Learning probabilistic latent process models from temporal data

Faicel Chamroukhi

Information Sciences and Systems Laboratory (LSIS), UMR CNRS 7296
University of the South Toulon-Var (USTV), France

email: chamroukhi@univ-tln.fr
web: chamroukhi.univ-tln.fr

ERMITES 2012, La Seyne/Mer

01 september 2012
Outline

1. Context and objectives
2. Probabilistic modeling with Hidden process for curves
3. Curves classification
4. Applications
5. Conclusions
**Contexts**

1. Supervised learning: data \((x,y)\)
2. Unsupervised learning: data \(x, y\)?
3. Semi-supervised, partially supervised, ...

**Generative/Discriminative**

1. Discriminative approach: Directly learn \(p(y|x)\)
2. Generative approach: Learn \(p(x, y) \Rightarrow p(y|x) \propto p(x|y)p(y)\)
   - "understand" the process generating the data
   - Easily adaptable to the unsupervised context

⇒ Latent data models
**Context**

- Temporal data (signals, sequences, functions, time series, ..)

![A curve](image1.png)

![A set of curves](image2.png)

![Sequence of curves](image3.png)

- impedance signal of a fuel cell
- Acceleration measures
Context

- Temporal data (signals, sequences, functions, time series, ..)

1. Several regimes over time (temporal aspect) ⇒ Abrupt and/or regime changes
2. Many curves to analyse
3. Multidimensional temporal data

Impedance signal of a fuel cell
Acceleration measures
Objectives

- Take into account the **class dispersion** and the **regime changes**
- Explicitly integrate this complexity (class dispersion, regimes,..)

⇒ Learning probabilistic generative models
## Objectives

- Take into account the **class dispersion** and the **regime changes**
- Explicitly integrate this complexity (class dispersion, regimes,..)
  
  ⇒ Learning probabilistic generative models

## Curves modeling (functions)

- Approaches based on hidden process regression
- Probabilistic formalization of regime changes

## Curves Classification (supervised and unsupervised)

- Mixture model-based approach for curve classification
  
  - model-based functional cluster analysis (in the unsupervised case)
  - model-based functional discriminant analysis (in the supervised case)
Regression-based models

Data: 1 curves $\mathbf{y} = (y_1, \ldots, y_m)$ regularly observed at instants $(t_1, \ldots, t_m)$

\[ y_j = f(t_j) + \sigma \epsilon_j, \quad \epsilon_j \sim \mathcal{N}(0,1) \]
Regression-based models

Data: 1 curves \( y = (y_1, \ldots, y_m) \) regularly observed at instants \( (t_1, \ldots, t_m) \)

\[
y_j = f(t_j) + \sigma \epsilon_j, \quad \epsilon_j \sim \mathcal{N}(0, 1)
\]

- simple polynomial regression: \( f \) est une fonction polynôme
Regression-based models

Data: 1 curves $\mathbf{y} = (y_1, \ldots, y_m)$ regularly observed at instants $(t_1, \ldots, t_m)$

$$y_j = f(t_j) + \sigma \epsilon_j, \quad \epsilon_j \sim \mathcal{N}(0, 1)$$

- simple polynomial regression: $f$ est une fonction polynôme
- polynomial spline regression (Deboor, 1978)
Regression-based models

Data: 1 curves \( y = (y_1, \ldots, y_m) \) regularly observed at instants \( (t_1, \ldots, t_m) \)

\[
y_j = f(t_j) + \sigma \epsilon_j, \quad \epsilon_j \sim \mathcal{N}(0,1)
\]

- simple polynomial regression: \( f \) est une fonction polynôme
- polynomial spline regression (Deboor, 1978)
- \( f \) is a piecewise polynomial function (McGee et Carleton, 1970)
  \( \Rightarrow \) exact Optimization with dynamic prog. (Bellman, 1961)
Regression-based models

Data: 1 curves $\mathbf{y} = (y_1, \ldots, y_m)$ regularly observed at instants $(t_1, \ldots, t_m)$

$$y_j = f(t_j) + \sigma \epsilon_j, \quad \epsilon_j \sim \mathcal{N}(0, 1)$$

- simple polynomial regression: $f$ est une fonction polynôme
- polynomial spline regression (Deboor, 1978)
- $f$ is a piecewise polynomial function (McGee et Carleton, 1970)
  \[ \Rightarrow \text{exact Optimization with dynamic prog. (Bellman, 1961)} \]
  - dynamic programming can be time consuming
  - Adapted to abrupt regime changes but a problem of continuity
Regression-based models

Data: 1 curves \( y = (y_1, \ldots, y_m) \) regularly observed at instants \( (t_1, \ldots, t_m) \)

\[ y_j = f(t_j) + \sigma \epsilon_j, \quad \epsilon_j \sim \mathcal{N}(0,1) \]

- simple polynomial regression: \( f \) est une fonction polynôme
- polynomial spline regression (Deboor, 1978)
- \( f \) is a piecewise polynomial function (McGee et Carleton, 1970)
  \( \Rightarrow \) exact Optimization with dynamic prog. (Bellman, 1961)
    - dynamic programming can be time consuming
    - Adapted to abrupt regime changes but a problem of continuity
- Polynomial regression governed by a Markov chain (Fridman, 1993)
  The model for a curve: \( y_j = \beta_z^T t_j + \sigma z_j \epsilon_j \quad (j = 1, \ldots, m) \)
    - Continuity not guaranteed
    - Not adapted to approximate a set of curves
The Regression model with a Hidden Logistic Process (RHLP)

