Mixture-model-based signal denoising

Allou Samé[1], Latifa Oukhellou[1,2], Etienne Côme[1], Patrice Aknin[1*]

[1]Institut National de Recherche sur les
Transports et leur Sécurité (INRETS)
2 avenue du général Malleret-Joinville
94114, Arcueil, France

[2]CERTES, Université Paris XII
61 avenue du général De Gaulle
94100 Créteil, France

Abstract

This paper proposes a new signal denoising methodology for dealing with asymmetrical noises. The adopted strategy is based on a regression model where the noise is supposed to be additive and distributed following a mixture of Gaussian densities. The parameters estimation is performed using a Generalized EM (GEM) algorithm. Experimental studies on simulated and real signals in the context of a diagnosis application in the railway domain reveal that the proposed approach performs better than the least-squares and wavelets methods.

Key words   Denoising, Asymmetrical noise, Regression, Gaussian mixture model, EM algorithm, GEM algorithm

1   Introduction

Denoising or noise reduction is a subject of research in both signal and image processing. The aim is to remove the additive noise from the signal without distorting it. Various linear and non linear methods have been proposed to solve this problem. They lead to different solutions depending on the noise model and the signal properties which are closely linked to applications.

One of the most popular and recent denoising approaches is based on wavelet analysis (Mallat 1989, 1998). It consists in applying thresholding algorithms to the wavelet coefficients in order to extract the most significant of them. This method allows an unknown function to be reconstructed from

---

[*]Corresponding author (aknin@inrets.fr)

noisy data, but its theoretical framework involves a normal noise density and requires an optimal choice of the threshold. For this purpose, several thresholding rules have been proposed and successfully applied (Dohono et al. (1996), Ogden and Parzen (1996).

Many statistical approaches like least-squares method have also been proposed. In this paper, we explore a strategy for denoising based on mixture models which are widely used in clustering (Dempster et al. 1977, McLachlan and Krishnan 1997). The basic idea of the proposed approach is to fit a regression model on the signal with the assumption that the error term (or additive noise) follows a mixture of normal distributions. The parameters of such a model are then estimated by means of a Generalized Expectation Maximization (GEM) algorithm.

It should be noticed that this approach is different from the wavelet method, which is dedicated to reconstruct an unknown signal in a white Gaussian noise. Here, we assume a parametric regression model of the signal and a mixture of normal distributions for the noise. In many situations where the normal assumption cannot be satisfied, the normal mixtures seem to be suitable because they can represent a wide variety of noise densities including both symmetrical and non-symmetrical cases (Marron and Wand 1992).

The paper is organized as follows. At first, we present the regression model and introduce some preliminary notations. Then, we detail the parameter estimation procedure via a GEM algorithm. Section four gives the performances of the approach illustrated on simulated examples and section five is concerned with an application in the railway domain.

2   Additive noise and mixture model

A signal is supposed to be represented by an independent sample $((x_1, y_1), \ldots, (x_n, y_n))$ where variables x and y, defined on $I\!R$, represent the dependent variable (for example the time) and the independent variable (the signal at time x), respectively. Polynomial regression functions have been chosen, for their simplicity and their coherence with many real signals. An observed signal is then modeled by

$$y = \boldsymbol{a}^T \boldsymbol{x} + \varepsilon, \tag{1}$$

where $\boldsymbol{a} = (a_0, a_1, a_2, \ldots, a_m)^T$, $\boldsymbol{x} = (1, x, x^2, \ldots, x^m)^T$, m is the polynomial order and ε is an additive noise whose distribution does not depend on x.

Usually, ε is assumed to be normally distributed with zero mean and variance σ^2, and the estimation of the model parameters proceeds by solving a classical least-squares problem, but in certain practical problems the presence of other sources of noise results in an asymmetrical non-normal noise which makes the least-squares method unsuitable.

To solve this problem, the noise ε is assumed to be centered and distributed according to a mixture of K normal densities

$$f(\varepsilon) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\varepsilon; \mu_k, \sigma_k^2) \tag{2}$$

where $\mathcal{N}(\cdot; \mu_k, \sigma_k^2)$ is the normal distribution density with mean μ_k and variance σ_k^2, the $\pi_k's$ are the proportions of the mixture. The proportions π_k and the means μ_k satisfy the constraints $\sum_{k=1}^{K} \pi_k = 1$ and $\sum_{k=1}^{K} \pi_k \mu_k = 0$. The component density $\mathcal{N}(\cdot; \mu_k, \sigma_k^2)$ can thus be regarded as one source of noise. The well known flexibility of mixture distributions allows the regression model to handle both symmetrical and non-symmetrical noises (Bartolucci and Scaccia 2005).

