Smoothing in Multiple Model Change Detection for Stochastic Systems

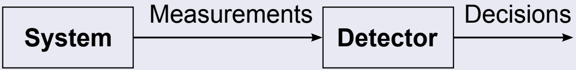
Ivo Punčochář, Jindřich Duník, Miroslav Šimandl

Department of Cybernetics
and
Research Centre Data – Algorithms – Decision Making
Faculty of Applied Sciences
University of West Bohemia
Pilsen, Czech Republic

Presentation Overview

Introduction

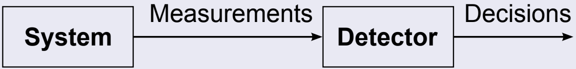
Change detection problem



- The primary task – recognize a change in an observed system as quick and reliable as possible
- Performance measures – the delay for detection, quality of detection, robustness with respect to disturbances, . . .
- Application areas – automatic control, signal processing, fault detection, . . .

Introduction

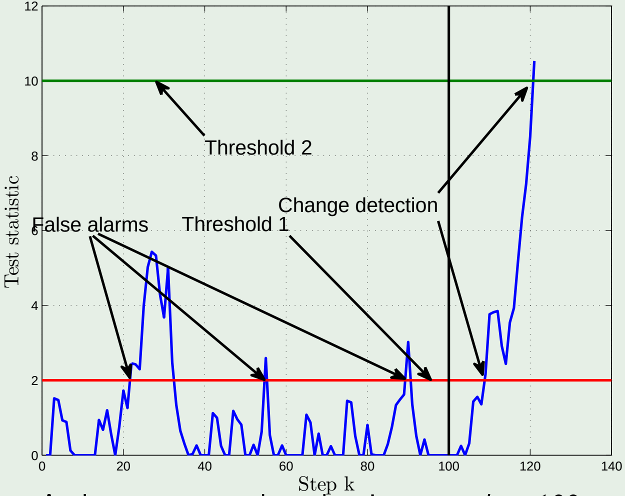
Change detection problem



- The primary task – recognize a change in an observed system as quick and reliable as possible
- Performance measures – the delay for detection, quality of detection, robustness with respect to disturbances, . . .
- Application areas – automatic control, signal processing, fault detection, . . .

A tradeoff between the delay for detection and the quality of decisions

An example of a detector based on a statistical test



A change occurred at the time step $k = 100$.

Introduction – cont'd

Multiple model change detection approach

- Use of the multiple model approach for system description
- Suitable for systems that may undergo abrupt changes and individual models are known (used e.g. in fault detection, state estimation, target tracking)
- Decisions are typically based on filtering estimates of the state (i.e. $p(\mathbf{x}_k|\mathbf{z}_0^k)$)

Deferred decisions

- In a specific application it is possible to defer decisions, obtain more measurements and use smoothing estimates (i.e. $p(\mathbf{x}_{k-\ell}|\mathbf{z}_0^k)$)

Introduction – cont'd

A motivational example – a production line

The decisions about past changes can be used to react to these changes. Other examples: pipelines, car engines, ...

Introduction – cont'd

Goals

- Formulate the problem of change detection with delayed decisions in the multiple model framework
- Design the optimal detector, that utilizes smoothing estimates, using the closed loop information processing strategy
- Present the smoothing algorithm used in the designed optimal detector

Note: The change detection problem with deferred decisions has not been considered in the literature yet.

Problem formulation

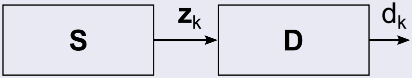
Description of the system at time steps $k \in \mathcal{T} = \{0, 1, \dots, F\}$

$$\begin{aligned}\mathbf{x}_{k+1} =& \mathbf{A}_{\mu_k}\mathbf{x}_k + \mathbf{G}_{\mu_k}\mathbf{w}_k \\ \mathbf{z}_k =& \mathbf{C}_{\mu_k}\mathbf{x}_k + \mathbf{H}_{\mu_k}\mathbf{v}_k\end{aligned}$$

