A regression model with a hidden logistic process for signal parametrization

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

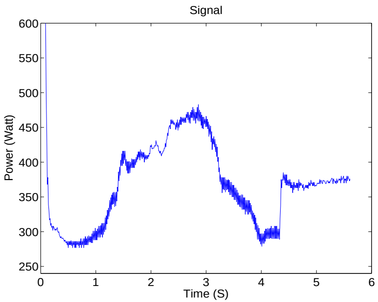
French National Institute for Transport and Safety Research (INRETS)
Compiègne University of Technology (UTC)
HEUDIASYC Laboratory - France

April 24 2009

1. Context
2. The piecewise regression approach
3. The proposed regression approach
4. Parameter estimation
5. Experiments

Context: Predictive maintenance of the French railway switches

- Available data: signals of consumed electrical power during switch operations;
- Each switch operation consists of 5 successive electromechanical motions.



1. Parameterize switch operations signals;
2. Exploit these parameters for identifying incipient faults.

Piecewise regression model [McGee et Carleton 70]

- The data: $\{(x_1, t_1), \ldots, (x_n, t_n)\}$ where x_i is the observation of the signal at time t_i;
- The piecewise polynomial regression model generating the signal $\mathbf{x}$ is:

$$\forall i = 1, \ldots, n, \quad x_i = \begin{cases} \boldsymbol{\beta}_1^T \boldsymbol{r}_i + \epsilon_{i1} & \text{if } i \in I_1 \\ \boldsymbol{\beta}_2^T \boldsymbol{r}_i + \epsilon_{i2} & \text{if } i \in I_2 \\ \vdots & \\ \boldsymbol{\beta}_K^T \boldsymbol{r}_i + \epsilon_{iK} & \text{if } i \in I_K \end{cases},$$

  - $I_k =]\gamma_k, \gamma_{k+1}]$: indexes of elements of segment k with $(\gamma_1 = 0$ and $\gamma_{K+1} = n)$.
  - $\boldsymbol{r}_i = (1, t_i, \ldots, t_i^p)^T$: time-dependant covariate vector in R^{p+1};
  - $\boldsymbol{\beta}_k \in R^{(p+1)}$: the regression parameters associated to the k^{th} segment;
  - $\epsilon_{ik} \sim \mathcal{N}(0, \sigma_k^2)$: independent additive Gaussian noise on the segment k.

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 of 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.

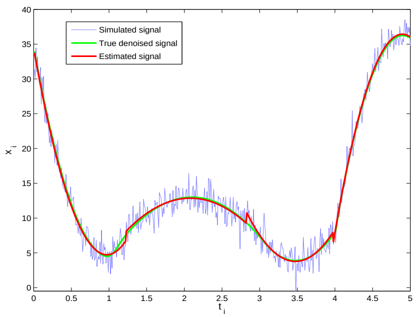
- x_i is approximated by a polynomial:

$$\hat{x}_i = \sum_{k=1}^{K} \hat{z}_{ik} \hat{\boldsymbol{\beta}}_k^T \boldsymbol{r}_i \quad ; \quad \forall i = 1, \ldots, n$$

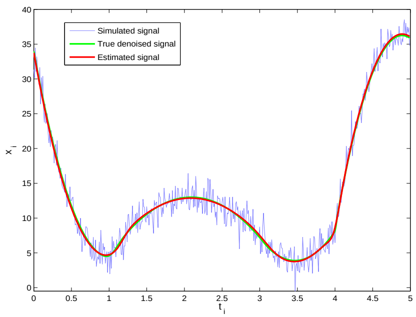
  - $\hat{z}_{ik} \in \{0, 1\}$
  - $(\hat{\gamma}_1, \ldots, \hat{\gamma}_{K+1}, \hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_K)$: the estimated model parameters.

Disadvantages

- Using dynamic programming is computationally expensive;
- Provides a hard partition ⇒ the estimated signal can present discontinuities.



Piecewise regression approach



Proposed approach

The proposed regression based on hidden process approach

The global regression model

- The data: $\{(t_1, x_1), \ldots, (t_n, x_n)\}$;

$$x_i = \boldsymbol{\beta}_{z_i}^T \boldsymbol{r}_i + \epsilon_i \quad ; \quad \forall i = 1, \ldots, n,$$

- $z_i \in \{1, \ldots, K\}$ hidden vaiable: the label of the regression model generating x_i;
- $\boldsymbol{\beta}_{z_i} \in R^{p+1}$: regression parameters of the regression model z_i;
- $\epsilon_i \sim \mathcal{N}(0, \sigma_{z_i}^2)$: independent Gaussian noise on the component z_i.

