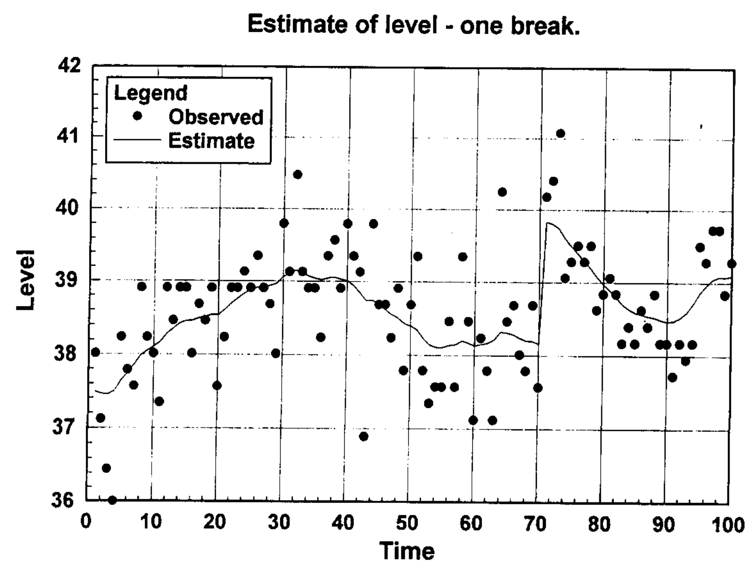

Figure 4. Example 1: optimal-smoothing estimates with break permitted at $k = 70$.

with constant observation-noise variance $R_k = 0{\cdot}01$. The input is modelled as a random walk

$$u_{k+1} = u_k + w_k \text{ with } \operatorname{var}(w_k) \equiv Q_k = 0{\cdot}01 \tag{31}$$

Figure 5 shows 200 output observations and the actual output and input; the example is designed so that the input changes at $k = 80$, $k = 150$ are not easily inferred from the output response. In Fig. 6 the probability of d_k attaining its computed value without a state jump is plotted as before, with a threshold of

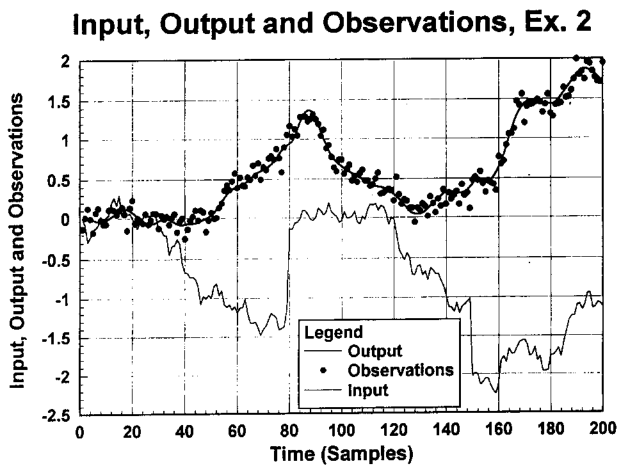

Figure 5. Example 2: actual input and output (interpolated) and observed output.

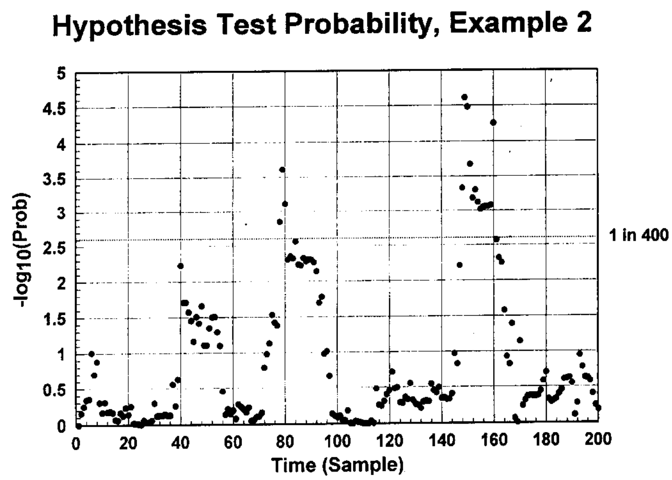

Figure 6. Example 2: $-\log_{10}$ prob (d_k *exceeds computed value with no step present*) against time k, with increment variance Q_k specified as constant.