The next section shows how the parameters of the proposed model can be estimated.

3   Parameters estimation via the GEM algorithm

The model parameters are estimated by maximizing the conditional likelihood or equivalently the conditional log-likelihood defined by

$$L(\boldsymbol{\Psi}) \;\; = \;\; \sum_{i=1}^{n} \log p(y_i | x_i; \boldsymbol{\Psi}), \tag{3}$$

where $\boldsymbol{\Psi} = (\boldsymbol{a}, \pi_1, \dots, \pi_K, \mu_1, \dots, \mu_K, \sigma_1^2, \dots, \sigma_K^2)$ is the parameter vector to be estimated. It can be easily verified that, conditionally on x, the variable y has the mixture distribution density

$$p(y|x; \boldsymbol{\Psi}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(y \; ; \; \boldsymbol{a}^T \boldsymbol{x} + \mu_k \; , \sigma_k^2).$$

The (observed) log-likelihood is then written

$$L(\boldsymbol{\Psi}) = \sum_{i=1}^{n} \log \Big[\sum_{k=1}^{K} \pi_k \mathcal{N}\big(y_i \; ; \; \boldsymbol{a}^T \boldsymbol{x}_i + \mu_k, \sigma_k^2\big) \Big], \tag{4}$$

where $\boldsymbol{x}_i = (1, x_i, x_i^2, \dots, x_i^m)$. This equation clearly highlights that the proposed model is simply a constrained Gaussian mixture of regressions (Aitkin and Wilson 1980, DeSarbo and Cron 1988) in which the coefficient vector $\boldsymbol{a}$ is common for each component.

The Expectation-Maximization (EM) algorithm (Dempster et al. 1977) is used to perform the maximization of the log-likelihood (4) which can not be solved directly. Let us recall that the EM algorithm requires the specification

3

of complete data whose log-likelihood can be maximized more easily than the observed log-likelihood (4). In this situation, the complete data are obtained by adding to each observation (x_i, y_i) its component membership variable $z_i \in \{1, \dots, K\}$. Using the binary coding of z_i, $z_{ik} = 1$ if $z_i = k$ and $z_{ik} = 0$ otherwise, the complete data log-likelihood is written as

$$L_C(\boldsymbol{\Psi}) = \sum_{i=1}^{n} \sum_{k=1}^{K} z_{ik} \log \left[\pi_k \mathcal{N}\left(y_i \; ; \; \boldsymbol{a}^T \boldsymbol{x}_i + \mu_k, \sigma_k^2\right) \right], \tag{5}$$

Given an initial parameter vector $\boldsymbol{\Psi}^{(0)}$, the EM algorithm consists in alternating the two following steps until convergence.

E step : Expectation

This step consists in evaluating the expectation of the complete data log-likelihood conditionally on the observed data and the current parameter vector $\boldsymbol{\Psi}^{(q)}$:

$$\begin{aligned} Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(q)}) &= E\left[L_C(\boldsymbol{\Psi}) | x_1, \dots, x_n, y_1, \dots, z_n; \boldsymbol{\Psi}^{(q)} \right] \\ &= \sum_{i=1}^{n} \sum_{k=1}^{K} t_{ik}^{(q)} \log \left[\pi_k \mathcal{N}\left(y_i \; ; \; \boldsymbol{a}^T \boldsymbol{x}_i + \mu_k, \sigma_k^2\right) \right] \end{aligned} \tag{6}$$

where $L_C(\boldsymbol{\Psi})$ is the complete data log-likelihood and

$$\begin{aligned} t_{ik}^{(q)} &= E(z_{ik} | x_i, y_i; \boldsymbol{\Psi}^{(q)}) \\ &= P(z_{ik} = 1 | x_i, y_i; \boldsymbol{\Psi}^{(q)}) \\ &= \frac{\pi_k^{(q)} \mathcal{N}(y_i; \; \boldsymbol{a}^{(q)T} \boldsymbol{x}_i + \mu_k^{(q)}, \sigma_k^{(q)2})}{\sum_{\ell=1}^{K} \pi_\ell^{(q)} \mathcal{N}(y_i; \; \boldsymbol{a}^{(q)T} \boldsymbol{x}_i + \mu_\ell^{(q)}, \sigma_\ell^{(q)2})} \end{aligned} \tag{7}$$

is the posterior probability that y_i originates from the k*th* mixture component, given x_i. Thus, the E step simply requires the computation of the posterior probabilities $t_{ik}^{(q)}$.