1. Context and objectives

2. Probabilistic modeling with Hidden process for curves
   - The proposed Regression model with a Hidden Logistic Process (RHLP)
   - Parameter estimation of the RHLP model
   - Experiments

3. Curves classification

4. Applications

5. Conclusions

Temporal data \( y = (y_1, \ldots, y_m) \) observed at instants \( t = (t_1, \ldots, t_m) \)

**Model definition**

\[
y_j = \mathbf{\beta}^T z_j + \sigma z_j \epsilon_j \quad ; \quad \epsilon_j \sim \mathcal{N}(0,1), \quad (j = 1, \ldots, m)
\]

- \( z_j \) hidden label of the regime of \( y_j \)
- \( z = (z_1, \ldots, z_n) \) is hidden logistic process

Temporal data \( y = (y_1, \ldots, y_m) \) observed at instants \( t = (t_1, \ldots, t_m) \)

Model definition

\[
y_j = \beta_{z_j}^T t_j + \sigma_{z_j} \epsilon_j \quad ; \quad \epsilon_j \sim \mathcal{N}(0,1), \quad (j = 1, \ldots, m)
\]

- \( z_j \) hidden label of the regime of \( y_j \)
- \( z = (z_1, \ldots, z_n) \) is hidden logistic process

\[
z_j | t_j \sim \mathcal{M}(1, \pi_1(t_j; w), \ldots, \pi_R(t_j; w)) \quad ; \quad \text{où}
\]

\[
\pi_r(t_j; w) = p(z_j = r|t_j; w) = \frac{\exp\left(\lambda_r(t_j + \gamma_r)\right)}{\sum_{\ell=1}^{R} \exp\left(\lambda_{\ell}(t_j + \gamma_{\ell})\right)}
\]

- \( w_r = (\lambda_r, \gamma_r)^T \) parameter of the \( r^{th} \) logistic function
- \( w = (w_1, \ldots, w_R) \) parameter of the \( R \) logistic functions
Flexibility of the logistic transformation

Temporal variation of the logistic function w.r.t $\mathbf{w}$

- $\pi_r(t_i; \mathbf{w}) = \frac{\exp(\lambda_r(t_i + \gamma_r))}{\sum_{\ell=1}^{R} \exp(\lambda_\ell(t_i + \gamma_\ell))}$

- Example with $K = 2$ regimes

$\Rightarrow$ The parameter $\lambda_r$ controls the quality of transitions between regimes

$\Rightarrow$ The parameter $\gamma_r$ is related to the transition time
Illustration of the principle of the method
Illustration of the principle of the method
Parameter estimation with ML via EM

- parameter vector of the model: \( \theta = (w, \beta_1, \ldots, \beta_K, \sigma_1^2, \ldots, \sigma_K^2) \)
- Log-likelihood: \( \mathcal{L}(\theta) = \sum_{j=1}^{m} \log \sum_{r=1}^{R} \pi_r(t_j; \theta) \mathcal{N}(y_j; \beta_r^T t_j, \sigma_r^2) \)
Parameter estimation with ML via EM

- Parameter vector of the model: \( \theta = (w, \beta_1, \ldots, \beta_K, \sigma_1^2, \ldots, \sigma_K^2) \)

- Log-likelihood: \( \mathcal{L}(\theta) = \sum_{j=1}^{m} \log \sum_{r=1}^{R} \pi_r(t_j; w) \mathcal{N}(y_j; \beta_r^T t_j, \sigma_r^2) \)

- Completed Log-likelihood:
  \[
  \mathcal{L}_c(\theta) = \sum_{j=1}^{m} \sum_{r=1}^{R} z_{jr} \log [\pi_r(t_j; w) \mathcal{N}(y_j; \beta_r^T t_j, \sigma_r^2)]
  \]
Parameter estimation with ML via EM

- Parameter vector of the model: \( \theta = (w, \beta_1, \ldots, \beta_K, \sigma^2_1, \ldots, \sigma^2_K) \)
- Log-likelihood: \( \mathcal{L}(\theta) = \sum_{j=1}^{m} \log \sum_{r=1}^{R} \pi_r(t_j; w) \mathcal{N}(y_j; \beta^T_r t_j, \sigma^2_r) \)
- Completed Log-likelihood: \( \mathcal{L}_c(\theta) = \sum_{j=1}^{m} \sum_{r=1}^{R} z_{jr} \log \left[ \pi_r(t_j; w) \mathcal{N}(y_j; \beta^T_r t_j, \sigma^2_r) \right] \)

**EM Algorithm for the proposed model**

Initial Parameter \( \theta^{(0)} \):

1. E-Step: Conditional expectation of the complete log-likelihood \( Q_{\theta^{(q)}}(\theta^{(q)}) \):

\[
\mathbb{E} \left[ \mathcal{L}_c(\theta) | y, t; \theta^{(q)} \right] = \sum_{r=1}^{R} \sum_{j=1}^{m} \tau_{jr}^{(q)} \log \pi_r(t_j; w) + \sum_{r=1}^{R} \sum_{j=1}^{m} \tau_{jr}^{(q)} \log \mathcal{N}(y_j; \beta^T_r t_j, \sigma^2_r)
\]

\[
Q_w \quad Q_{\theta r}
\]