0·0025 marked. Jumps at about $k = 80$ and $k = 150$, the centres of the periods over which the threshold is exceeded, are indicated. The influences of the two equal-sized jumps on the probability computed from $\{d_k\}$ differ markedly because of noise, but both are detected. Allowing free jumps at $k = 80$ and $k = 150$ brings the corresponding test statistic peaks below the threshold, with little effect elsewhere, as in Fig. 7. The optimally smoothed input estimate, Fig. 8, is again much better close to the jumps than that from the original constant-Q_k model, Fig. 9.

5. Scope of method; conclusions

The technique presented here is suitable whenever a change can be related to a step in an observable state variable (augmenting the state if necessary). Steps in input (e.g. feed changes) or output (e.g. load) are readily covered, and exponential, ramp

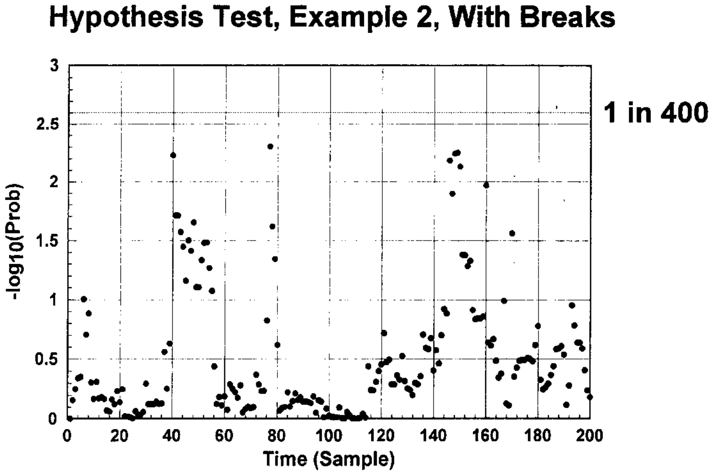

Figure 7. Example 2: $-\log_{10}$ prob (d_k *exceeds computed value with no step present*) against time k, with increment variance Q_k infinite at $k = 80, 150$ and otherwise constant.

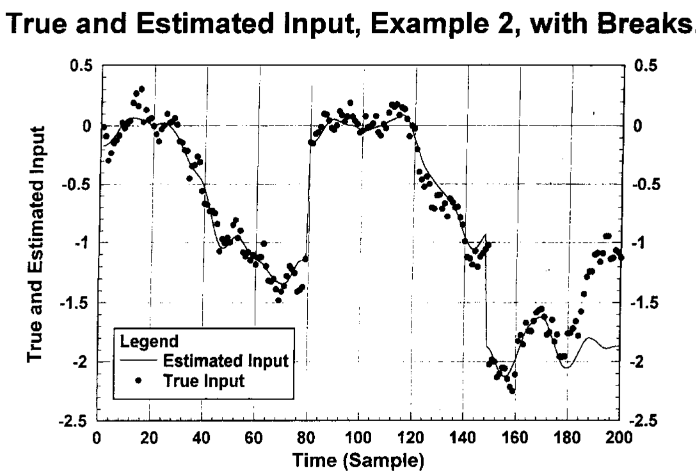

Figure 8. Example 2: actual input and optimal-smoothing estimates (interpolated) with breaks permitted at $k = 80, 150$.

or more complicated changes may be treated as due to abrupt forcing of a known auxiliary state-variable model.

The approach also extends to abrupt changes the well-established use of optimal smoothing for identification of changes in parameters of linear models (Norton 1975, 1976, 1986, Young 1984), treating the parameters as state variables. Its ability to detect and estimate simultaneous changes in a number of unknowns is essential when a change in a physically significant parameter such as a gain or time constant affects several parameters of a discrete-time model.