M step : Maximization

This step consists in computing the parameter vector $\boldsymbol{\Psi}^{(q+1)}$ that maximizes with respect to $\boldsymbol{\Psi}$, the quantity $Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(q)})$ which, in our situation, can be written

$$\begin{aligned} Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(q)}) &= \sum_{k,i} t_{ik}^{(q)} \log \pi_k - \\ &\frac{1}{2} \sum_{k,i} t_{ik}^{(q)} \left[\log(2\pi) + \log \sigma_k^2 + \frac{(y_i - (\boldsymbol{a}^T \boldsymbol{x}_i + \mu_k))^2}{\sigma_k^2} \right]. \end{aligned} \tag{8}$$

4

As in the classical Gaussian mixture situation, it can be verified that the proportions maximizing Q are given by

$$\pi_k^{(q+1)} = \frac{\sum_{i=1}^n t_{ik}^{(q)}}{n}. \tag{9}$$

The simultaneous maximization with respect to the parameters $\boldsymbol{a}$, (μ_k) and (σ_k^2) cannot be performed analytically. But it can be shown that, for fixed parameters (σ_k^2), the optimum parameter vector $[\boldsymbol{a}^T, \mu_1, \ldots, \mu_K]$ is given by

$$[\boldsymbol{a}^T, \mu_1, \ldots, \mu_K]^T = \Big[\sum_{k=1}^K \mathcal{H}_k^T \mathcal{H}_k\Big]^{-1} \Big[\sum_{k=1}^K \mathcal{H}_k^T \mathcal{Y}_k\Big], \tag{10}$$

with

$$\mathcal{H}_k = \frac{1}{\sigma_k} \cdot diag\left(\sqrt{t_{1k}^{(q)}}, \ldots, \sqrt{t_{nk}^{(q)}}\right) \cdot \mathbf{X}_k$$

$$\mathcal{Y}_k = \frac{1}{\sigma_k} \cdot diag\left(\sqrt{t_{1k}^{(q)}}, \ldots, \sqrt{t_{nk}^{(q)}}\right) \cdot \mathbf{Y},$$

where $diag(e_1, \ldots, e_n)$ is the diagonal matrix with diagonal elements $e_1, \ldots, e_n$, and

$$\mathbf{X}_k = \begin{bmatrix} \boldsymbol{x}_1^T & 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{x}_i^T & 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{x}_n^T & 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix}.$$

Notice that $\mathbf{X}_k$ is a $n \times (K+m+1)$ matrix with all elements equal 0 except those in the columns $1, \ldots, m+1$ and $k+m+1$.

On the other hand, for fixed parameters $\boldsymbol{a}$ and (μ_k), it can be shown that the parameter σ_k^2 maximizing Q is given by

$$\sigma_k^2 = \frac{\sum_{i=1}^n t_{ik}^{(q)} \left(y_i - (\boldsymbol{a}^T \boldsymbol{x}_i + \mu_k)\right)^2}{\sum_{i=1}^n t_{ik}^{(q)}}. \tag{11}$$

Thus, we propose to compute the parameter vector $\boldsymbol{\Psi}^{(q+1)}$ as follows:

(a) compute $\boldsymbol{a}^{(q+1)}, \mu_1^{(q+1)}, \ldots, \mu_K^{(q+1)}$ using equation (10), where the old estimates $\sigma_k^{(q)2}, \ldots, \sigma_K^{(q)2}$ have been inserted in the right hand side;

(b) compute $\sigma_k^{(q+1)2}$ using equation (11), where the parameters $\boldsymbol{a}^{(q+1)}$ and $\mu_k^{(q+1)}$ have been inserted in the right hand side.

5

The strategy described by (a) and (b), which fulfil $Q(\boldsymbol{\Psi}^{(q+1)}; \boldsymbol{\Psi}^{(q)}) \geq Q(\boldsymbol{\Psi}^{(q)}; \boldsymbol{\Psi}^{(q)})$, is commonly used when the maximization cannot be directly performed and also known as a conditional-M step (McLachlan and Krishnan 1997). It provides a Generalized EM algorithm (GEM) (Dempster et al. 1977) whose convergence properties are the same as those of the EM algorithm. Usually, the GEM algorithm is iterated for a number of random initial parameters $\boldsymbol{\Psi}^{(0)}$ and only the parameter that provides the largest log-likelihood value is adopted.

