A regression model with a hidden logistic process for feature extraction from time series

F. Chamroukhi, A. Samé, G. Govaert and P. Aknin

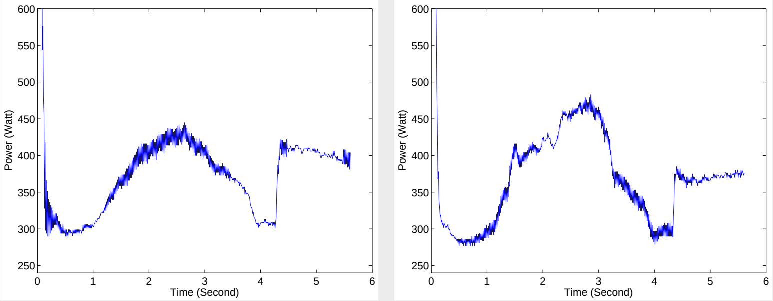
French National Institute for Transport and Safety Research (INRETS)

Compiègne University of Technology (UTC)
HEUDIASYC Laboratory - France

June 15 2009

Context: Feature extraction from the switch operation signals

- Signals of the consumed electrical power during switch operations



- Each switch operation consists of 5 successive electromechanical motions
- The signals present smooth or abrupt changes between different regimes
- The proposed solution: use an adapted regression model whose parameters will be used as the feature vector for each signal.

Piecewise polynomial regression model [McGee & Carleton 70]

- The data: $\{(x_1, t_1), \ldots, (x_n, t_n)\}$
  - x_i: real dependant variable: the observation of the signal
  - t_i: independant variable representing the time
- The piecewise polynomial regression model generating the signal $\mathbf{x}$ is:

$$\forall i = 1, \ldots, n, \quad x_i = \boldsymbol{\beta}_k^T \boldsymbol{r}_i + \sigma_k \epsilon_i \quad ; \quad \epsilon_i \sim \mathcal{N}(0, 1)$$

  - k satisfies $i \in I_k = (\gamma_k, \gamma_{k+1}]$: indexes of elements in segment k
  - $\boldsymbol{r}_i = (1, t_i, \ldots, t_i^p)^T$: time-dependant covariate vector in R^{p+1}
  - $\boldsymbol{\beta}_k$: regression coefficients vector $\in R^{(p+1)}$ for the k^{th} segment

The model parameters

$(\boldsymbol{\psi}, \boldsymbol{\gamma})$ with $\boldsymbol{\psi} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K, \sigma_1^2, \ldots, \sigma_K^2)$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{K+1})$.

Parameter estimation for the piecewise regression model

Maximize the likelihood of (ψ, γ) or equivalently minimize, with respect to (ψ, γ):

$$J(\psi, \gamma) = \sum_{k=1}^{K} \sum_{i \in I_k} \left[\log \sigma_k^2 + \frac{(x_i - \boldsymbol{\beta}_k^T \boldsymbol{r}_i)^2}{\sigma_k^2} \right].$$

- Global optimization using Fisher's algorithm [Fisher 58] based on dynamic programming [Bellman 61; Lechevallier 90] since the criterion J is additive on k
- Local optimization using an iterative variant of Fisher's Algorithm [Samé et al. 07]

Time series approximation and segmentation

- $\hat{x}_i = \sum_{k=1}^{K} \hat{z}_{ik} \hat{\boldsymbol{\beta}}_k^T \boldsymbol{r}_i \quad ; \quad \forall i = 1, \ldots, n$
- $\hat{z}_{ik} = 1$ if $i \in (\hat{\gamma}_k, \hat{\gamma}_{k+1}]$ (x_i belongs to the k^{th} segment) and $\hat{z}_{ik} = 0$ otherwise

- Using dynamic programming can be computationally expensive
- Provides a hard partition ⇒ adapted for regimes with abrupt changes

The proposed regression based on hidden process approach

The global regression model

$$\forall i = 1, \ldots, n, , \quad x_i = \boldsymbol{\beta}_{z_i}^T \boldsymbol{r}_i + \sigma_{z_i} \epsilon_i \quad ; \quad \epsilon_i \sim \mathcal{N}(0, 1),$$

- $z_i \in \{1, \ldots, K\}$ hidden variable: the class label of the regression model generating x_i

$\mathbf{z} = (z_1, \ldots, z_n)$ is a hidden discrete process

$z_i \sim \mathcal{M}\left(1, \pi_{i1}(\mathbf{w}), \ldots, \pi_{iK}(\mathbf{w})\right)$; where