$\mathbf{z}_k \in \mathbb{R}^{n_z}$ – measurements, $\bar{\mathbf{x}}_k^T = [\mathbf{x}_k^T, \mu_k]$ – system state,
$\mathbf{x}_k \in \mathbb{R}^{n_x}$ – common continuous state,
$\mu_k \in \mathcal{M} = \{1, 2, \dots, N\}$ – a scalar index into the set of models
$P(\mu_{k+1} = j | \mu_k = i) = \pi_{ij}$ – transition probabilities
$\mathbf{w}_k$, $\mathbf{v}_k$ – noises with standard Gaussian distribution $\mathcal{N}\{\mathbf{0}, \mathbf{E}\}$
$\mathbf{x}_0$ – initial state with Gaussian distribution $\mathcal{N}\{\mathbf{x}_{0|-1}, \mathbf{P}_{0|-1}\}$
μ_0 – initial model with probabilities $P(\mu_0)$
$\mathbf{A}_{\mu_k}$, $\mathbf{G}_{\mu_k}$, $\mathbf{C}_{\mu_k}$, and $\mathbf{H}_{\mu_k}$ – given matrices

Problem formulation – cont'd

Description of the optimal detector at time steps $k \in \mathcal{T}$



$$\mathbf{D}\text{: } d_k = \sigma_k\left(\mathbf{I}_0^k\right)$$

d_k – a decision at the time step k,
$\sigma_k(\mathbf{I}_0^k)$ – an unknown function that describes the detector
$\mathbf{I}_0^k = \left[\mathbf{z}_0^k, d_0^{k-1}\right]$ – all available information at the time step k

Problem formulation – cont'd

A criterion for non-delayed decisions

$$J\left(\sigma_0^F\right) = \mathrm{E}\left\{\sum_{k=0}^{F} L_k^{\mathrm{d}}\left(\mu_k, d_k\right)\right\}$$

$L_k^{\mathrm{d}}(\mu_k, d_k)$ – a cost function that assesses the decision d_k with respect to the current model μ_k

Šimandl, M. and Punčochář, I. (2009)

Active fault detection and control: Unified formulation and optimal design. *Automatica*, 45(9), 2052–2059.

Problem formulation – cont'd

A criterion for decisions delayed by $\ell \geq 0$ steps

$$J\left(\sigma_\ell^F\right) = \mathrm{E}\left\{\sum_{k=\ell}^{F} L_k^{\mathrm{d}}\left(\mu_{k-\ell}, d_k\right)\right\}$$

$L_k^{\mathrm{d}}(\mu_{k-\ell}, d_k)$ – a cost function that assesses the decision d_k with respect to the past model $\mu_{k-\ell}$

Note: The detector is not defined at the time steps 0 to $\ell - 1$.

An example of the function $L_k^{\mathrm{d}}(\mu_{k-\ell}, d_k)$

$$L_k^{\mathrm{d}}(\mu_{k-\ell}, d_k) = \begin{cases} 0 & \text{if } d_k = \mu_{k-\ell} \\ 1 & \text{otherwise} \end{cases}$$

Optimal detector design – cont'd

Three fundamental information processing strategies

Open Loop (OL) – Only an a priori information is used.

Open Loop Feedback (OLF) – An a priori information and measurements received up to the current time step are used. No further measurements will be received in the future.

Closed Loop (CL) – An a priori information and measurements received up to the current time step are used. Further measurements will be received and utilized in the future.

$$J^{\mathrm{OL}} \geq J^{\mathrm{OLF}} \geq J^{\mathrm{CL}}$$

Optimal detector design – cont'd

Three fundamental information processing strategies

Open Loop (OL) – Only an a priori information is used.

Open Loop Feedback (OLF) – An a priori information and measurements received up to the current time step are used. No further measurements will be received in the future.

Closed Loop (CL) – An a priori information and measurements received up to the current time step are used. Further measurements will be received and utilized in the future.

$$J^{\mathrm{OL}} \geq J^{\mathrm{OLF}} = J^{\mathrm{CL}}$$

Optimal detector design – cont'd

Backward recursive equation $k = F, F-1, \ldots, 0$

$$V_k^*\left(\mathbf{z}_0^k\right) = \underbrace{\min_{d_k \in \mathcal{M}} \mathrm{E}\left\{L_k^{\mathrm{d}}\left(\mu_{k-\ell}, d_k\right) + V_{k+1}^*\left(\mathbf{z}_0^{k+1}\right) \mid \mathbf{z}_0^k, d_k\right\}}_{d_k^* = \arg\min_{d_k \in \mathcal{M}}}$$