4 Experiments on simulated signals

This section is designed to evaluate the performance of the proposed GEM algorithm as regards the quality of estimation, by using simulated signals. For this purpose, results obtained with GEM are compared with those of the least-squares and wavelets methods. Let us remind that least-squares and wavelet denoising methods (Donoho et al. 1996) are based on the assumption that the noise has a zero mean normal distribution.

4.1 Simulation protocol

Each simulated signal is generated according to a polynomial regression function $g(x)$ (the denoised signal that is to be recovered) and an asymmetrically distributed noise. For all simulated signals, we consider a sample size $n = 600$ and the value $x_i = i$ for the dependant variable ($i = 1, ..., n$). In the following, we describe the setting of the polynomial coefficients, the type of asymmetrical noise distribution, the criteria used to measure the estimation quality, and the performed experiments.

Polynomial coefficients

We focus here only on second degree polynomial functions as the denoised signal, in accordance with the real measured signals described in section 5. Since no significant influence of the regression coefficients on the quality of estimation has been observed, only one polynomial function with coefficients vector $\boldsymbol{a} = (1.2; 3 \cdot 10^{-5}; -5 \cdot 10^{-5})$ is considered in all the described simulations below, i.e., $g(x) = 1.2 + 3 \cdot 10^{-5}x - 5 \cdot 10^{-5}x^2$. Relatively small values ($-5 \cdot 10^{-5}$ and $3 \cdot 10^{-5}$) of the polynomial coefficients have been chosen to get simulated signals whose amplitude is very close to the real signals amplitude.

Asymmetrical noise distributions

Two different types of asymmetric distributions have been adopted for the noise generation:

- a two-components Gaussian mixture distribution with density

$$f(\varepsilon; \pi_1, \pi_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = \pi_1 \mathcal{N}(\varepsilon; \mu_1, \sigma_1^2) + \pi_2 \mathcal{N}(\varepsilon; \mu_2, \sigma_2^2) \qquad (12)$$

with mean $\sum_{k=1}^{2} \pi_k \mu_k = 0$;

- a three-parameters Weibull distribution

$$f(\varepsilon; \kappa, \lambda, \alpha) = \frac{\kappa}{\lambda} \left(\frac{\varepsilon - \alpha}{\lambda} \right)^{\kappa-1} \exp\left[-\left(\frac{\varepsilon - \alpha}{\lambda} \right)^{\kappa} \right] \mathbf{1}_{]\alpha; +\infty[}(\varepsilon), \qquad (13)$$

with $\alpha = -\lambda \Gamma(1 + \frac{1}{\kappa})$, Γ being the gamma function. This particular choice of the parameter α ensures a zero mean distribution.

Figures 1 and 2 show examples of signals simulated according to the adopted polynomial regression and the two noise distributions (Gaussian mixture and Weibull).
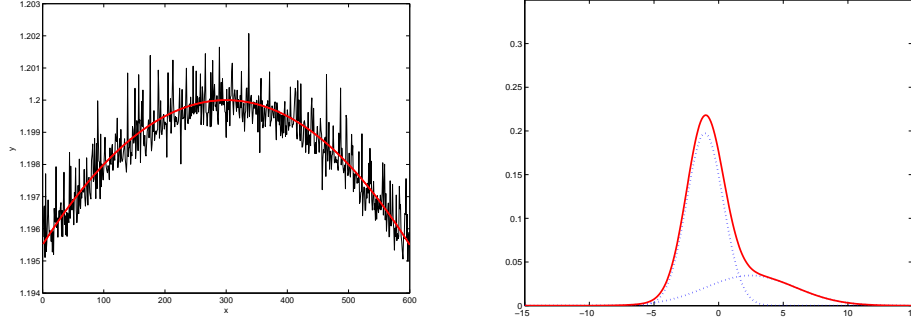

Figure 1: (left) Example of a signal simulated according to a second degree polynomial function and a Gaussian mixture distribution, and its true denoised signal; (right) corresponding noise mixture distribution and its component densities

Evaluation criteria

The quality of estimation provided by GEM, least-squares and wavelets is jointly evaluated by the following criteria measuring the quality of the estimated (denoised) signal and the quality of the noise estimation, respectively:

- The L_2 distance between the true regression function $g(\boldsymbol{x}_i)$ with parameter $\boldsymbol{a}$ and the estimated regression function $\widehat{g}(\boldsymbol{x}_i)$ with parameter $\widehat{\boldsymbol{a}}$, defined by

$$D_2(g, \widehat{g}) = \frac{1}{n} \sum_{i=1}^{n} [\boldsymbol{a}^T \boldsymbol{x}_i - \widehat{\boldsymbol{a}}^T \boldsymbol{x}_i]^2; \qquad (14)$$
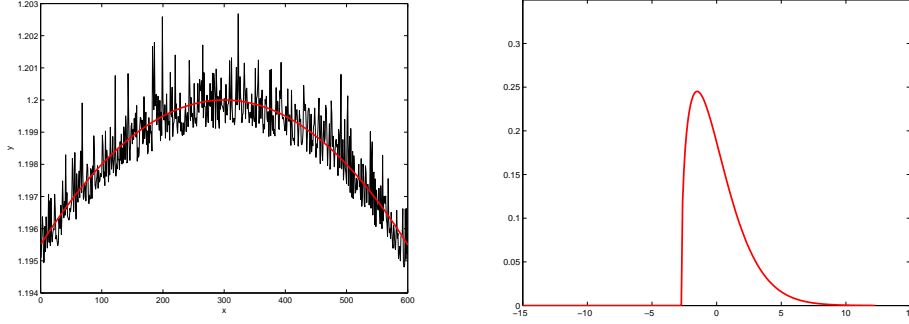
Figure 2: (left) Example of a signal simulated according to a second degree polynomial function and a Weibull distribution, and its true denoised signal; (right) corresponding noise Weibull distribution

- The Kullback-Leibler divergence between the estimated noise distribution $\widehat{f}$ and the true noise distribution f,

$$KL(f, \widehat{f}) = \int_{\mathbb{R}} f(\boldsymbol{x}) \log \frac{f(\boldsymbol{x})}{\widehat{f}(\boldsymbol{x})} d\boldsymbol{x}. \tag{15}$$

The Kullback-Leibler divergence is computed using Monte Carlo integration method since its evaluation is not analytically straightforward in the case of Gaussian mixtures.

Performed experiments

Four sets of experiments have been performed for this simulation study.

- The first set of experiments studies the quality of the estimates when the ratio σ_1^2/σ_2^2 of the Gaussian mixture variances varies in the set $\{1/5, 1/4, 1/3, 1/2, 1\}$, for equal proportions $\pi_1 = \pi_2 = 0.5$, variance $\sigma_1^2 = 2$, and $\Delta = |\mu_1 - \mu_2| = 4$.

- The second set studies the quality of the estimates when the ratio π_1/π_2 of the Gaussian mixture proportions varies in the set $\{1/3, 1/2, 1, 2, 3\}$, for equal variances $\sigma_1^2 = \sigma_2^2 = 4$, and $\Delta = 4$.

- The third set studies the quality of the estimates when the distance Δ between the means of the Gaussian mixture components varies in the set $\{0, 5, 10, 16, 21\}$, for proportions $\pi_1 = 0.7$, $\pi_2 = 0.3$ and variances $\sigma_1^2 = 2$, $\sigma_2^2 = 10$.

- The fourth set of experiments studies the quality of the estimates when the parameter λ of the Weibull distribution varies in the set $\{5, 10, 15, 20, 25, 30\}$, for a parameter $\kappa = 1.4$.

8

4.2 Results

Table 1 represents, for the four sets of experiments, the D_2 distance and the KL divergence obtained with the methods GEM (applied to $K = 2$ and $K = 3$ components), least-squares and wavelets. Notice that each value of D_2 and KL is an average over 30 different simulated samples (Monte-Carlo simulations). As it can be seen, GEM applied with $K = 2$ or $K = 3$ performs globally better than its competitors, for any modification of noise distribution parameters σ_1^2/σ_2^2, π_1/π_2, Δ or α, in terms of the D_2 distance. Not surprisingly, the KL divergence for GEM is distinctly better for any situation; this behavior illustrates that the mixture model is more suitable for dealing with various situation of asymmetrical noise distribution. When the distance Δ or the Weibull parameter λ increases, it can be observed that the D_2 distance for all methods increases. In practice, this phenomenon, even more pronounced for least-squares and wavelets, is due to an increase of the noise variance, in addition to the noise asymmetry. It can also be observed that the number of components $K = 3$ is more adapted for Weibull distribution.

5   Application to defect diagnosis in the railway domain

In the railway domain, signal denoising is an important preprocessing task for defect diagnosis. Most of the automatic diagnosis systems are indeed based on the analysis of measurement signals which are often corrupted with noise. Our application concerns the diagnosis of the track/vehicle transmission system in the French high speed lines (Aknin et al. 2003). The identification of two sources of noise led us to consider that the noise is distributed according to a mixture of two normal densities, rather than a single normal density. The first task consists in segmenting the complete signal into local arches with a piecewise approach, and the second task is focused on the denoising procedure on each arch. Figure 3 shows an example of a measurement signal.