Niedźwiecki (1994) derived a parameter estimate as an optimal linear combination of estimates based on two random-walk models, each having a prior probability, relating parameters respectively to their predecessors and successors within a

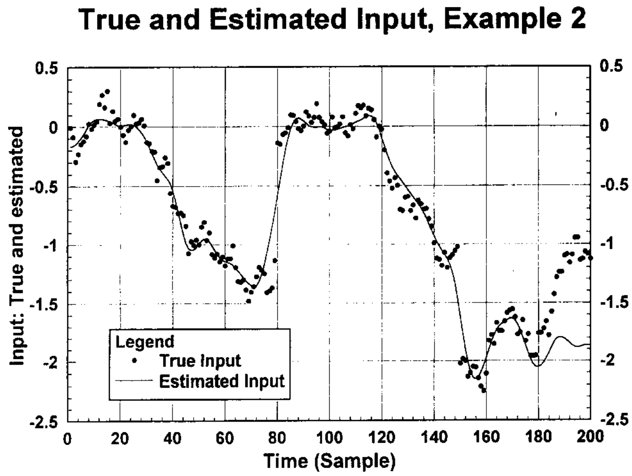

Figure 9. Example 2: actual input and optimal-smoothing estimates (interpolated) with increment variance Q_k specified as constant.

window centred at the present. The resulting algorithm, called competitive smoothing, makes simplifying assumptions but requires no intervention once a window width and values for Q have been selected. It shares with the technique presented here its use of two filters, one operating each side of the present, and its essentially offline nature. By switching between competing evolution models without explicit change detection, it risks spurious steps; Niedźwiecki suggests using a 'low-pass' model of parameter evolution, such as an integrated random walk (Norton 1976) to overcome this drawback.

In summary, the technique presented here offers a simple and flexible way to find sudden changes in state, input or parameters of a linear, discrete-time model, with a minimum of prior information and at negligible computational cost over that of optimal smoothing.

Appendix: Fixed-interval optimal smoothing (Bryson and Ho 1969, Norton 1975, Bierman 1977, Maybeck 1982 and their references)

The optimal smoothing problem is to compute estimates $\hat{x}_k$, $k = 0, 1, \ldots, N$, each based on the whole observation set $\{y_j, j = 1, 2, \ldots, N\}$ and on $\hat{x}_0^{\mathrm{f}}$ and P_0^{f}.

The many implementations of fixed-interval optimal smoothing developed in the 1960s and 1970s differ in numerical properties (including stability, Norton 1975), in flexibility (e.g. permitting the excitation matrix to have less than full rank), and in ease of interpretation. The change detection technique is most easily explained by reference to quantities in the Fraser–Potter algorithm (Fraser and Potter 1969, Maybeck 1982) but employs quantities computed by the modified Bryson–Frazier algorithm (Bryson and Frazier 1963, Bryson and Ho 1969, Bierman 1977).

The Fraser–Potter version combines the state estimates yielded by separate forwards and backwards passes through the observations. The forwards pass is standard Kalman filtering, while the backwards pass (Maybeck 1982) produces $\{\xi_N, \xi'_{N-1}, \xi_{N-1}, \xi'_{N-2}, \ldots, \xi_0\}$ and $\{M_N, M'_{N-1}, M_{N-1}, \ldots, M_0\}$, from which the covariance of the backwards state estimate is obtained by