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

The proposed regression based on hidden process approach

The global regression model

- The data: $\{(t_1, x_1), \ldots, (t_n, x_n)\}$;

$$x_i = \boldsymbol{\beta}_{z_i}^T \boldsymbol{r}_i + \epsilon_i \quad ; \quad \forall i = 1, \ldots, n,$$

- $z_i \in \{1, \ldots, K\}$ hidden vaiable: the label of the regression model generating x_i;
- $\boldsymbol{\beta}_{z_i} \in R^{p+1}$: regression parameters of the regression model z_i;
- $\epsilon_i \sim \mathcal{N}(0, \sigma_{z_i}^2)$: independent Gaussian noise on the component z_i.

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

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

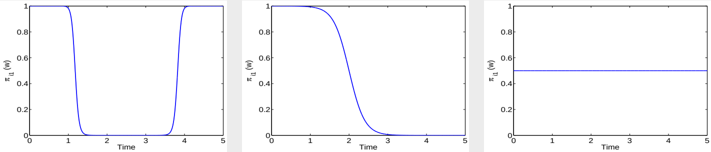
$$\pi_{ik}(\mathbf{w}) = p(z_i = k; \mathbf{w}) = \frac{\exp\left(\boldsymbol{w}_k^T \boldsymbol{v}_i\right)}{\sum_{\ell=1}^{K} \exp\left(\boldsymbol{w}_\ell^T \boldsymbol{v}_i\right)},$$

- $\boldsymbol{v}_i = (1, t_i, \ldots, t_i^q)^T$ time-dependant covariate vector in $R^{(q+1)}$;
- $\mathbf{w} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K)$ the parameter vector of the K regression models.

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

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

$q = 2$ | $q = 1$ | $q = 0$

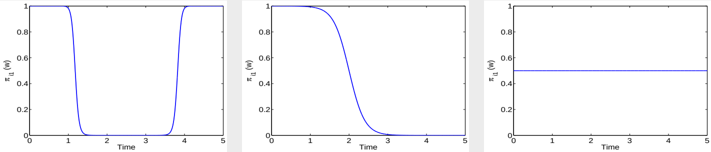


$\Rightarrow q = 1$ garantees a segmentation into homogenous parts.

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

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

$q = 2$ | $q = 1$ | $q = 0$

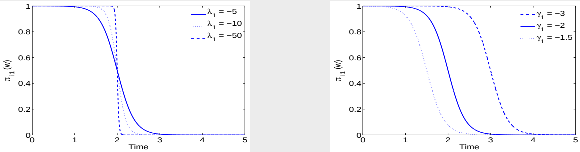


$\Rightarrow$ $q = 1$ garantees a segmentation into homogenous parts.

2. Evolution of $\pi_{ik}(\mathbf{w})$ during time in relation to $\boldsymbol{w}_k$ for $q = 1$:

   - Parametrize $\boldsymbol{w}_k = (\boldsymbol{w}_{k0}, \boldsymbol{w}_{k1})^T$ with $\boldsymbol{w}_k = \lambda_k(\gamma_k, 1)^T$.

$\gamma_1 = -2$ | $\lambda_1 = -5$



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

Parameter estimation

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

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

Parameter estimation

- Maximum Likelihood method;
- Log-likelihood of $\boldsymbol{\theta}$:

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

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

Dedicated EM algorithm

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

Repeat until convergence:

Dedicated EM algorithm

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

Repeat until convergence:

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 $L(\boldsymbol{\theta}; \mathbf{x}, \mathbf{z})$

$$
\begin{aligned}
Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m)}) &= E\left[L(\boldsymbol{\theta}; \mathbf{x}, \mathbf{z})|\mathbf{x}, \boldsymbol{\theta}^{(m)}\right] \\
&= \underbrace{\sum_{i=1}^{n}\sum_{k=1}^{K} \tau_{ik}^{(m)} \log \pi_{ik}(\mathbf{w})}_{Q_1(\mathbf{w})} + \underbrace{\sum_{i=1}^{n}\sum_{k=1}^{K} \tau_{ik}^{(m)} \log \mathcal{N}\left(x_i; \boldsymbol{\beta}_k^T \boldsymbol{r}_i, \sigma_k^2\right)}_{Q_2(\boldsymbol{\beta}_k, \sigma_k^2; k=1,\ldots,K)},
\end{aligned}
$$

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 $L(\boldsymbol{\theta}; \mathbf{x}, \mathbf{z})$

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

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

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

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

M step (suite)

