F2E5216/TS1002 Adaptive Filtering and Change Detection

Fredrik Gustafsson (LiTH) and Bo Wahlberg (KTH)

Linköpings universitet

Lecture 6

Change detection based on filter banks (Chapter 7)

- Segmentation
- Local tree search

Filter Banks for Parametric Changes

A changing regression model

Formulation 1:

$$y_t = \varphi_t^T \theta(i) + e_t, \quad \text{when } k_{i-1} < t \leq k_i.$$

Formulation 2:

$$\begin{aligned} \theta_{t+1} &= (1-\delta_t)\theta_t + \delta_t v_t \\ y_t &= \varphi_t^T \theta_t + e_t. \end{aligned}$$

The measurement covariance is $\lambda(i)R_t$, where $\lambda(i)$ may be unknown.

Notation

Parameter estimate in segment i:

$$\begin{aligned} \hat{\theta}(i) &= P(i) \sum_{t=k_{i-1}+1}^{k_i} \varphi_t R_t^{-1} y_t \\ P(i) &= \left(\sum_{t=k_{i-1}+1}^{k_i} \varphi_t R_t^{-1} \varphi_t^T \right)^{-1} \end{aligned}$$

Sufficient statistics in segment i:

$$\begin{aligned} V(i) &= \sum_{t=k_{i-1}+1}^{k_i} (y_t - \varphi_t^T \hat{\theta}(i))^T R_t^{-1} (y_t - \varphi_t^T \hat{\theta}(i)) \\ D(i) &= -\log\det P(i) \\ N(i) &= k_i - k_{i-1} + 1 \end{aligned}$$

Segmentation

Off-line formulation of the segmentation problem:

$$\widehat{k^n} = \arg \max_{n,k_1,\ldots,k_n} V(k^n).$$

The loss function depends on the sufficient statistics below:

Data	$\underbrace{y_1, y_2, .., y_{k_1}}$	$\underbrace{y_{k_1+1}, .., y_{k_2}}$	$\cdots$	$\underbrace{y_{k_{n-1}+1}, .., y_{k_n}}$
Segmentation	Segment 1	Segment 2	$\cdots$	Segment n
LS estimates	$\hat{\theta}(1), P(1)$	$\hat{\theta}(2), P(2)$	$\cdots$	$\hat{\theta}(n), P(n)$
Sufficient statistics	$V(1), D(1), N(1)$	$V(2), D(2), N(2)$	$\cdots$	$V(n), D(n), N(n)$

Optimality Criteria

Possible loss functions $V(k^n)$

- Statistical criterion: The maximum likelihood or maximum *a posteriori* estimate of k^n is studied.

- Information based criterion: The information of data is $\sum_i V(i)$ (the sum of squared residuals). Degenerated solution $k^n = 1, 2, 3, ..., N$ gives $\sum_i V(i) = 0$. A penalty term is needed.

Computations

- All loss functions $V(k^n)$ can be computed recursively using RLS.

- There are 2^N loss functions $V(k^n)$. Numerical or local search approximation.

Likelihoods and posteriors

Likelihood given everything:

$$-2 \log p(y^N | k^n, \theta^n, \lambda^n) = Np \log(2\pi)$$
$$+ \left(\sum_{t=1}^N \log \det R_t \right) \sum_{i=1}^n N(i) \log(\lambda(i)^p)$$
$$+ \sum_{i=1}^n \frac{\sum_{t=k_{i-1}+1}^{k_i} (y_t - \varphi_t^T \theta(i))^T R_t^{-1} (y_t - \varphi_t^T \theta(i))}{\lambda(i)}$$

Likelihood given only change times (marginalization)

$$p(y^N | k^n) = \int_{\theta^n, \lambda^n} p(y^N | k^n, \theta^n, \lambda^n) p(\theta^n | \lambda^n) p(\lambda^n) d\theta^n d\lambda^n$$

Possible to include prior information about change times using MAP estimate: $p(k^n | y^N) = p(y^N | k^n) \frac{p(k^n)}{p(y^N)}$.

Filter bank Segmentation

- Examine every possible segmentation, parameterized in the number of jumps n and jump times k^n
- For each segment compute the LS parameter estimates and their covariance matrices
- Compute the sum of the squared prediction errors $V(i)$ and $D(i) = -\log\det P(i)$
- Minimize the posterior probabilities!

Explicit forms of posterior probabilities: Known noise variance:

$$\widehat{k^n} = \arg\min_{k^n,n} \sum_{i=1}^{n} (D(i) + V(i)) + 2n \log \frac{1-q}{q}$$

Unknown and constant noise variance $\lambda(i) = \lambda$:

$$\widehat{k^n} = \arg\min_{k^n,n} \sum_{i=1}^{n} D(i) + (Np - nd - 2) \log \sum_{i=1}^{n} \frac{V(i)}{Np - nd - 4} + 2n \log \frac{1-q}{q}$$

Unknown and abruptly changing noise variance $\lambda(i)$:

$$\widehat{k^n} = \arg\min_{k^n,n} \sum_{i=1}^{n} \left(D(i) + (N(i)p - d - 2) \log \frac{V(i)}{N(i)p - d - 4} \right) + 2n \log \frac{1-q}{q}$$

Prior: Assume a fixed probability of jump q at each new time instant (Bernoulli distribution)

Lindley's paradox: The more non-informative prior (large $P(0)$), the more the zero-hypothesis is favored.