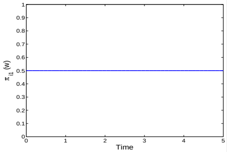
$$\pi_{ik}(\mathbf{w}) = p(z_i = k; \mathbf{w}) = \frac{\exp\left(\boldsymbol{w}_k^T \boldsymbol{v}_i\right)}{\sum_{\ell=1}^{K} \exp\left(\boldsymbol{w}_\ell^T \boldsymbol{v}_i\right)},$$

- $\boldsymbol{v}_i = (1, t_i, \ldots, t_i^q)^T$ time-dependant covariate vector $\in R^{(q+1)}$
- $\mathbf{w} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K)$ the parameter vector for the K k logistic functions $\in R^{K \times (q+1)}$

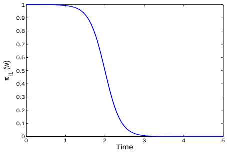
Flexibility of the logistic transformation: Example for $K = 2$:

1. $\pi_{ik}(\mathbf{w})$ in relation to the dimension q of $\boldsymbol{w}_k$:

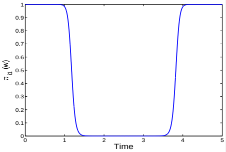
$q = 0$



$q = 1$



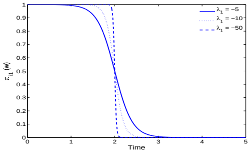
$q = 2$



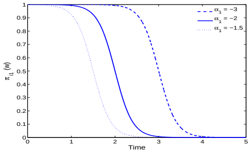
$\Rightarrow q = 1$ guarantees segmentation into contiguous segments

2. $\pi_{ik}(\mathbf{w})$ in relation to $\boldsymbol{w}_k$ for $q = 1$; we parametrize $\boldsymbol{w}_k$ by $\boldsymbol{w}_k = \lambda_k(\alpha_k, 1)^T$

$\alpha_1 = -2$



$\lambda_1 = -5$



$\Rightarrow$ The parameter λ_k controls the quality of transitions between classes
$\Rightarrow$ The parameter α_k controls the transition time point.

Parameter estimation by maximum likelihood

- Derived mixture density

$$p(x_i;\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_{ik}(\mathbf{w})\mathcal{N}(x_i;\boldsymbol{\beta}_k^T \boldsymbol{r}_i, \sigma_k^2)$$

- Model parameters

$$\boldsymbol{\theta} = \big(\mathbf{w}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K, \sigma_1^2, \ldots, \sigma_K^2\big)$$

- Log-likelihood of $\boldsymbol{\theta}$:

$$L(\boldsymbol{\theta};\mathbf{x}) = \sum_{i=1}^{n} \log \sum_{k=1}^{K} \pi_{ik}(\boldsymbol{w})\mathcal{N}(x_i;\boldsymbol{\beta}_k^T \boldsymbol{r}_i, \sigma_k^2).$$

- Maximization of $L(\boldsymbol{\theta};\mathbf{x})$ by a dedicated Expectation-Maximization (EM) algorithm [Dempster et al. 77].

Dedicated EM algorithm

Initialization: $\boldsymbol{\theta}^{(0)}$

Repeat until convergence:

1. **E step: Expectation** (at iteration m)

   Compute the conditional expectation of the complete log-likelihood $L(\boldsymbol{\theta};\mathbf{x},\mathbf{z})$

$$Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(m)}) = E\Big[L(\boldsymbol{\theta};\mathbf{x},\mathbf{z})\Big|\mathbf{x},\boldsymbol{\theta}^{(m)}\Big]$$

$$= \underbrace{\sum_{i=1}^{n}\sum_{k=1}^{K}\tau_{ik}^{(m)}\log \pi_{ik}(\mathbf{w})}_{Q_1(\mathbf{w})} + \underbrace{\sum_{i=1}^{n}\sum_{k=1}^{K}\tau_{ik}^{(m)}\log \mathcal{N}\Big(x_i;\boldsymbol{\beta}_k^T\boldsymbol{r}_i,\sigma_k^2\Big)}_{Q_2(\boldsymbol{\beta}_k,\sigma_k^2;k=1,\ldots,K)},$$

2. **M step: Maximization** (at iteration m)

   Compute $\boldsymbol{\theta}^{(m+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(m)})$

Details of the M step

1. Maximization of Q_2 with respect to $\boldsymbol{\beta}_k$s: Analytic solutions of K separate polynomial regression problems weighted by the $\tau_{ik}^{(m)}$s:
   - $\beta_k^{(m+1)} = (\mathbf{M}^T \boldsymbol{\Gamma}_k^{(m)} \mathbf{M})^{-1} \mathbf{M}^T \boldsymbol{\Gamma}_k^{(m)} \mathbf{x}$

   where $\mathbf{M}$ is the design matrix and $\boldsymbol{\Gamma}_k^{(m)} = \text{diag}\,(\tau_{1k}^{(m)}, \ldots, \tau_{nk}^{(m)})$.

