F2E5216/TS1002 Adaptive Filtering and Change Detection

Fredrik Gustafsson (LiTH) and Bo Wahlberg (KTH)

Lecture 8

Filter Banks for State Changes

- Explicit modeling of additive change: GLR and MLR
- Multiple models: pruning, merging and off-line algorithms

Likelihood Ratio based Change Detection Tests

Hypothesis test:

$$\begin{aligned} H_0 &: \quad \text{no jump} \\ H_1(k,\nu) &: \quad \text{a jump of magnitude } \nu \text{ at time } k. \end{aligned}$$

Likelihood ratio: In previous notation,

$$g_t(k) = \frac{p(y^k)p(y^t_{k+1})}{p(y^t)}$$

- $g_t(k)$ is just a normalized version of the likelihood.
- $g_t(k)$ is a **distance measure** between H_0 and $H_1(k)$.
- $\nu = \theta_1$ when $\theta_0 = 0$ is assumed.

Recursive Formulation

$$\begin{aligned} g_t(k) &= \frac{p(y^k)p(y^t_{k+1})}{p(y^t)} \\ &= g_{t-1}(k)\frac{p(y_t|y^{t-1}_{k+1})}{p(y_t|y^{t-1})} \end{aligned}$$

or in the negative logarithm

$$\underbrace{-\log g_t(k)}_{\bar{g}_t(k)} = \underbrace{-\log g_{t-1}(k)}_{\bar{g}_{t-1}(k)} + \underbrace{(-\log p(y_t|y^{t-1}_{k+1}) + \log p(y_t|y^{t-1}))}_{\bar{s}_t(k)}$$

Fits the general stopping rule framework.

Gaussian Case

The jump ν can be ML estimated (the *generalized likelihood ratio* test) or marginalized (the *marginalized likelihood ratio* test)

$$\begin{aligned} g_t^{GLR}(k) &= \frac{\hat{\nu}^2(k)}{R/(t-k)} \underset{H_1}{\overset{H_0}{\lessgtr}} h \\ g_t^{MLR}(k) &= \frac{\hat{\nu}^2(k)}{R/(t-k)} - \log(2\pi R) \underset{H_1}{\overset{H_0}{\lessgtr}} 0 \end{aligned}$$

The noise variance R is assumed known.
Remark 1: It is the product Rh that determines the performance of GLR.
Remark 2: There is no threshold to design in MLR (implicitly given by R).

Implementation Aspects

All $0 < k < t$ are involved in the test.

Approximation 1: Consider only change times in a sliding window $t - L < k < t$.

Approximation 2: Consider only one change time $k = t - L$ (Brandt's GLR).

Off-line algorithm:

1. Forward filter computes $p(y^k), \forall k$.
2. Backward filter computes $p(y_{k+1}^N), \forall k$.
3. MLR combines these as $\frac{p(y^k)p(y_{k+1}^N)}{p(y^N)}$.

Data Models

Explicit modeling of additive **pulse** change (Ch. 9 and 11):

$$\begin{aligned} x_{t+1} &= A_t x_t + B_{u,t} u_t + B_{v,t} v_t + \delta_{t-k} B_\theta \nu \\ y_t &= C_t x_t + e_t + D_{u,t} u_t + \delta_{t-k} D_{\theta,t} \nu. \end{aligned}$$

Step changes are modeled by changing notation $\delta \leftrightarrow \sigma$ (step function).

Multiple models with *mode* parameter δ, usually 0 or 1 in Ch. 10, or Markov chain in *jump Markov models*

$$\begin{aligned} x_{t+1} &= A_t(\delta) x_t + B_{u,t}(\delta) u_t + B_{v,t}(\delta) v_t \\ y_t &= C_t(\delta) x_t + D_{u,t}(\delta) u_t + e_t \\ v_t &\in \mathrm{N}(m_{v,t}(\delta), Q_t(\delta)) \\ e_t &\in \mathrm{N}(m_{e,t}(\delta), R_t(\delta)). \end{aligned}$$

