F2E5216/TS1002 Adaptive Filtering and Change Detection

Fredrik Gustafsson (LiTH) and Bo Wahlberg (KTH)

Linköpings universitet

Lecture 5

Change detection based on sliding windows (Chapter 6)

- Distance measures
- Detection and Isolation

Model Validation

Data are taken from

- a sliding window (typical application).

$$\text{Data}: \underbrace{y_1, y_2, ..., y_{t-L}}_{\text{Model } \theta_0}, \underbrace{y_{t-L+1}, .., y_t}_{\text{Model } \theta \text{ ,Data } Y}$$

- an increasing window.

Nominal model (parameter vector θ_0) known from:

- recursive identification from past data.
- a nominal model, obtained by physical modeling or system identification.

Data Model

Linear regression (in vector notation, $\Phi = (\varphi_{t-L+1}, \varphi_2, ..., \varphi_t)^T$).

$$Y = \Phi\theta + E,$$

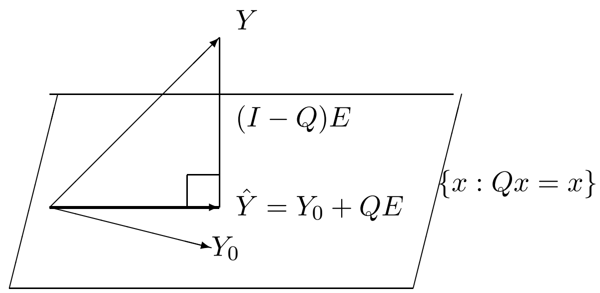
Least squares estimate:

$$\hat{\theta} = (\Phi^T\Phi)^{-1}\Phi^T Y, \quad \hat{Y} = \Phi\hat{\theta} = \underbrace{\Phi(\Phi^T\Phi)^{-1}\Phi^T}_{Q} Y$$

Nominal data $Y_0 = \Phi\theta_0$,

Notice that $QY_0 = Y_0$, since $Q\Phi = \Phi$

Geometric Interpretations



General Form of Consistency Tests

1: Parameter estimation error

$$\hat{\theta} = (\Phi^T\Phi)^{-1}\Phi^T Y \quad \Rightarrow$$
$$(\Phi^T\Phi)^{-1}\Phi^T Y - \theta_0 \in \mathrm{N}(\tilde{\theta}, P)$$

2: Norm of the simulation error

$$\|Y_0 - Y\|_Q^2 = (Y_0 - Y)^T Q (Y_0 - Y)$$

where Q is a projection matrix ($QQ = Q$).

Ex: Linear Regression and Model Differences

Simulation difference between nominal and estimated model

$$\begin{aligned} V^{MD} &= \|\Phi\theta_0 - \Phi\hat{\theta}\|_2^2 \\ &= \|QY_0 - QY\|_2^2 \\ &= \|Y_0 - Y\|_Q^2 \end{aligned}$$

Ex: Linear Regression and Parameter Norm

Standard calculations (with deterministic Φ) give

$$\tilde{\theta} = \hat{\theta} - \theta_0 = \left(\Phi^T\Phi\right)^{-1}\Phi^T Y - \theta_0 \in \mathrm{N}(0, \underbrace{\lambda\left(\Phi^T\Phi\right)^{-1}}_{P}), \quad \text{under } H_0.$$

Test statistic:

$$\|\tilde{\theta}\|_{P^{-1}}^2 \in \chi^2(d), \quad \text{under } H_0,$$

The test:

$$\|\tilde{\theta}\|_{P^{-1}}^2 \underset{H_1}{\overset{H_0}{\lessgtr}} h_\alpha.$$

Example, cont'd

Alternative formulation with

$$Y_0 = \Phi\theta_0 \Rightarrow \theta_0 = \left(\Phi^T\Phi\right)^{-1}\Phi^T Y_0 \quad \Rightarrow$$

$$\begin{aligned} V^{PN} &= \|\tilde{\theta}\|_{P^{-1}} \\ &= (Y_0 - Y)^T\Phi\left(\Phi^T\Phi\right)^{-1}\frac{1}{\lambda}\left(\Phi^T\Phi\right)\left(\Phi^T\Phi\right)^{-1}\Phi^T(Y_0 - Y) \\ &= \frac{1}{\lambda}(Y_0 - Y)^T\underbrace{\Phi\left(\Phi^T\Phi\right)^{-1}\Phi^T}_{Q}(Y_0 - Y) \\ &= \frac{1}{\lambda}(Y_0 - Y)^T Q (Y_0 - Y). \end{aligned}$$