Still, priority is often given to non-informative priors in order to reduce the number of design parameters.

Comparing On-line and Off-line

The density function for the measurements can equivalently be computed from the residuals. With a little abuse of notation, $p(y^N) = p(\varepsilon^N)$. The sum of squared on-line residuals relates to the sum of squared off-line residuals as:

$$\sum_{t=1}^{N} (y_t - \varphi_t^T \hat{\theta}(t-1))^T \left(\varphi_t^T P(t-1)\varphi_t + R_t\right)^{-1} (y_t - \varphi_t^T \hat{\theta}(t-1))$$

$$= \sum_{t=1}^{N} (y_t - \varphi_t^T \hat{\theta}(N))^T R_t^{-1} (y_t - \varphi_t^T \hat{\theta}(N))$$

$$+ (\theta - \hat{\theta}(N))^T P(0)^{-1} (\theta - \hat{\theta}(N)).$$

Comparing On-line and Off-line

The on-line residual covariances relate to the *a posteriori* parameter covariance as:

$$-\sum_{t=1}^{N} D(t) = \sum_{t=1}^{N} \log\det\left(\varphi_t^T P(t-1)\varphi_t + R_t\right)$$
$$= \sum_{t=1}^{N} \log\det R_t + \log\det P(0) - \log\det P(N).$$

Together, this means that the likelihood given all available measurements can be computed from off-line statistics as:

$$p(y^N) \sim p(y^N|\hat{\theta}_N)p_\theta(\hat{\theta}_N)\left(\det P_N\right)^{1/2},$$

which holds for both a Gaussian prior $p_\theta(x)$ and a flat non-informative one $p_\theta(x) = 1$.

Off-line Global Optimization of $V(k^n)$

Approaches

- Gradient based methods where small changes in the current estimate k^n are evaluated.
- MCMC based methods, where random change points are computed.
- The EM algorithm, which alternates between estimating the parameters conditioned on the change points (trivial) and estimating the change points, given the parameters.

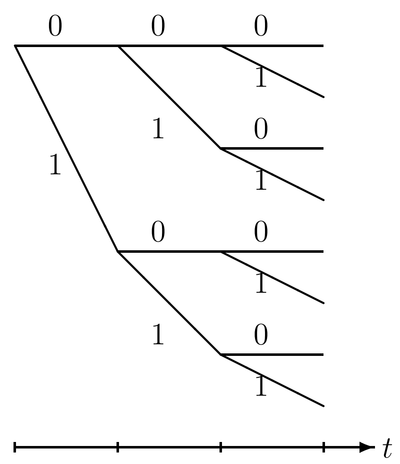
On-line Local Optimization of $V(k^n)$

Algorithm

1. Choose an optimality criterion.
2. Compute $V(k^n)$ using a bank of filters matched to k^n.
3. Pruning and splitting rules to keep M fixed.

a. Let only the most probable sequence split.

b. Cut off the least probable sequence.



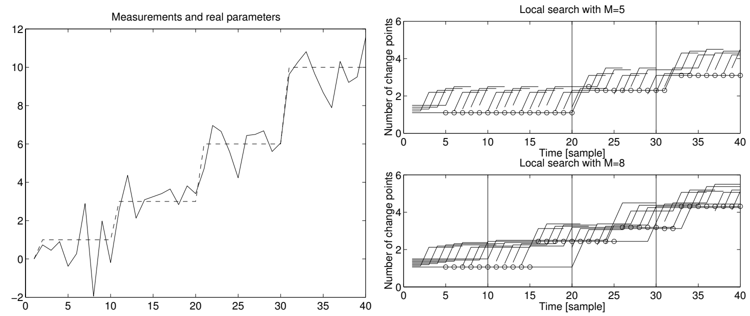
Additional rules

c. A minimum segment length

d. A minimum length of life

Example

```
[jumphat,thseg,lamseg,threc,Alfa,Xn]=
detectM(y,-1,[1 1],0.5,M,0);
hypplot(Xn,jumphat)
```

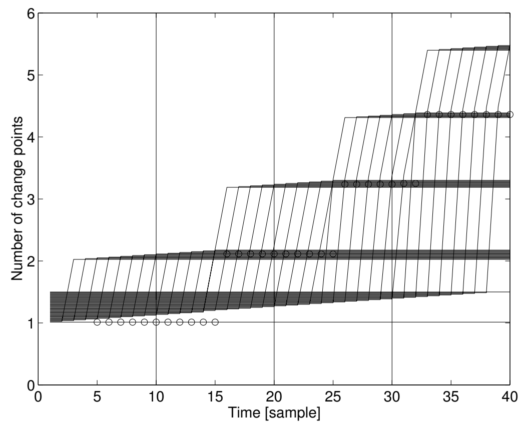



Optimal Segmentation: A First Inequality

Important Viterbi-like inequality:

If we condition on a jump at time k_0, then we only need to consider sequences which start with the best possible subsequence

Consequence: a filter bank with $M = N$ (and length of life $N-1$) finds the global optimum using $N^2/2$ evaluations of the loss function.