\( \Rightarrow \) calculating the posterior probabilities \( \tau_{jr}^{(q)} = p(z_j = r | y_j, t_j; \theta^{(q)}) \)
Parameter estimation with ML via EM

- Parameter vector of the model: \( \theta = (w, \beta_1, \ldots, \beta_K, \sigma_1^2, \ldots, \sigma_K^2) \)
- Log-likelihood: \( \mathcal{L}(\theta) = \sum_{j=1}^{m} \log \sum_{r=1}^{R} \pi_r(t_j; w) \mathcal{N}(y_j; \beta_r^T t_j, \sigma_r^2) \)
- Completed Log-likelihood: \( \mathcal{L}_c(\theta) = \sum_{j=1}^{m} \sum_{r=1}^{R} z_{jr} \log [\pi_r(t_j; w) \mathcal{N}(y_j; \beta_r^T t_j, \sigma_r^2)] \)

EM Algorithm for the proposed model

Initial Parameter \( \theta^{(0)} \):

1. E-Step: Conditional expectation of the complete log-likelihood: \( Q(\theta, \theta^{(q)}) = \mathbb{E} \left[ \mathcal{L}_c(\theta) | y, t; \theta^{(q)} \right] \)

\[
\mathbb{E} \left[ \mathcal{L}_c(\theta) | y, t; \theta^{(q)} \right] = \sum_{r=1}^{R} \sum_{j=1}^{m} \tau_{jr}^{(q)} \log \pi_r(t_j; w) + \sum_{r=1}^{R} \sum_{j=1}^{m} \tau_{jr}^{(q)} \log \mathcal{N}(y_j; \beta_r^T t_j, \sigma_r^2)
\]

\( Q_w \) \hspace{2cm} \( Q_{\theta_r} \)

⇒ calculating the posterior probabilities \( \tau_{jr}^{(q)} = p(z_j = r | y_j, t_j; \theta^{(q)}) \)

2. M-Step: \( \theta^{(q+1)} = \arg\max_{\theta} Q(\theta, \theta^{(q)}) = Q_w(w, \theta^{(q)}) + \sum_{r=1}^{R} Q_{\theta_r}(\theta, \theta^{(q)}) \)

⇒ Separate Maximizations of \( Q_{\theta_r} \) and \( Q_w \)
Maximisation of $Q_{\theta_r}$: exact solutions of weighted regressions (weights $\tau_{jr}$)

$$
\beta_r^{(q+1)} = \left[ \sum_{j=1}^{m} \tau_{jr}^{(q)} t_j^T t_j \right]^{-1} \sum_{j=1}^{m} \tau_{jr}^{(q)} y_j t_j
$$

$$
\sigma_r^2(q+1) = \frac{1}{\sum_{j=1}^{m} \tau_{jr}^{(q)}} \sum_{j=1}^{m} \tau_{jr}^{(q)} (x_{ij} - \beta_{gkr}^{T(q+1)} t_j)^2.
$$

- Maximisation of $Q_w$: a convex problem of multi-class logistic regression weighted by $\tau_{jr}^{(q)} \Rightarrow$ iterative method: IRLS

$$
w^{(q,l+1)} = w^{(l)} - \left[ \frac{\partial^2 Q_w(w, \theta^{(q)})}{\partial w \partial w^T} \right]^{-1} \frac{\partial Q_w(w, \theta^{(q)})}{\partial w} \bigg|_{w=w^{(l)}}
$$
Curve approximation and segmentation

### Curve approximation

\[
\mathbb{E}[y_j|t_j; \hat{\theta}] = \sum_{r=1}^{R} \pi_r(t_j; \hat{w}) \hat{\beta}_r^T t_j
\]

Weighted sum of polynomials weighted by logistic functions

⇒ Adapted to abrupt and smooth regime change

⇒ Ensure the continuity and the regularity of the estimated curve

### Curve Segmentation

\[
\hat{z}_j = \arg \max_r \pi_r(t_j; \hat{w}), \quad (j = 1, \ldots, m)
\]

Choice of \((R, p) \Rightarrow BIC(R, p) = \mathcal{L}(\hat{\theta}) - \frac{v_\theta \log(m)}{2}\)
Evaluation in terms of modeling and segmentation

Approximation error as a function of the speed of transitions

Computing time
Evolution of the approximation and segmentation errors

**influence of \( m \)**

- Cubic spline
- Cubic B-spline
- PWPR
- HMMR
- RHLP

**influence of \( \sigma \)**

- Cubic spline
- Cubic B-spline
- PWPR
- HMMR
- RHLP

Faicel Chamroukhi (USTV/LSIS - France)
Curves classification

1. Context and objectives

2. Probabilistic modeling with Hidden process for curves

3. Curves classification
   - Curves Classification
   - Curves Clustering

4. Applications

5. Conclusions
Curves classification

Curves clustering

1. Model-based clustering:
   Regression mixtures, splines, B-splines (Gaffney, 2004; James and Sugar, 2003; Liu and Yang, 2009),
   Mixture of HMMs (Smyth, 1996)