$$P_k^{\mathrm{b}} = M_k^{-1} \tag{A 1.1}$$

then the state estimate itself by

$$\hat{x}_k^{\mathrm{b}} = P_k^{\mathrm{b}} \xi_k \tag{A 1.2}$$

There is no correlation between the errors in $x_k^{\mathrm{f}'}$ and $\hat{x}_k^{\mathrm{b}}$; that in $\hat{x}_k^{\mathrm{f}'}$ is due only to the error in $\hat{x}_0^{\mathrm{f}}$ and to $\{w_0, v_1, w_1, \ldots, v_{k-1}, w_{k-1}\}$, while the error in $\hat{x}_k^{\mathrm{b}}$ is due only to $\{v_N, w_{N-1}, v_{N-1}, \ldots, w_k, v_k\}$. They are also unbiased, minimum-covariance estimates, so the information matrix of their minimum-covariance linear combination $\hat{x}_k$, the optimal smoothed estimate, is the sum of their information matrices:

$$P_k^{-1} = P_k^{\mathrm{f}'-1} + P_k^{\mathrm{b}-1} \equiv P_k^{\mathrm{f}'-1} + M_k \tag{A 1.3}$$

and (Bryson and Frazier 1963, Norton 1986)

$$P_k^{-1}\hat{x}_k = P_k^{\mathrm{f}'-1}\hat{x}_k^{\mathrm{f}'} + P_k^{\mathrm{b}-1}\hat{x}_k^{\mathrm{b}} \equiv P_k^{\mathrm{f}'-1}\hat{x}_k^{\mathrm{f}'} + \xi_k \tag{A 1.4}$$

(Alternatively, stepping past the processing of y_k, the primed and unprimed quantities may be reversed.)

References

Basseville, M., 1988, Detecting changes in signals and systems—a survey. *Automatica*, **24**, 309–326.

Basseville, M., and Benveniste, A., 1986, *Detection of Abrupt Changes in Signals and Dynamical Systems*, LNCIS No. 77 (Berlin: Springer-Verlag).

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

Bryson, A. E., and Frazier, M., 1963, Smoothing for linear and nonlinear dynamic systems. Proceedings of the Optimum Systems Synthesis Conference, USAF Technical Report ASD-TDR–063-119.

Bryson, A. E., and Ho, Y. C., 1969, *Applied Optimal Control* (Waltham, Massachusetts, U.S.A.: Ginn (Blaisdell)).

Fraser, D. C., and Potter, J. E., 1969, The optimal linear smoother as a combination of two optimal linear filters. *IEEE Transactions on Automatic Control*, **14**, 387–390.

Frank, P., 1990, Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy—a survey and some new results. *Automaticsa*, **26**, 459–474.

Gertler, J. J., 1988, Survey of model-based failure detection and isolation in complex plant. *IEEE Control Systems Magazine*, December, 3–11.

Isermann, R., 1984, Process fault detection based on modelling and estimation methods. *Automatica*, **20**, 387–404.

Jun, D. B., 1989, On detecting and estimating a major level or slope change in general exponential smoothing. *Journal of Forecasting*, **8**, 55–64.

Maybeck, P. S., 1982, *Stochastic Models, Estimation, and Control*, Vol. 2 (New York: Academic Press).

Niedźwiecki, M., 1994, Identification of time-varying systems with abrupt parameter changes, *Automatica*, **30**, 447–459.

Norton, J. P., 1975, Optimal smoothing in the identification of linear time-varying systems. *Proceedings of the Institution of Electrical Engineers*, **122**, 663–668; 1976, Identification by optimal smoothing using integrated random walks. *Ibid.* **123**, 451–452; 1986, *An Introduction to Identification* (London: Academic Press).

Tsay, R. S., 1988, Outliers, level shifts, and variance changes in time series. *Journal of Forecasting*, **7**, 1–20.

Willsky, A. S., 1976, A survey of design methods for failure detection in dynamic systems. *Automatica*, **12**, 601–611.

Willsky, A. S., and Jones, H. L., 1976, A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *IEEE Transactions on Automatic Control*, **21**, 108–112.

Young, P. C., 1984, *Recursive Estimation and Time-Series Analysis* (Berlin: Springer-Verlag); 1994, Time-variable parameter and trend estimation in non-stationary economic time series. *Journal of Forecasting*, **13**, 179–210.

Young, P. C., and Ng, C., 1989, Variance intervention. *Journal of Forecasting*, **8**, 399–416.