Warning: Never use $\|\tilde{\theta}\|^2$ or $\|Y - Y_0\|^2$

Ex: GLR and Linear Regression

Likelihood ratio assuming Gaussian noise:

$$LR = \frac{P(Y|\theta)}{P(Y|\theta_0)} = \frac{(2\pi\lambda)^{-L/2} e^{-\frac{\|Y-\Phi\theta\|^2}{2\lambda}}}{(2\pi\lambda)^{-L/2} e^{-\frac{\|Y-\Phi\theta_0\|^2}{2\lambda}}}$$

log likelihood ratio

$$2\lambda \cdot LLR = \|Y - \Phi\theta_0\|^2 - \|Y - \Phi\theta\|^2.$$

Generalized Likelihood Ratio:

$$\begin{aligned} V^{GLR} &= 2\lambda \cdot GLR = \max_\theta 2\lambda \cdot LLR \\ &= \|Y - \Phi\theta_0\|^2 - \|Y - \Phi\hat{\theta}\|^2 = ... = \\ &= (Y_0 - Y)^T Q (Y_0 - Y). \end{aligned}$$

Summary of Examples

Gaussian noise or asymptotic arguments lead to

$$\|Y_0 - Y\|_Q^2 \in \chi^2(d)$$

and

$$\tilde{\theta} \in \mathrm{N}(0, P) \Leftrightarrow \tilde{\theta}^T P^{-1} \tilde{\theta} \in \chi^2(d)$$

when the noise variance is known.

More General Distance Functions

Assume Gaussian noise and consider the noise variance λ as a parameter (here and in the book N=L).

Simulation error

$$\frac{\|Y - \Phi\theta_0\|^2}{\lambda_0} - N.$$

Generalized likelihood ratio test

$$N \log \frac{\lambda_0}{\lambda_1} + \frac{\|Y - \Phi\theta_0\|^2}{\lambda_0} - \frac{\|Y - \Phi\theta_1\|^2}{\lambda_1}$$

The Kullback Divergence test

$$N\left(\frac{\lambda_0}{\lambda_1} - 1\right) + \left(1 + \frac{\lambda_0}{\lambda_1}\right) \frac{\|Y - \Phi\theta_0\|^2}{\lambda_0} - 2\frac{(Y - \Phi\theta_0)^T (Y - \Phi\theta_1)}{\lambda_1}$$

The Asymptotic Local Approach

Hypothesis test:

$$\begin{aligned} H_0 &: \text{No change} \quad \theta_L = \theta_0 \\ H_1 &: \text{Change} \quad \theta_L = \theta_0 + \frac{1}{\sqrt{L}}\nu. \end{aligned}$$

Why decreasing change??

- There is no physical reason
- The test is made more sensitive when the number of data increases. The estimated change ν will have covariance of constant size.

The asymptotic local approach

- is standard in statistics
- can be generalized to *e.g.* non-linear models.

See Basseville and Nikiforov: *Detection of abrupt changes*, 1993.

The Asymptotic Local Approach: Linear Regression

Model: $y_t = \varphi_t^T \theta + e_t$

$H_0: \quad \theta_L = \theta_0,$

$H_1: \quad \theta_L = \theta_0 + \frac{1}{\sqrt{L}}\nu$

Data: $Z_t = (\varphi_t^T, y_t)^T$

Primary residual: $K(Z_t, \theta_0) = \varphi_t(y_t - \varphi_t^T \theta_0)$, such that $\mathsf{E}(K(Z_t, \theta_0)) = 0$ under H_0, and nonzero under H_1.

Quasi-score:

$$n_L(\theta_0) = \frac{1}{\sqrt{L}} \sum_{t=1}^{L} \varphi_t(y_t - \varphi_t^T \theta_0)$$

$$n_L(\theta_0) = \frac{1}{\sqrt{L}} \sum_{t=1}^{L} \varphi_t e_t, \quad \text{under } H_0$$

$$n_L(\theta_0) = \frac{1}{L} \left(\sum_{t=1}^{L} \varphi_t \varphi_t^T \right) \nu + \frac{1}{\sqrt{L}} \sum_{t=1}^{L} \varphi_t e_t \quad \text{under } H_1$$

Observe that $\frac{1}{L}\sum_{t=1}^{L} \varphi_t \varphi_t^T \approx \mathsf{E}(\varphi_t \varphi_t^T) = M$ is almost independent of the data length L.

Let $\Sigma = \mathsf{E}(n_L(\theta_0) n_L^T(\theta_0)) = \lambda \mathsf{E}(\varphi_t \varphi_t^T)$

Asymptotic Distribution

$$n_L(\theta_0) \in \begin{cases} \text{AsN}(0, \Sigma) & \text{under } H_0 \\ \text{AsN}(M\nu, \Sigma) & \text{under } H_1 \end{cases}$$

For a scalar $n_L(\theta_0)$ any standard test can be used.