2. Distance-based approach (K-means like) (Hébrail et al., 2010)
Curves classification

Curves clustering

1. Model-based clustering:
   Regression mixtures, splines, B-splines (Gaffney, 2004; James and Sugar, 2003; Liu and Yang, 2009),
   Mixture of HMMs (Smyth, 1996)

2. Distance-based approach (K-means like) (Hébrail et al., 2010)

Curves classification

1. Functional Linear Discriminant Analysis (James and Hastie, 2001)
2. Functional Mixture Discriminant Analysis (Gui and Li, 2003) (B-splines)
Curves clustering

Data: $n$ independent curves $(y_1, \ldots, y_n)$ (observed)
hidden classes $(h_1, \ldots, h_n)$, hidden regimes $(z_{1k}, \ldots, z_{mk})$ of class $k$
Curves clustering

Data: \( n \) independent curves \((y_1, \ldots, y_n)\) (observed)
hidden classes \((h_1, \ldots, h_n)\), hidden regimes \((z_{1k}, \ldots, z_{mk})\) of class \( k \)

Unsupervised learning, clustering and segmentation
(Advances in Data Analysis and Classification (ADAC) 5(4) : 301-321, 2011.)

- Mixture of RHLP models (MixRHLP)

\[
p(y_i|\mathbf{t}; \Psi) = \sum_{k=1}^{K} \alpha_k \prod_{j=1}^{m} \sum_{r=1}^{R} \pi_{kr}(t_j; \mathbf{w}_k) \mathcal{N}(y_{ij}; \mathbf{\beta}_{kr}^T \mathbf{t}_j, \sigma_{kr}^2)
\]

- Cluster: RHLP component density
- Cluster prob.
- Regime prob.
- Noisy polynomial regime

Faicel Chamroukhi (USTV/LSIS - France)
Curves clustering

Data: \( n \) independent curves \((y_1, \ldots, y_n)\) (observed)
hidden classes \((h_1, \ldots, h_n)\), hidden regimes \((z_{1k}, \ldots, z_{mk})\) of class \( k \)

Unsupervised learning, clustering and segmentation
(Advances in Data Analysis and Classification (ADAC) 5(4) : 301-321, 2011.)

- Mixture of RHLP models (MixRHLP)

\[
p(y_i|t; \Psi) = \sum_{k=1}^{K} \alpha_k \prod_{j=1}^{m} \sum_{r=1}^{R} \pi_{kr}(t_j; w_k) \mathcal{N}(y_{ij}; \beta_{kr}^T t_j, \sigma_{kr}^2)
\]

- Log-likelihood:

\[
\mathcal{L}(\Psi; Y, t) = \sum_{i=1}^{n} \log \sum_{k=1}^{K} \alpha_k \prod_{j=1}^{m} \sum_{r=1}^{R} \pi_{kr}(t_j; w_k) \mathcal{N}(y_{ij}; \beta_{kr}^T t_j, \sigma_{kr}^2)
\]
Curves clustering

Data: \( n \) independent curves \((y_1, \ldots, y_n)\) (observed)
hidden classes \((h_1, \ldots, h_n)\), hidden regimes \((z_{1k}, \ldots, z_{mk})\) of class \( k \)

Unsupervised learning, clustering and segmentation
(Advances in Data Analysis and Classification (ADAC) 5(4) : 301-321, 2011.)

- Mixture of RHLP models (MixRHLP)

\[
p(y_i|t; \Psi) = \sum_{k=1}^{K} \alpha_k \prod_{j=1}^{m} \sum_{r=1}^{R} \pi_{kr}(t_j; w_k) N(y_{ij}; \beta_{kr}^T t_j, \sigma_{kr}^2)
\]

- Log-likelihood:

\[
\mathcal{L}(\Psi; Y, t) = \sum_{i=1}^{n} \log \sum_{k=1}^{K} \alpha_k \prod_{j=1}^{m} \sum_{r=1}^{R} \pi_{kr}(t_j; w_k) N(y_{ij}; \beta_{kr}^T t_j, \sigma_{kr}^2)
\]

- Maximisation of the log-likelihood by using the EM algorithm (ADAC, 2011)
EM Algorithm

- **Initialisation**: $\Psi^{(0)}$, $q \leftarrow 0$ ($q$ itération)
EM Algorithm

- **Initialisation :** $\Psi^{(0)}$, $q \leftarrow 0$ ($q$ itération)

1. **Étape E : Expectation**

$$Q(\Psi, \Psi^{(q)}) = \mathbb{E} \left[ \mathcal{L}_c(\Psi; Y, t, h, z_1, \ldots, z_K) | Y, t; \Psi^{(q)} \right]$$

$$= \sum_{k=1}^{K} \sum_{i=1}^{n} \tau^{(q)}_{ik} \log \alpha_g + \sum_{k=1}^{K} \sum_{r=1}^{R_k} \sum_{i=1}^{n} \sum_{j=1}^{m} \tau^{(q)}_{ik} \gamma_{ijkr}^{(q)} \log \pi_{kr}(t_j; w_k)$$

$$+ \sum_{k=1}^{K} \sum_{r=1}^{R_k} \sum_{i=1}^{n} \sum_{j=1}^{m} \tau^{(q)}_{ik} \gamma_{ijkr}^{(q)} \log \mathcal{N}(y_{ij}; \beta^T_{kr}r_j, \sigma^2_{kr})$$

$$\tau^{(q)}_{ik} = p(h_i = k|y_i; \Psi^{(q)}) : \text{Posterior prob that } y_i \text{ belongs to class } k$$

$$\gamma_{ijkr}^{(q)} = p(z_{jk} = r|y_{ij}; \Psi^{(q)}) : \text{Posterior prob that } y_{ij} \text{ belong to regime } r \text{ of class } k$$
**EM Algorithm**