A Direct Approach

Assume **step** changes. Augmented state space model

$$\begin{aligned} \bar{x}_{t+1} &= \begin{pmatrix} x_{t+1} \\ \theta_{t+1} \end{pmatrix} = \begin{pmatrix} A_t & B_{\theta,t} \\ 0 & I \end{pmatrix} \bar{x}_t + \begin{pmatrix} B_{u,t} \\ 0 \end{pmatrix} u_t \\ &+ \begin{pmatrix} B_{v,t} & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} v_t \\ \delta_{t-(k-1)} \nu \end{pmatrix} \\ y_t &= \begin{pmatrix} C_t & D_{\theta,t} \end{pmatrix} \bar{x}_t + e_t + D_{u,t} u_t \\ \bar{x}_{0|0} &= \begin{pmatrix} x_0 \\ 0 \end{pmatrix} \\ \bar{P}_{0|0} &= \begin{pmatrix} P_0 & 0 \\ 0 & 0 \end{pmatrix} \end{aligned}$$

Adaptive Filter or Whiteness Test Approach

Disregards explicit use of δ_{t-k} changes. Parameter (change) estimator:

$$\hat{\theta}_{t+1|t} = \hat{\theta}_{t|t-1} + K_t^\theta (y_t - C_t \hat{x}_{t|t-1} - D_{\theta,t} \hat{\theta}_{t|t-1} - D_{u,t} u_t),$$

$$K_t = \begin{pmatrix} K_t^x \\ K_t^\theta \end{pmatrix}, \quad P_t = \begin{pmatrix} P_t^{xx} & P_t^{x\theta} \\ P_t^{\theta x} & P_t^{\theta\theta} \end{pmatrix}.$$

- Adaptive filtering with state noise covariance

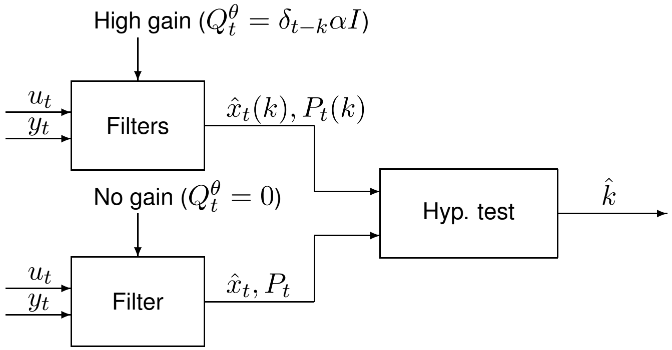
$$\bar{Q}_t = \begin{pmatrix} Q_t & 0 \\ 0 & Q_t^\theta \end{pmatrix}$$

- Whiteness based residual test, where Q_t^θ is momentarily increased when a change is detected.

Multiple-Model Approach

Run N matched filters (standard KF) to each hypothesis $H_1(k)$.

Compare likelihoods (or likelihood ratios) computed from $\varepsilon_t(k)$ and $S_t(k)$.



Idea of GLR

Kalman filter matched to $H_0 \;\longrightarrow\; \hat{x}_t,\; K_t$ (gain), $\varepsilon_t,\; S_t = \mathrm{Cov}(\varepsilon_t)$

Kalman filter matched to $H_1(k) \;\longrightarrow\; \hat{x}_t(k),\; \varepsilon_t(k),\; \varphi_t(k),\; \mu_t(k)$

Identification under $H_1(k) \;\longrightarrow\; \varepsilon_t(k) = \varphi_t^T(k)\nu(k) + e_t$

Compensation under $H_1(k) \;\longrightarrow\; \hat{x}_t(k) \approx \hat{x}_t + \mu_t(k)\nu(k)$.

- Note: linear regression for change magnitude!
- Need: one KF and t RLS filters $\Rightarrow \hat{\nu}(k)$
- First: update equations for $\varepsilon_t(k)$ and $\mu_t(k)$.