For the vector case apply a χ_2 test on

$$\Sigma^{-1/2} n_L(\theta_0)$$

Should be modified if $\nu = T\bar{\nu}$, where dim$\bar{\nu} <$ dimθ, see page 217.

Isolation

Split parameter vector.

$$Y = \Phi^T \theta = ((\Phi^a)^T, \ (\Phi^b)^T) \begin{pmatrix} \theta^a \\ \theta^b \end{pmatrix}$$

Detection	$H_0: \ \theta = \theta_0$	$H_1: \ \theta \neq \theta_0$
Isolation	$H_0: \ \theta^a = \theta_0^a$	$H_1^a: \ \theta^a \neq \theta_0^a$

Assume that $\theta_0 = 0$, otherwise replace measurements by nominal residuals $Y - \Phi\theta_0$.

Diagnosis: Multiple hypothesis test of different θ^a.

Alt. 1: Isolation (GLR)

$\theta^b = 0$, so the fault in θ^a does not influence the other elements in the parameter vector (extreme marginalization).

$$\begin{aligned} V^1 &= -2\log\left(\frac{\max_{\theta^a} p\left(Y|\begin{pmatrix}\theta^a\\0\end{pmatrix}\right)}{p\left(Y|\begin{pmatrix}0\\0\end{pmatrix}\right)}\right) \\ &= Y^T \underbrace{\Phi^a\left((\Phi^a)^T\Phi^a\right)^{-1}(\Phi^a)^T}_{Q^a} Y \\ &= Y^T Q^a Y, \end{aligned}$$

which is the simulation error projected on the subspace generated by the fault hypothesis.

Alt. 2: Isolation (GLR)

θ^b is nuisance, and its value is unknown and irrelevant.

$$\begin{aligned} V^2 &= -2\log\left(\frac{\max_{\theta^a,\theta^b} p\left(Y|\begin{pmatrix}\theta^a\\\theta^b\end{pmatrix}\right)}{\max_{\theta^b} p\left(Y|\begin{pmatrix}0\\\theta^b\end{pmatrix}\right)}\right) \\ &= Y^T Q Y - Y^T Q^b Y \\ &= Y^T (Q - Q^b) Y \end{aligned}$$

Remarks

- Q^a and Q^b are two subspaces in Q
- Geometrical interpretation: V^1 is the part of V that belongs to Q^a. V^2 is V subtracted by the part that belongs to Q^b.
- $Q - Q^b$ is no projection matrix.
- We have $Q \geq Q^b + Q^a$, with equality only when $\Phi^a \Phi^b = 0$. This means that the second alternative gives a smaller test statistic and a more cautious test.

Shortcoming:

- There is no penalty in $\dim \theta$, which implies that nested fault models cannot be handled.

Diagnosis of Sensor Faults

Simple model with two measurements:

$$y_t = \begin{pmatrix}\theta^a\\\theta^b\end{pmatrix} + e_t = \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\begin{pmatrix}\theta^a\\\theta^b\end{pmatrix} + e_t.$$

Let $Y = (y_{t-L+1}^T, \dots, y_t^T)^T$, $\Phi = \begin{pmatrix}1 & 0 & \dots & 1 & 0\\0 & 1 & \dots & 0 & 1\end{pmatrix}$

Nominal model: no sensor offset $\theta_0 = 0$ and $Y_0 = 0$.

1. Detection:

$$Q = \frac{1}{L}\Phi^T\Phi \Rightarrow V = Y^T Q Y = \frac{1}{L}\sum_{i=t-L+1}^{t}\sum_{j=1}^{2}(y_i^{(j)})^2.$$

which is $\chi_2(2L)$ under H_0.

2. Isolation: Both approaches give the same result:

$$Q^a = \frac{1}{L}(1\ 0 \dots 1\ 0)^T (1\ 0 \dots 1\ 0) \quad \Rightarrow \quad Y^T Q^a Y = \sum_{i=t-L+1}^{t} (y_i^{(1)})^2$$

$$Q^b = \frac{1}{L}(0\ 1 \dots 0\ 1)^T (0\ 1 \dots 0\ 1) \quad \Rightarrow \quad Y^T Q^b Y = \sum_{i=t-L+1}^{t} (y_i^{(2)})^2.$$

Fault	θ^a	θ^b
V^a	"large"	"small"
V^b	"small"	"large"