- **Initialisation** : $\Psi^{(0)}$, $q \leftarrow 0$ ($q$ itération)

1. **Étape E : Expectation**

   \[
   Q(\Psi, \Psi^{(q)}) = \mathbb{E} \left[ \mathcal{L}_c(\Psi; Y, t, h, z_1, \ldots, z_K) | Y, t; \Psi^{(q)} \right]
   \]

   \[
   = \sum_{k=1}^{K} \tau_{ik}^{(q)} \log \alpha_g + \sum_{k=1}^{K} R_k \sum_{r=1}^{R_k} \sum_{i=1}^{n} \sum_{j=1}^{m} \tau_{ik}^{(q)} \gamma_{ijkr}^{(q)} \log \pi_{kr}(t_j; w_k)
   \]

   \[
   + \sum_{k=1}^{K} \sum_{r=1}^{R_k} \sum_{i=1}^{n} \sum_{j=1}^{m} \tau_{ik}^{(q)} \gamma_{ijkr}^{(q)} \log \mathcal{N}(y_{ij}; \beta_{kr}^T r_j, \sigma_{kr}^2)
   \]

   \[
   \tau_{ik}^{(q)} = p(h_i = k|y_i; \Psi^{(q)}) : \text{Posterior prob that } y_i \text{ belongs to class } k
   \]

   \[
   \gamma_{ijkr}^{(q)} = p(z_{jk} = r|y_{ij}; \Psi^{(q)}) : \text{Posterior prob that } y_{ij} \text{ belong to regime } r \text{ of class } k
   \]

2. **Étape M : Maximisation** : $\Psi^{(q+1)} = \arg \max_\Psi Q(\Psi, \Psi^{(q)})$

Faicel Chamroukhi (USTV/LSIS - France)
**EM Algorithm**

- **Initialisation**: $\Psi^{(0)}$, $q \leftarrow 0$ ($q$ itération)

1. **Étape E : Expectation**

$$Q(\Psi, \Psi^{(q)}) = \mathbb{E} \left[ \mathcal{L}_c(\Psi; \mathbf{Y}, t, h, z_1, \ldots, z_K) | \mathbf{Y}, t; \Psi^{(q)} \right]$$

$$= \sum_{k=1}^{K} \sum_{i=1}^{n} \tau_{ik}^{(q)} \log \alpha_g + \sum_{k=1}^{K} \sum_{r=1}^{R_k} \sum_{i=1}^{n} \sum_{j=1}^{m} \tau_{ik}^{(q)} \gamma_{ijkr}^{(q)} \log \pi_{kr}(t_j; \mathbf{w}_k)$$

$$+ \sum_{k=1}^{K} \sum_{r=1}^{R_k} \sum_{i=1}^{n} \sum_{j=1}^{m} \tau_{ik}^{(q)} \gamma_{ijkr}^{(q)} \log \mathcal{N}(y_{ij}; \beta_{kr}^T r_j, \sigma_{kr}^2)$$

$$\tau_{ik}^{(q)} = p(h_i = k | y_i; \Psi^{(q)}) : \text{Posterior prob that } y_i \text{ belongs to class } k$$

$$\gamma_{ijkr}^{(q)} = p(z_{jk} = r | y_{ij}; \Psi^{(q)}) : \text{Posterior prob that } y_{ij} \text{ belong to regime } r \text{ of class } k$$

2. **Étape M : Maximisation**: $\Psi^{(q+1)} = \operatorname*{arg\,max}_{\Psi} Q(\Psi, \Psi^{(q)})$

- $q \leftarrow q + 1$
M-step:

1. separate maximizations w.r.t the mixing proportions $(\alpha_1, \ldots, \alpha_K)$, the regression parameters $\{\beta_{kr}, \sigma^2_{kr}\}$ and the processes’ parameters $\{w_k\}$.
M-step:

1. separate maximizations w.r.t the mixing proportions \((a_1, \ldots, a_K)\), the regression parameters \(\{\beta_{kr}, \sigma^2_{kr}\}\) and the processes' parameters \(\{w_k\}\).

2. Updating the mixing proportions: 
   \[
   a_1^{(q+1)} = \frac{1}{n} \sum_{i=1}^{n} \gamma^{(q)}_{ik} \quad (k = 1, \ldots, K),
   \]
M-step:

1. separate maximizations w.r.t the mixing proportions \( (\alpha_1, \ldots, \alpha_K) \), the regression parameters \( \{ \beta_{kr}, \sigma^2_{kr} \} \) and the processes' parameters \( \{ \mathbf{w}_k \} \).