- $V_k^*(\mathbf{z}_0^k)$ – the cost-to-go (Bellman) function
- $V_{F+1}^* = 0$ – the initial condition
- $J^{\mathrm{CL}} = J(\sigma_\ell^{F*}) = \mathrm{E}\{V_0^*(\mathbf{z}_0)\}$

Optimal detector design – cont'd

Optimal detector law

$$d_k^* = \sigma_k^*\left(\mathbf{z}_0^k\right) = \arg\min_{d_k \in \mathcal{M}} \mathrm{E}\left\{ L_k^{\mathrm{d}}\left(\mu_{k-\ell}, d_k\right) \,\middle|\, \mathbf{z}_0^k, d_k \right\}$$

- It is not necessary to compute the cost-to-go function $V_k^*(\mathbf{z}_0^k)$ because the decision d_k does not influence the future costs
- The smoothing probability $P(\mu_{k-\ell}|\mathbf{z}_0^k)$ is needed for evaluation the conditional expectation $\mathrm{E}\{\cdot|\cdot\}$

Filtering and smoothing algorithms

Filtering algorithm

- The aim is to compute the probability $P(\mu_k|\mathbf{z}_0^k)$ and the pdf $p(\mathbf{x}_k|\mathbf{z}_0^k)$
- Given the model sequences μ_0^k the Kalman filters are used to compute the pdfs $p(\mathbf{x}_k|\mathbf{z}_0^k, \mu_0^k)$
- Then filtering probability is given as $P(\mu_k|\mathbf{z}_0^k) = \sum_{\mu_0^{k-1}} P(\mu_0^k|\mathbf{z}_0^k)$, where

$$P(\mu_0^k|\mathbf{z}_0^k) = \frac{p\left(\mathbf{z}_k|\mathbf{z}_0^{k-1}, \mu_0^k\right) P\left(\mu_k|\mu_{k-1}\right) P\left(\mu_0^{k-1}|\mathbf{z}_0^{k-1}\right)}{p\left(\mathbf{z}_k|\mathbf{z}_0^{k-1}\right)}$$

Filtering and smoothing algorithms – cont'd

Smoothing algorithm

- The aim is to compute the probability $P(\mu_{k-\ell}|\mathbf{z}_0^k)$ and optionally the pdf $p(\mathbf{x}_{k-\ell}|\mathbf{z}_0^k)$ for $\ell > 0$
- Given the model sequences μ_0^k the Rauch-Tung-Striebel smoothers are used to obtain the pdfs $p(\mathbf{x}_{k-\ell}|\mathbf{z}_0^k, \mu_0^k)$
- The smoothing probability $P(\mu_{k-\ell}|\mathbf{z}_0^k)$ can directly be computed by marginalization as

$$P(\mu_{k-\ell}|\mathbf{z}_0^k) = \sum_{\mu_0^{k-\ell-1}, \mu_{k-\ell+1}^k} P(\mu_0^k|\mathbf{z}_0^k)$$

Filtering and smoothing algorithms – cont'd

Notes on estimation algorithms

- The number of sequences increases according to N^{k+1}
- Merging with depth $h \geq \ell$ based on the moment matching technique is used to limit computational demands

$$P\left(\mu_{k-h}^k | \mathbf{z}_0^k\right) = \sum_{\mu_0^{k-h-1}} P\left(\mu_0^k | \mathbf{z}_0^k\right)$$

$$p\left(\mathbf{x}_k | \mathbf{z}_0^k, \mu_{k-h}^k\right) = \sum_{\mu_0^{k-h-1}} P\left(\mu_0^{k-h-1} | \mathbf{z}_0^k, \mu_{k-h}^k\right) \times p\left(\mathbf{x}_k | \mathbf{z}_0^k, \mu_0^k\right) \approx \mathcal{N}\{\tilde{\mathbf{x}}, \tilde{\mathbf{P}}\}$$