3. Fault identification: $\hat{\theta}^a = \frac{1}{L} \sum_{i=t-L+1}^{t} y_i^{(1)}$

Likelihood Based Methods

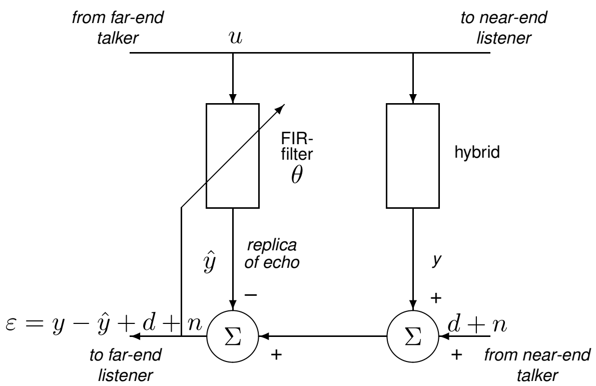
Compute and compare likelihoods of **any** hypothesis

$$p(Y|H_i) = p(Y|\theta = \theta_i, \lambda = \lambda_i).$$

Marginalization or maximization of nuisance parameters.
All formulas are found in the appendices of Chapter 5!

Detection and isolation jointly solved. See application example below.

Echo Path change and Double Talk Detection



FIR echo cancellation works well except

- after abrupt echo path changes (*Recovery:* increase adaption gain!)
- after double talk (*Recovery:* stop adaptation at once!)

Correct isolation is far more important then just detection.

Parameter or variance change

$$\begin{aligned} H_0: &\quad \theta^0 = \theta^1 \quad \text{and } \mathrm{Var}(w^0) = \mathrm{Var}(w^1) \\ H_1: &\quad \theta^0 \neq \theta^1 \quad \text{and } \mathrm{Var}(w^0) = \mathrm{Var}(w^1) \\ H_2: &\quad \theta^0 = \theta^1 \quad \text{and } \mathrm{Var}(w^0) \neq \mathrm{Var}(w^1) \end{aligned}$$

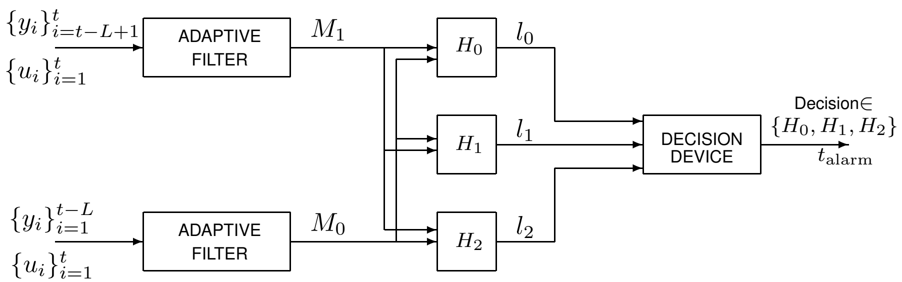
Probability q_1 for H_1 and q_2 for H_2 can be used.

Wishart prior on variance include parameters m and σ.

Sufficient Statistics

Data	$\underbrace{y(1), y(2), .., y(t-L)}$,	$\underbrace{y(t-L+1), .., y(t)}$
Model	M_0	M_1
Time interval	T_0	T_1
RLS quantities	$\theta^0,\ P_0$	$\theta^1,\ P_1$
Loss function	V_0	V_1
Number of data	$n_0 = t - L$	$n_1 = L$

Likelihood Approach to Detection and Isolation



A Posteriori Likelihoods

$$\begin{aligned} l_0 &\approx (n_0 + n_1 - 2 + m)\log\left(\frac{V_0(\theta^0) + V_1(\theta^0) + \sigma}{n_0 + n_1 - 4}\right) \\ &\quad + \log\det(P_0^{-1} + P_1^{-1}) + 2\log(q_0), \\ l_1 &\approx (n_0 + n_1 - 2 + m)\log\left(\frac{V_0(\theta^0) + V_1(\theta^1) + \sigma}{n_0 + n_1 - 4}\right) \\ &\quad - \log\det P_0 - \log\det P_1 + 2\log(q_1), \\ l_2 &\approx (n_0 - 2 + m)\log\left(\frac{V_0(\theta^0) + \sigma}{n_0 - 4}\right) \\ &\quad + (n_1 - 2 + m)\log\left(\frac{V_1(\theta^0) + \sigma}{n_1 - 4}\right) \\ &\quad - 2\log\det P_0 + 2\log(q_2), \end{aligned}$$

Exercises + Next time

Exercises: 31, 33

Next time: Chapter 7 Change detection based on filter banks