2. Updating the mixing proportions: \( \alpha_1^{(q+1)} = \frac{1}{n} \sum_{i=1}^{n} \gamma_{ik}^{(q)} \quad (k = 1, \ldots, K) \),

3. for the regression parameters: separate analytic solutions of weighted least-squares problems

\[
\begin{align*}
\beta_{kr}^{(q+1)} &= \left[ \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ik}^{(q)} \tau_{ijkr} t_j t_j^T \right]^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ik}^{(q)} \tau_{ijkr} y_{ij} t_j \\
\sigma^2_{kr}^{(q+1)} &= \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ik}^{(q)} \tau_{ijkr} (y_{ij} - \beta_{kr}^{T(q+1)} t_j)^2}{\sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ik}^{(q)} \tau_{ijkr}}.
\end{align*}
\]
M-step :

1. separate maximizations w.r.t the mixing proportions $\{\alpha_1, \ldots, \alpha_K\}$, the regression parameters $\{\beta_{kr}, \sigma^2_{kr}\}$ and the processes’ parameters $\{w_k\}$.

2. Updating the mixing proportions : $\alpha_1^{(q+1)} = \frac{1}{n} \sum_{i=1}^{n} \gamma_{ik}^{(q)}$ $\quad (k = 1, \ldots, K)$,

3. for the regression parameters : separate analytic solutions of weighted least-squares problems

   $$\beta_{kr}^{(q+1)} = \left[ \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ik}^{(q)} \tau_{ijkr} t_j t_j^T \right]^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ik}^{(q)} \tau_{ijkr} y_{ij} t_j$$

   $$\sigma^2_{kr}^{(q+1)} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ik}^{(q)} \tau_{ijkr} (y_{ij} - \beta_{kr}^T (q+1) t_j)^2}{\sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ik}^{(q)} \tau_{ijkr}}.$$

4. the maximization w.r.t the logistic processes parameters $\{w_{gk}\}$ consists in solving multinomial logistic regression problems weighted by $\gamma_{igk}^{(q)} \tau_{ijgkr} \Rightarrow$ solved with a multi-class IRLS algorithm.
Experiments on simulated data

Faicel Chamroukhi  (USTV/LSIS - France)
Experiments on simulated data
Curve classification (Discrimination)

Data: \( n \) independent labeled curves \(((y_1, c_1), \ldots, (y_n, c_n))\)

- **Generative Functional discriminant analysis**
- Assign a (new) curve \( y_i \) to the class \( c_i \) using the MAP rule:

\[
\arg \max_{1 \leq g \leq G} \frac{w_g p(y_i | c_i = g, t; \Psi_g)}{\text{Cst}}
\]

- **Classification directly in the space of curves**
Curve classification (Discrimination)

Data: \( n \) independent labeled curves \(((y_1, c_1), \ldots, (y_n, c_n))\)

- Generative Functional discriminant analysis
- Assign a (new) curve \( y_i \) to the class \( c_i \) using the MAP rule:
  \[
  c_i = \arg \max_{1 \leq g \leq G} w_g \underbrace{p(y_i|c_i = g, t; \Psi_g)}_{\text{conditional}} \underbrace{\frac{\text{prior}}{\text{Cst}}}_{\text{prior}}
  \]

Classification directly in the space of curves

There are different ways to model the conditional density \( p(y_i|c_i = g, t; \Psi_g) \):

1. Functional Linear (or Quadratic) Discriminant Analysis (FLDA) (James 2001))
Curve classification (Discrimination)

Homogeneous classes: Functional Linear Discriminant Analysis

- Summarize a class of curves in a curve "model" (the model expectation)
- Distribution of a homogeneous class of curves (RHLP)
  \[ p(\{y_i\}|c_i = g, t; \theta_g) = \prod_i \prod_{j=1}^m \sum_{r=1}^R \pi_{gr}(t_j; w) \mathcal{N}(y_{ij}; \beta_{gr}^T t_j, \sigma_{gr}^2) \]
Curve classification (Discrimination)

Homogeneous classes: Functional Linear Discriminant Analysis

- Summarize a class of curves in a curve “model” (the model expectation)
- Distribution of a homogeneous class of curves (RHLP)
  \[ p(\{y_i\}|c_i = g, t; \theta_g) = \Pi_i \Pi_{j=1}^m \sum_{r=1}^{R} \pi_{gr}(t_j; w) \mathcal{N}(y_{ij}; \beta_{gr}^T t_j, \sigma_{gr}^2) \]
- Parameter Estimation with EM similar to the case of a single curve
  \[ \text{(Neurocomputing - Elsevier, 73(7-9):1210-1221, 2010.)} \]
Curve classification (Discrimination)

Homogeneous classes : Functional Linear Discriminant Analysis

- Summarize a class of curves in a curve “model” (the model expectation)
- Distribution of a homogeneous class of curves (RHLP)
  \[ p(\{y_i\}|c_i = g, t; \theta_g) = \prod_i \prod_{j=1}^m \sum_{r=1}^R \pi_{gr}(t_j; w) \mathcal{N}(y_{ij}; \beta_{gr}^T t_j, \sigma_{gr}^2) \]
- Parameter Estimation with EM similar to the case of a single curve
  \[(Neurocomputing - Elsevier, 73(7-9) :1210-1221, 2010.)\]

Dispersed classes : Functional Mixture Discriminant Analysis

- Summarise the curves by several models, each model is associated with a subclass
- Mixture distribution for each class of curves (MixRHLP)
  \[ p(y_i|c_i = g, t; \Psi_g) = \sum_{k=1}^{K_g} \alpha_{gk} \prod_{j=1}^m \sum_{r=1}^{R_{gk}} \pi_{gkr}(t_j; w_{gk}) \mathcal{N}(y_{ij}; \beta_{gkr}^T t_j, \sigma_{gkr}^2) \]
Curve classification (Discrimination)