GLR Lemma

Linear model $\longrightarrow$ influence of change linear $\longrightarrow$ **postulate**

$$\begin{aligned}\hat{x}_{t|t}(k) &= \hat{x}_{t|t} + \mu_t(k)\nu \\ \varepsilon_t(k) &= \varepsilon_t + \varphi_t^T(k)\nu.\end{aligned}$$

Lemma Update recursion

$$\begin{aligned}\varphi_{t+1}^T(k) &= C_{t+1}\left(\prod_{i=k}^{t} A_i - A_t\mu_t(k)\right) \\ \mu_{t+1}(k) &= A_t\mu_t(k) + K_{t+1}\varphi_{t+1}^T(k),\end{aligned}$$

initialized by $\mu_k(k) = 0$ and $\varphi_k(k) = 0$.

GLR Algorithm

Main filter: Kalman filter assuming no jump.

Filter bank: Regressors $\varphi_t(k)$ and the LS quantities
$R_t(k) = \sum_{i=1}^{t} \varphi_i(k) S_i^{-1} \varphi_i^T(k)$ and
$f_t(k) = \sum_{i=1}^{t} \varphi_i(k) S_i^{-1} \varepsilon_i$ *for each* k, $1 \le k \le t$.

GLR Test: At time $t = N$, the test statistic is given by
$l_N(k, \hat{\nu}(k)) = f_N^T(k) R_N^{-1}(k) f_N(k)$.
A jump candidate is given by $\hat{k} = \arg\max l_N(k, \hat{\nu}(k))$.
It is accepted if $l_N(\hat{k}, \hat{\nu}(\hat{k})) > h$

Identification: $\hat{\nu}_N(\hat{k}) = R_N^{-1}(\hat{k}) f_N(\hat{k})$.

Comments on GLR

- The system is (often) not *persistently excited*. That is, φ_t decays to zero. Intuitively, this means that the KF compensates itself, making identification of ν unnecessary after a while.
- Test statistic χ^2 distributed.
- Regressors pre-computable, decay rather fast to zero for many systems and depend only on $t-k$ for time-invariant systems. $\longrightarrow$ Efficient implementations might exist.
- RLS better to use $\longrightarrow$ matrix inversion of $R_N(k)$ not needed:

$$l_t(k, \hat{\nu}(k)) = f_t^T(k)\hat{\nu}_t(k),$$

MLR versus GLR

- In GLR, the threshold is sensitive to incorrectly specified noise scalings (which does not affect the KF).

$$\overline{R} = \lambda R, \quad \overline{P}_0 = \lambda P_0, \quad \overline{Q} = \lambda Q \Rightarrow \bar{l}_N(k) = l_N(k)/\lambda \underset{H_1}{\overset{H_0}{\lessgtr}} h.$$

In MLR, there is no threshold. The noise scaling can be incorporated as a nuissance parameter.

- Complexity. GLR requires N^2 filter updates. Sliding window approximation requires NL filter updates. Two-filter MLR requires $2N$ filter updates.

Multiple Models

$$\begin{aligned}
x_{t+1} &= A_t(\delta)x_t + B_{u,t}(\delta)u_t + B_{v,t}(\delta)v_t \\
y_t &= C_t(\delta)x_t + D_{u,t}(\delta)u_t + e_t \\
v_t &\in \mathrm{N}(m_{v,t}(\delta), Q_t(\delta)) \\
e_t &\in \mathrm{N}(m_{e,t}(\delta), R_t(\delta)).
\end{aligned}$$

Discrete parameter δ is the mode, or discrete state, of the system.

1. δ_t has S possible outcomes. Mostly $S=2$ considered.

2. δ_t has S Markov states with transition matrix Π (jump Markov models, hidden Markov model (HMM)). Difficult on-line. EM-algorithm or Baum-Welch method off-line.

Example 1: $Q(\delta) = (1+9\delta)Q_0$ can be used to model additive state changes implicitly.

Example 2: $A(\delta)$ can be used to model different turn rates in target tracking, with $(x_1, \dot{x}_1, x_2, \dot{x}_2)^T$ as states.