   Maximization of Q_2 with respect to σ_k^2s
   - $\sigma_k^{2(m+1)} = \frac{1}{\sum_{i=1}^n \tau_{ik}^{(m)}} \sum_{i=1}^n \tau_{ik}^{(m)} (x_i - \boldsymbol{\beta}_k^{T(m+1)} \boldsymbol{r}_i)^2.$

2. Maximize Q_1 with respect to $\mathbf{w}$: Solve a multiclass convex logistic regression problem weighted by the $\tau_{ik}^{(m)}$s $\Rightarrow$ IRLS algorithm [Chen 99, Green 84, Krishnapuram 05]

$$\mathbf{w}^{(c+1)} = \mathbf{w}^{(c)} - \left[\frac{\partial^2 Q_1(\mathbf{w})}{\partial \mathbf{w} \partial \mathbf{w}^T}\right]^{-1}_{\mathbf{w}=\mathbf{w}^{(c)}} \frac{\partial Q_1(\mathbf{w})}{\partial \mathbf{w}}\Big|_{\mathbf{w}=\mathbf{w}^{(c)}} ,$$

$\Rightarrow$ Applying the IRLS algorithm provides the parameter $\mathbf{w}^{(m+1)}$.

Time series approximation and segmentation with the proposed model

Time series approximation

- As in standard regression, given the estimated parameters, x_i is approximated by its expectation $\forall i = 1, \ldots, n$:

$$\hat{x}_i \quad = \quad E(x_i; \hat{\boldsymbol{\theta}}) = \int_R x_i p(x_i; \hat{\boldsymbol{\theta}}) dx_i = \sum_{k=1}^{K} \pi_{ik}(\hat{\mathbf{w}}) \hat{\boldsymbol{\beta}}_k^T \boldsymbol{r}_i \quad .$$

A sum of polynomials weighted by the logistic probabilities $\pi_{ik}(\hat{\mathbf{w}})$'s.

$\Rightarrow$ Adapted for a smooth or abrupt transitions between the regression models.

Time series segmentation

- The estimated class label $\hat{z}_i$ of x_i is given by the rule:

$$\hat{z}_i = \arg \max_{1 \leq k \leq K} \pi_{ik}(\hat{\mathbf{w}}).$$

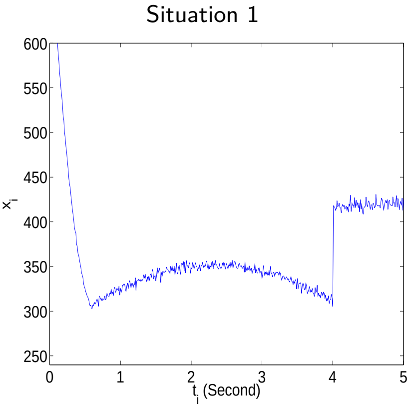
Experiments using simulated data

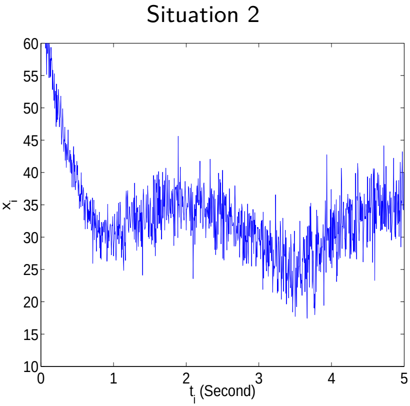
- Evaluation criteria:
  1. Misclassification error rate (segmentation error)
  2. Error between the true simulated curve without noise and the estimated curve (Denoising error):

$$\frac{1}{n}\sum_{i=1}^{n}(E(x_i;\boldsymbol{\theta}) - E(x_i;\hat{\boldsymbol{\theta}}))^2$$