Homogeneous classes: Functional Linear Discriminant Analysis

- Summarize a class of curves in a curve "model" (the model expectation)
- Distribution of a homogeneous class of curves (RHLP)
  \[ p(\{y_i\}|c_i = g, t; \theta_g) = \prod_i \prod_{j=1}^{m} \sum_{r=1}^{R} \pi_{gr}(t_j; \mathbf{w}) \mathcal{N}(y_{ij}; \beta_{gr}^T t_j, \sigma_{gr}^2) \]
- Parameter Estimation with EM similar to the case of a single curve
  \[ (\text{Neurocomputing - Elsevier, 73(7-9):1210-1221, 2010}) \]

Dispersed classes: Functional Mixture Discriminant Analysis

- Summarise the curves by several models, each model is associated with a subclass
- Mixture distribution for each class of curves (MixRHLP)
  \[ p(y_i|c_i = g, t; \Psi_g) = \sum_{k=1}^{K_g} \alpha_{gk} \prod_{j=1}^{m} \sum_{r=1}^{R_{gk}} \pi_{gkr}(t_j; \mathbf{w}_{gk}) \mathcal{N}(y_{ij}; \beta_{gkr}^T t_j, \sigma_{gkr}^2) \]
- Parameter estimation via EM as in the clustering case
  \[ (\text{ESANN 2012, IJCNN 2012}) \]
Experiments

- evaluation of the proposed FMDA-MixRHLP approach on simulated data
- comparisons with FLDA approaches using a polynomial regression (FLDA-PR) or a spline regression (FLDA-SR) model (James 2001), and the one that uses a single RHLP model (FLDA-RHLP) (Chamroukhi et al. 2010).
- comparisons with alternative FMDA approaches that use polynomial regression mixtures (FMDA-PRM) (Gaffney 2004), and spline regression mixtures (FMDA-SRM) (Gui 2003)

Evaluation criteria

- the misclassification error rate computed by a 5-fold cross-validation
- the intra-class inertia:
  - For FLDA: inertia $= \sum_g \sum_{i|y_i=g} \|x_i - m_g\|^2$
  - for FMDA: inertia $= \sum_g \sum_{i|y_i=g} \sum_{k=1}^{K_g} \|x_i - m_{gk}\|^2$
Experiments using simulated data

- simulated curves issued from two classes of piecewise noisy functions
- Including a complex shaped class composed of three sub-classes, and a homogeneous class
- Each curve consists of three piecewise regimes and is composed of 200 points
Simulation results

**Figure:** The estimated sub-classes colored according to the partition given by the EM algorithm for the proposed approach (top); Then are presented separately each sub-class of curves with the estimated mean curve in bold line (top sub-plot) and the corresponding logistic probabilities that govern the hidden regimes.
Simulation Results

<table>
<thead>
<tr>
<th>Discrimination Approach</th>
<th>Classif. error rate (%)</th>
<th>Intra-class inertia ($\times 10^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLDA-PR</td>
<td>21</td>
<td>7.1364</td>
</tr>
<tr>
<td>FLDA-SR</td>
<td>19.3</td>
<td>6.9640</td>
</tr>
<tr>
<td>FLDA-RHLP</td>
<td>18.5</td>
<td>6.4485</td>
</tr>
<tr>
<td>FMDA-PRM</td>
<td>11</td>
<td>6.1735</td>
</tr>
<tr>
<td>FMDA-SRM</td>
<td>9.5</td>
<td>5.3570</td>
</tr>
<tr>
<td>FMDA-MixRHLP</td>
<td>5.3</td>
<td>3.8095</td>
</tr>
</tbody>
</table>

- FMDA approaches provide better results compared to FLDA approaches.
- Using a single model for complex-shaped classes (i.e., when using FLDA approaches) is not adapted for complex-shaped classes.
- The proposed FMDA approach based on hidden logistic process regression (FMDA-MixRHLP) outperforms the alternative FMDA approaches thanks to the flexibility of the logistic process well adapted for curves with regime changes.
Application to the study of a railway system

1. Context and objectives

2. Probabilistic modeling with Hidden process for curves

3. Curves classification

4. Applications
   - Switch diagnosis and monitoring
   - Energy of Transportation
   - Assistive robotics

5. Conclusions
**Context**

- Collaboration with la SNCF
- Diagnosis and monitoring of a component of the railway infrastructure

**Switch mechanism**

**Objectives**

- Estimating the state of component (diagnosis)
- Monitor its status over time
Visualization of modeling results on real data

One curves

A class of curves
Visualization of modeling results on real data

One curves

A class of curves
Classification of switch operation signals

- Diagnosis results

<table>
<thead>
<tr>
<th>Approaches</th>
<th>Classification error rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSR-MDA</td>
<td>$13 \pm (4.5)$</td>
</tr>
<tr>
<td>PWPR-MDA</td>
<td>$12 \pm (1.7)$</td>
</tr>
<tr>
<td>HMMR-MDA</td>
<td>$9 \pm (2.25)$</td>
</tr>
<tr>
<td>RHLP-MDA</td>
<td>$4 \pm (1.33)$</td>
</tr>
<tr>
<td>Functional LDA</td>
<td></td>
</tr>
<tr>
<td>PSR-MAP</td>
<td>$7.3 \pm (4.36)$</td>
</tr>
<tr>
<td>PWPR-MAP</td>
<td>$1.82 \pm (5.74)$</td>
</tr>
<tr>
<td>RHLP-MAP</td>
<td>$1.67 \pm (2.28)$</td>
</tr>
</tbody>
</table>
Classification of real data by FMDA