Formulation incorporates: change detection, segmentation, model structure selection, equalization, blind equalization, outliers and missing data!

Basic Strategy

1. Conditional Kalman filter given the mode sequence gives

$$\hat{x}_{t|t}(\delta^t), \quad P_{t|t}(\delta^t).$$

2. Compute the posterior probability of the mode sequence

$$p(\delta^t|y^t).$$

3. There are S^t different sequences δ^t, labelled $\delta^t(i)$, $i = 1, 2, ..., S^t$. Theorem of total probability gives the Gaussian mixture:

$$p(x_t|y^t) = \frac{1}{\sum_{i=1}^{S^t} p(\delta^t(i)|y^t)} \sum_{i=1}^{S^t} p(\delta^t(i)|y^t) \, \mathrm{N}\left(\hat{x}_{t|t}(\delta^t(i)), P_{t|t}(\delta^t(i))\right).$$

Approximations

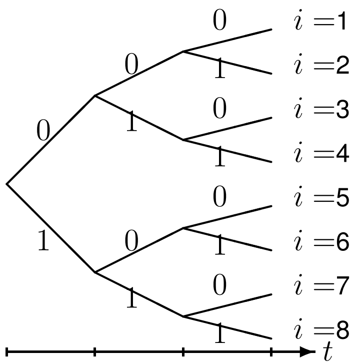
$$p(x_t|y^t) = \frac{1}{\sum_{i=1}^{S^t} p(\delta^t(i)|y^t)} \sum_{i=1}^{S^t} p(\delta^t(i)|y^t) \, \mathrm{N}\left(\hat{x}_{t|t}(\delta^t(i)), P_{t|t}(\delta^t(i))\right).$$

4. **On-line**: Merging (`imm`) Add overlapping distributions $N\left(\hat{x}_{t|t}(\delta^t(i)), P_{t|t}(\delta^t(i))\right)$

Pruning sequences (`detectM`) Remove components with small coefficients $p(\delta^t(i)|y^t)$

5. **Off-line**: numerical approaches based on the EM algorithm and MCMC methods (`mcmc, gibbs`).

Pruning versus Merging



Pruning: cut off branches.
Merging: represent several branches by one.

A Merging Formula

The best approximation of a sum of L Gaussian distributions

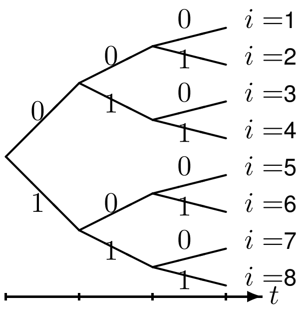
$$p(x) = \sum_{i=1}^{L} \alpha(i) \, \mathrm{N}(\hat{x}^j, P^j) \approx \alpha \, \mathrm{N}(\hat{x}, P),$$

where
$$\alpha = \sum_{i=1}^{L} \alpha(i), \quad \hat{x} = \frac{1}{\alpha} \sum_{i=1}^{L} \alpha(i)\hat{x}(i)$$
$$P = \frac{1}{\alpha} \sum_{i=1}^{L} \alpha(i) \left(P(i) + (\hat{x}(i) - \hat{x})(\hat{x}(i) - \hat{x})^T\right)$$

Second term: *spread of the mean.*

GPB Merging Strategy

Generalized Pseudo Bayesian



GPB(n): n is the size of sliding memory ($n = 0$ standard)
GPB(0): merge all sequences (1-8).
GPB(1): merge sequences (1,3,5,7) and (2,4,6,8).
GPB(2): merge sequences (1,5), (2,6), (3,7) and (4,8).

IMM

Interacting Multiple Model (IMM) by Bar-Shalom and Li. Essentially as GPB, but merging after time update, rather than after measurement update.

Summary: State Detection

Abrupt state changes can be detected and isolated with either:

- Likelihood ratio (MLR, GLR) hypothesis test, using the statistical approach.
- Multiple models (IMM,GPB)

Exercises:

41, 42 (should be (8.100) in 2000-edition), 43.

Next Time

Parity space change detection (deterministic approach)