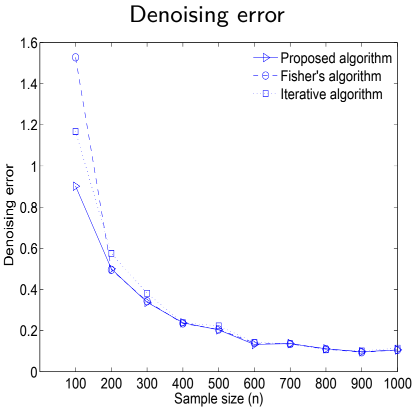
- Comparison with the two piecewise regression approahces
- 2 situations of signals with
  - $K = 3$, $p = 2$, $q = 1$
  - Varying the sample size $n = 100, 200, \ldots, 1000$
  - $\sigma_1^2 = 4$, $\sigma_2^2 = 10$, $\sigma_3^2 = 15$
- Assessment criteria are averaged over 20 samples for each value of n

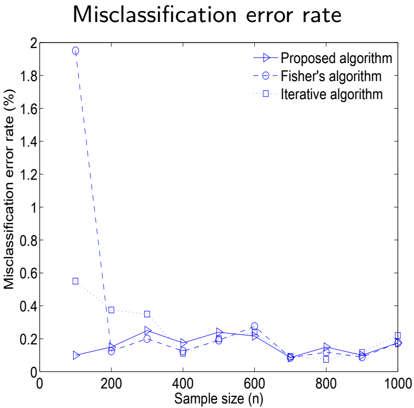
Example of simulated signals



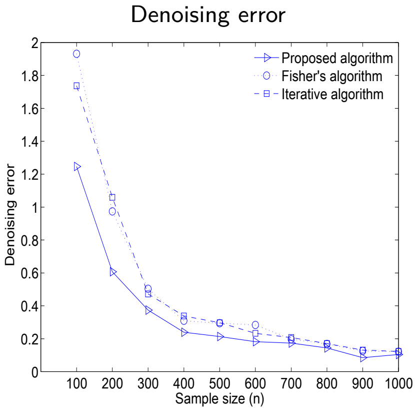


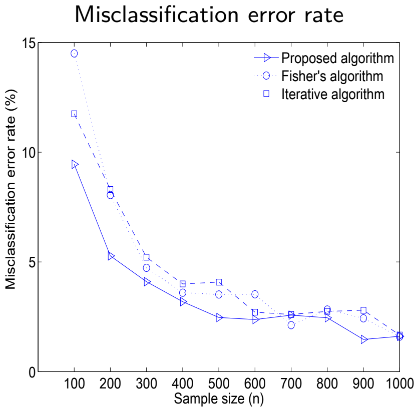
Results 1 (Situation 1)



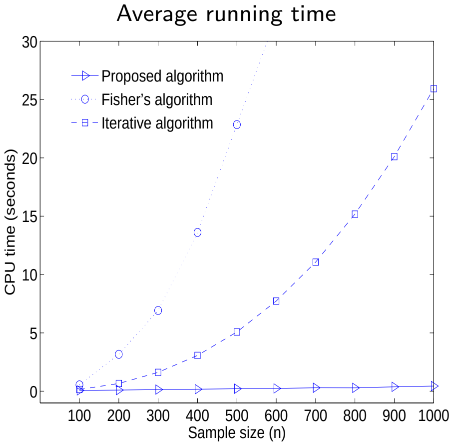


Results 2 (Situation 2)





Computing time

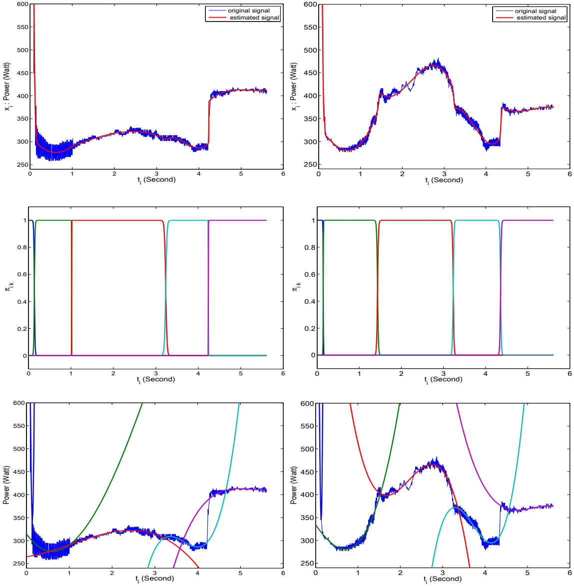


Application to real signals

Original signal and estimated signal

Probabilities of the regression models

Corresponding regression models



Conclusion

- In contrast with the basic polynomial regression, the proposed approach authorizes the regression parameters to vary over time

  $\Rightarrow$ Accurate modeling of nonlinear signals

- The proposed model integrates a logistic process which makes possible to change smoothly within various possible regression models

- In addition to feature extraction, this approach can be used to denoise and segment time series (or signals)

- Computationally efficient.

Thank you!