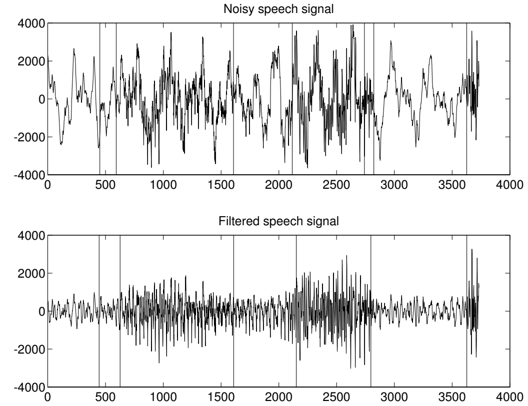
For MAP estimation, we have $t+1$ candidates for the global optimum of the jump sequence at time t.

$$\begin{aligned}\delta(0) &= (0,0,...,0)\\ \delta(k) &= (\hat{\delta}^{k-1}_{MAP},1,0,...,0) \quad k=1,2,...,t\end{aligned}$$

Step 1 in local search algorithm (let only the most probable branch split) is optimal.

Application: Speech Segmentation

A speech signal recorded in a car (upper plot) and a high-pass filtered version (lower plot). Vertical lines indicate the segmentation



Comparison of local search with $q=0.5$ and $M=10$ and a variant of `detect2` using the CUSUM test, with different design parameter for voiced and unvoiced sounds.

Signal	Method	AR	Estimated change times								
Noisy	Divergence	16	451	611	1450	1900	2125		2830		3626
Noisy	Brandt's GLR	16	451	611	1450	1900	2125		2830		3626
Noisy	Brandt's GLR	2		593	1450		2125		2830		3626
Noisy	Approx ML	2	451	593	1608		2116	2741	2822		3626
Filtered	Divergence	16	445	645	1550	1800	2151		2797		3626
Filtered	Brandt's GLR	16	445	645	1550	1800	2151		2797		3626
Filtered	Brandt's GLR	2	445	645	1550	1750	2151		2797	3400	3626
Filtered	Approx ML	2	445	626	1609		2151		2797		3627

Conclusions: Under-modeling works well. Essentially the same result for local search as the expert tuned model validation methods (five design parameters).

Application: Path Segmentation

Measurements of angular velocity of left and right (non-driven) wheels.

$$\begin{aligned}v(t) &= \frac{\omega_l(t)+\omega_r(t)}{2}r\\ R^{-1}(t) &= \frac{1-\frac{\omega_l(t)}{\omega_r(t)}(1+\varepsilon)}{L}\end{aligned}$$

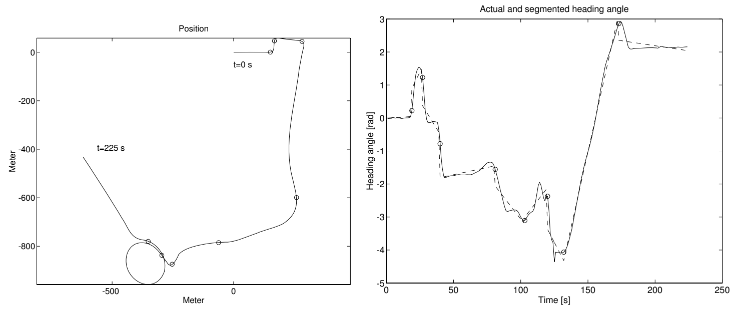
L wheel base
r nominal wheel radius
ε relative difference in wheel radius The heading angle $\psi(t)$ and global position $(X(t),Y(t))$ computed from

$$\begin{aligned}\psi(t+1) &= \psi(t)+v(t)T_sR^{-1}(t)\\ X(t+1) &= X(t)+v(t)T_s\cos(\psi(t))\\ Y(t+1) &= Y(t)+v(t)T_s\sin(\psi(t))\end{aligned}$$

Data model. Assumption that steering wheel angle is piecewise constant:

$$\begin{aligned}\theta_{t+1} &= (1-\delta_t)\theta_t + \delta_t v_t \\ \psi_t &= \theta_t^1 + \theta_t^2 t + e_t \\ \mathrm{E}e_t^2 &= \lambda_t\end{aligned}$$

$\lambda = 0.05$, M=10, lifelength=6 and minseg=0 give:



Exercises for Lectures 6

Exercise: 28, 29

Next Time

We have now covered the first three parts of the book. Next time, we will start to study Part IV: State estimation, i.e. methods based on Kalman filtering!