A second order system described by two models

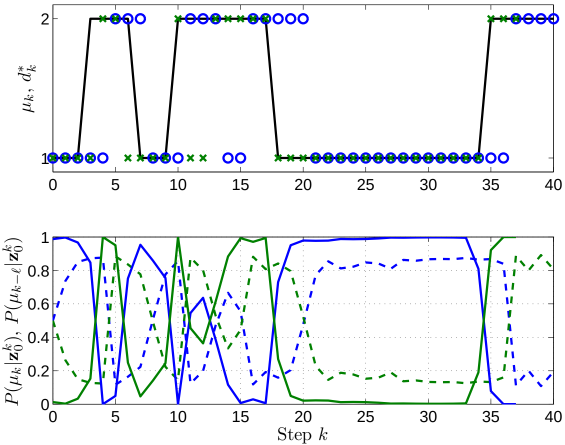
$$\mathbf{A}_1 = \begin{bmatrix} 0.9 & 1 \\ 0 & 0.9 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 0.8 & 1 \\ 0 & 0.9 \end{bmatrix},$$
$$\mathbf{G}_1 = 0.01\mathbf{E}_2, \qquad \mathbf{G}_2 = 0.1\mathbf{E}_2,$$
$$\mathbf{C}_1 = \mathbf{C}_2 = [1\ 0], \qquad \mathbf{H}_1 = \mathbf{H}_2 = 0.01$$

- Horizon $F = 40$
- Initial condition $\mathbf{x}_{0|-1} = [1\ 0]^T$, $\mathbf{P}_{0|-1} = 0.1\mathbf{E}_2$
- Initial probabilities $P(\mu_0 = 1) = P(\mu_0 = 2) = 0.5$
- Transition probabilities $\pi_{1,1} = \pi_{2,2} = 0.95$
- The cost function $L_k^{\mathrm{d}}(\mu_{k-\ell}, d_k) = L_k^{\mathrm{d1}}(\mu_{k-\ell}, d_k) + L^{\mathrm{d2}}\ell$
  - $L_k^{\mathrm{d1}}(\mu_{k-\ell}, d_k)$ – the zero-one cost function
  - L^{d2} – a constant cost of deferring the decision by one time step

Numerical example (Scenario 1)

Comparison of decisions based on filtering and smoothing

- Comparison within an individual realization of random process
- Chosen parameters
  - The cost of deferring decision $L^{\mathrm{d}2} = 0.01$
  - The depth for merging $h = 3$ and the lag $\ell = 3$

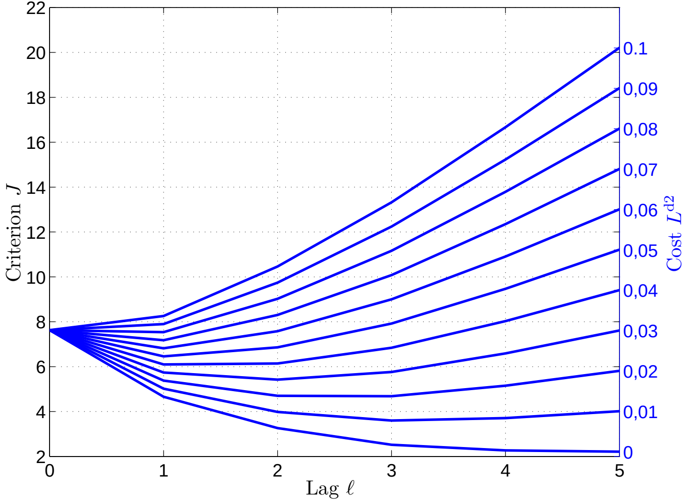


True model – black line, Filtering decisions/probabilities – blue circles/lines, Smoothing decisions/probabilities – green x-marks/lines

Numerical example (Scenario 2)

Dependence of the criterion J on the lag ℓ

- The value of the criterion J is evaluated on the interval 0 to $F - \ell_{\max}$ using 1000 Monte Carlo simulations
- The maximum considered lag $\ell_{\max} = 5$
- The depth for merging $h = \ell_{\max}$
- Considered costs of deferring decision by one time step $L^{\mathrm{d2}} = \{0, 0.01, 0.02, \ldots, 0.1\}$


22
20
18
16
14
12
10
8
6
4
2
Criterion J
0
1
2
3
4
5
Lag ℓ
0.1
0,09
0,08
0,07
0,06
0,05
0,04
0,03
0,02
0,01
0
Cost L^{d2}

Concluding Remarks

- The core idea – delay decisions, gather more measurements and thus improve change detection
- An innovative formulation of considered problem was provided and the new optimal detector with deferred decisions was derived using closed loop information processing strategy
- The approach was applied in change detection with multiple linear Gaussian models
- It was demonstrated that the quality of decisions increases as the lag increases when cost of deferring the decisions is zero