- Database issued from a real french railway diagnosis application
  - two classes:
    - Class 1: curves with no defect or with a minor defect
    - Class 2: curves with a critical defect

**Figure:** 75 switch operation curves from the first class (left) and 45 curves from the second class (right).
Results

Figure: Results obtained with the proposed model for the real curves. The estimated sub-classes for class 1 (top-left) and the corresponding mean curves (top) provided by the proposed approach; Then, we show separately each sub-class of class 1 with the estimated mean curve presented in a bold line (top sub-plot), the polynomial regressors (degree $p = 3$), the corresponding logistic proportions that govern the hidden process, and finally in the bottom plots we show the same results for class 2.


Results

Class 1

Sub-Class 1 of Class 1

Sub-Class 2 of Class 1

Class 2

Regime probabilities

Figure: Results obtained with the proposed model for the real curves. The estimated sub-classes for class 1 (top-left) and the corresponding mean curves (top) provided by the proposed approach; Then, we show separately each sub-class of class 1 with the estimated mean curve presented in a bold line (top sub-plot), the polynomial regressors (degree $p = 3$), the corresponding logistic proportions that govern the hidden process, and finally in the bottom plots we show the same results for class 2.
Results

Figure: Results obtained with the proposed model for the real curves. The estimated sub-classes for class 1 (top-left) and the corresponding mean curves (top) provided by the proposed approach; then, we show separately each sub-class of class 1 with the estimated mean curve presented in a bold line (top subplot), the polynomial regressors (degree $p = 3$), the corresponding logistic proportions that govern the hidden process, and finally in the bottom plots we show the same results for class 2.
## Results

<table>
<thead>
<tr>
<th>Discrimination approach</th>
<th>Classif. error rate (%)</th>
<th>Intra-class inertia ($\times 10^9$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLDA-PR</td>
<td>11.5</td>
<td>10.7350</td>
</tr>
<tr>
<td>FLDA-SR</td>
<td>9.53</td>
<td>9.4503</td>
</tr>
<tr>
<td>FLDA-RHLP</td>
<td>8.62</td>
<td>8.7633</td>
</tr>
<tr>
<td>FMDA-PRM</td>
<td>9.02</td>
<td>7.9450</td>
</tr>
<tr>
<td>FMDA-SRM</td>
<td>8.50</td>
<td>5.8312</td>
</tr>
<tr>
<td><strong>FMDA-MixRHLP</strong></td>
<td><strong>6.25</strong></td>
<td><strong>3.2012</strong></td>
</tr>
</tbody>
</table>

**Table:** Obtained results for the real curves.
Curves clustering

Real Data: 115 signals
Curves clustering

Real data: graphical results
Temporal monitoring of the curves

non-homogeneous autoregressive HMM : prediction results

Sequence of curves

prediction prof the system state

True sequence

Estimated sequence
**Modeling of impedance spectra of Fuel Cells**

- **Context**: energy of transportation
- **Objective**: Estimation of the fuel cell lifetime.
- **Representation of impedance spectrum data**
Modeling of impedance spectra of Fuel Cells

- **Context**: energy of transportation
- **Objective**: Estimation of the fuel cell lifetime.
- **Representation of impedance spectrum data**
  - a probabilistic approach: the RHLP model (*IEEE ICMLA 2009*)
Activity recognition from acceleration data

- **Context**: Assistive Robotics
- **Objective**: Gesture recognition for helping ageing people
- **Data**: accelerations, ...

![Images showing various activities](a.png b.png c.png d.png e.png f.png)
Unsupervised classification of acceleration data

- Joint segmentation of temporal multidimensional data
- Multidimensional regression with a hidden process process
  - *IEEE Transactions on Automation Science and Engineering (Accepted)*
  - *Neurocomputing (in revision)*
Conclusions and perspectives

1. Context and objectives
2. Probabilistic modeling with Hidden process for curves
3. Curves classification
4. Applications
5. Conclusions
Conclusions

1. Curves modeling:
   - A Hidden process regression model for curves with regime changes
     - Suited to abrupt and/or smooth regime changes
     - Unsupervised Learning based on the EM algorithm
Conclusions

1. Curves modeling:
   - A Hidden process regression model for curves with regime changes
     - Suited to abrupt and/or smooth regime changes
     - Unsupervised Learning based on the EM algorithm

2. Curves classification (Discrimination and clustering)
   - Direct Classification of curves without preliminary representation
   - Adapted to heterogeneous classes
Conclusions

1. Curves modeling:
   - A Hidden process regression model for curves with regime changes
     - Suited to abrupt and/or smooth regime changes
     - Unsupervised Learning based on the EM algorithm

2. Curves classification (Discrimination and clustering)
   - Direct Classification of curves without preliminary representation
   - Adapted to heterogeneous classes

3. Application to real problems
Merci de votre attention!