1. Maximize Q_2 with respect to $\boldsymbol{\beta}_k$s and σ_k^2s: Analytic solutions of K separate polynomial regression problems weighted by the $\tau_{ik}^{(m)}$s:
   - $\boldsymbol{\beta}_k^{(m+1)} = (\mathbf{M}^T\boldsymbol{\Gamma}_k^{(m)}\mathbf{M})^{-1}\mathbf{M}^T\boldsymbol{\Gamma}_k^{(m)}\mathbf{x}$

     where $\mathbf{M}$ is the design matrix and $\boldsymbol{\Gamma}_k^{(m)} = \text{diag}\,(\tau_{1k}^{(m)}, \ldots, \tau_{nk}^{(m)})$.
   - $\sigma_k^{2(m+1)} = \frac{1}{\sum_{i=1}^n \tau_{ik}^{(m)}} \sum_{i=1}^n \tau_{ik}^{(m)}(x_i - \boldsymbol{\beta}_k^{T(m+1)}\boldsymbol{r}_i)^2.$

M step (suite)

1. Maximize Q_2 with respect to $\boldsymbol{\beta}_k$s and σ_k^2s: Analytic solutions of K separate polynomial regression problems weighted by the $\tau_{ik}^{(m)}$s:
   - $\boldsymbol{\beta}_k^{(m+1)} = (\mathbf{M}^T \boldsymbol{\Gamma}_k^{(m)} \mathbf{M})^{-1} \mathbf{M}^T \boldsymbol{\Gamma}_k^{(m)} \mathbf{x}$

     where $\mathbf{M}$ is the design matrix and $\boldsymbol{\Gamma}_k^{(m)} = \mathrm{diag}\,(\tau_{1k}^{(m)}, \ldots, \tau_{nk}^{(m)})$.
   - $\sigma_k^{2(m+1)} = \frac{1}{\sum_{i=1}^n \tau_{ik}^{(m)}} \sum_{i=1}^n \tau_{ik}^{(m)} (x_i - \boldsymbol{\beta}_k^{T(m+1)} \boldsymbol{r}_i)^2.$
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 $\mathbf{w}^{(m+1)}$.

Estimated signal

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

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

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

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

Signal segmentation

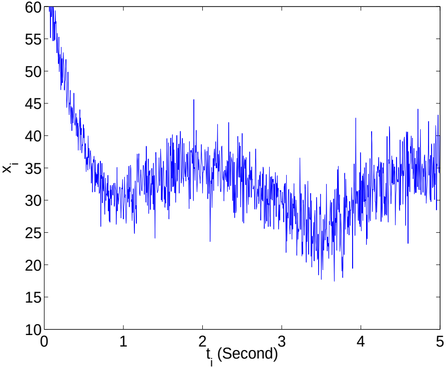
- The estimated label $\hat{z}_i$ of x_i is given by the rule:

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

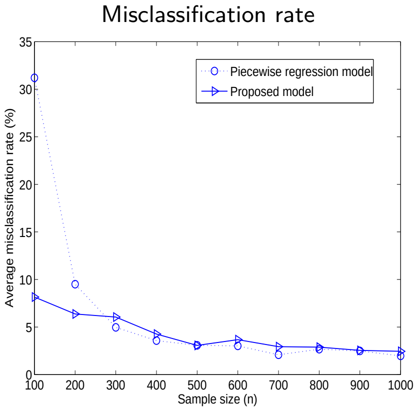
Experiments using simulated data

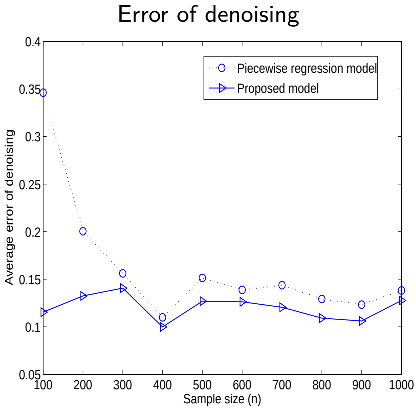
- Expermiment 1 ($K = 3$,$p = 2$,$q = 1$);
  - Varying the sample size $n = 100, 200, \ldots, 1000$;
  - $\sigma_1^2 = 4$, $\sigma_2^2 = 10$, $\sigma_3^2 = 15$;
- Expermiment 2 ($K = 3$,$p = 2$,$q = 1$);
  - Varying the noise level $\sigma = 0.5, 1, 2, \ldots, 7$
  - $n = 500$;
- Assessment criteria (averaged over 20 samples for each value of n and σ):
  - Misclassification rate;
  - Error between the true curve and the estimated curve (Error of denoising): $\frac{1}{n}\sum_{i=1}^{n}(E(x_i;\boldsymbol{\theta}) - E(x_i;\hat{\boldsymbol{\theta}}))^2$.

Example of simulated signal

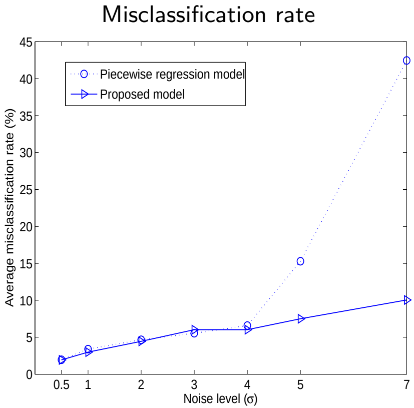


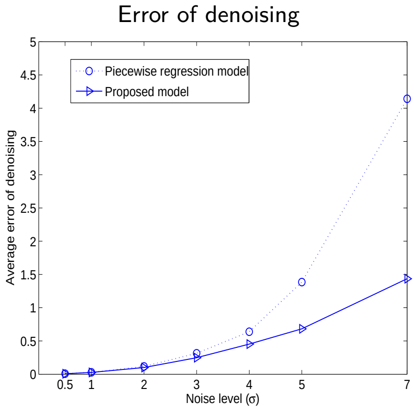
Results 1 (varying sample size n)





Results 2 (varying noise level σ)





Results in terms of computing time

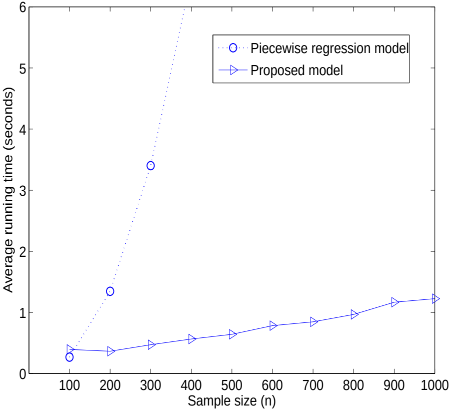


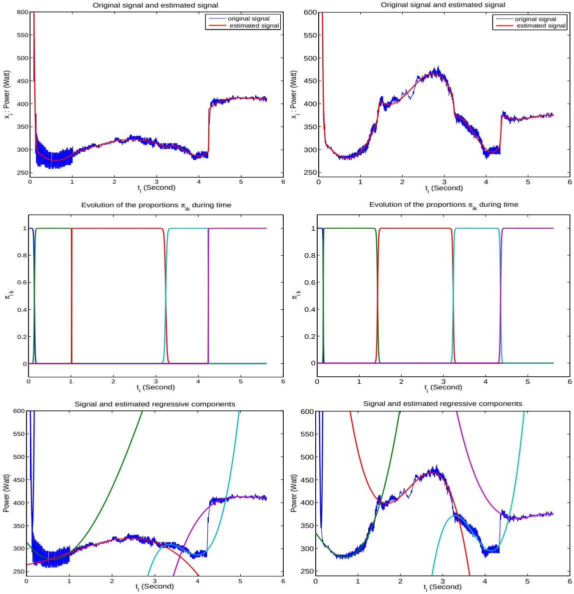
Figure: Average running time

Experiments on real signals

Original signal and estimated signal

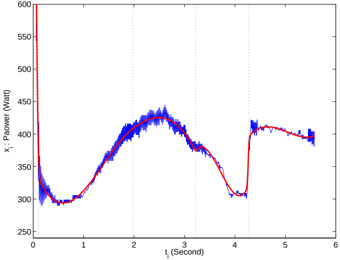
Probabilities of the regression models

Corresponding regression models

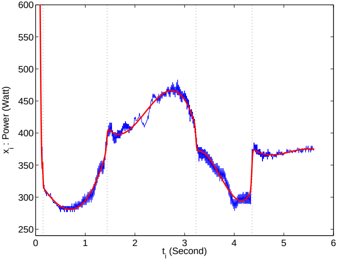


MSE between the original signal and the estimated signal

Proposed approach

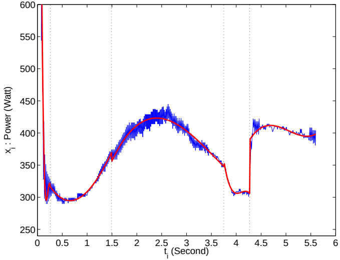


(a)$MSE = 353.319$

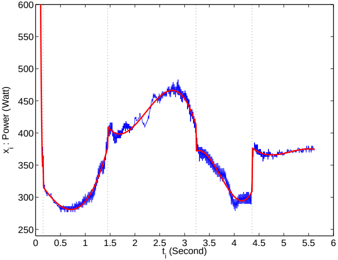


(b) $MSE = 307.753$

Piecewise regression approach



(c) $MSE = 497.561$



(d) $MSE = 310.251$

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;

- Makes possible to change smoothly within various possible regression models;

- In addition to feature extraction, this approach can be used for signals segmentation and denoising;

- Computationally efficient.

Thank you!