The evaluation of the proposed approach performances on these experimental signals is not possible because the original signals and noise distributions are not available. For this reason, we choose to test the proposed denoising algorithm on signals whose original denoised signals are given by an electrical model, which allows to achieve realistic simulations of the system (Aknin et al. 2003). We use two different denoised signals called DS1 and DS2, composed of $n = 778$ obervations (see figure 4). The same noise distributions (Gaussian mixture and Weibull) as for the simulation study are added to DS1 and DS2. The quality of denoised signals reconstruction is evaluated using the D_2 distance. Unlike the cases studied in section 4, the signals considered in this part are generated with non polynomial regression

Table 1: D_2 distance and KL divergence in relation to σ_1^2/σ_2^2, π_1/π_2, Δ and parameter λ of the Weibull distribution, obtained with GEM (applied with $K = 2$ and $K = 3$), least-squares and wavelets.

$\frac{\sigma_1^2}{\sigma_2^2}$	D_2				KL			
	GEM (K=2)	GEM (K=3)	LS	Wav.	GEM (K=2)	GEM (K=3)	LS	Wav.
1/5	0.029	0.028	0.043	0.137	0.005	0.009	0.083	0.083
1/4	0.029	0.034	0.037	0.124	0.005	0.007	0.069	0.070
1/3	0.027	0.031	0.032	0.112	0.004	0.009	0.059	0.052
1/2	0.024	0.027	0.023	0.088	0.003	0.007	0.032	0.032
1	0.019	0.019	0.026	0.079	0.004	0.007	0.029	0.029
$\frac{\pi_1}{\pi_2}$	D_2				KL			
	GEM (K=2)	GEM (K=3)	LS	Wav.	GEM (K=2)	GEM (K=3)	LS	Wav.
1/3	0.034	0.037	0.042	0.175	0.004	0.007	0.037	0.037
1/2	0.060	0.062	0.065	0.161	0.005	0.008	0.034	0.034
1	0.042	0.046	0.052	0.151	0.004	0.009	0.029	0.029
2	0.043	0.051	0.052	0.184	0.004	0.008	0.033	0.033
3	0.038	0.047	0.041	0.117	0.004	0.009	0.037	0.037
Δ	D_2				KL			
	GEM (K=2)	GEM (K=3)	LS	Wav.	GEM (K=2)	GEM (K=3)	LS	Wav.
0	0.017	0.019	0.019	0.068	0.006	0.007	0.042	0.043
5	0.024	0.025	0.034	0.103	0.005	0.008	0.130	0.131
10	0.022	0.023	0.045	0.164	0.004	0.008	0.212	0.212
16	0.046	0.047	0.093	0.236	0.009	0.012	0.357	0.356
21	0.082	0.082	0.201	0.529	0.013	0.015	0.539	0.540
λ	D_2				KL			
	GEM (K=2)	GEM (K=3)	LS	Wav.	GEM (K=2)	GEM (K=3)	LS	Wav.
4	0.026	0.025	0.036	0.118	0.068	0.039	0.176	0.176
6	0.050	0.036	0.076	0.237	0.067	0.038	0.176	0.177
8	0.106	0.093	0.146	0.443	0.068	0.037	0.175	0.176
10	0.140	0.127	0.181	0.524	0.069	0.038	0.176	0.177
12	0.183	0.129	0.258	1.172	0.068	0.038	0.176	0.177

functions but we try to approximate them with second degree polynomial functions.

Results obtained for DS1 and DS2 using the same noise distribution parameters as in the simulation study are shown in tables 2 and 3. It will be observed that, once again, our proposed GEM algorithm performs better than least-squares and wavelets. The relative behavior of the compared methods are almost the same as for the simulation study.

In rare situations where variances have small values, wavelets denoising has better performances than GEM and least-squares. This remark can be illustrated, for instance, in the case of the denoised signal DS2, by setting $\sigma_1^2 = 5 \cdot 10^{-4}$, $\pi_1 = \pi_2 = 0.5$, $\Delta = 0.05$ and choosing values of σ_2^2 corresponding to different ratios σ_1^2/σ_2^2 (see Table 4). The good performances of wavelets in this case, can be attributed to the non polynomial nature of the denoised signal DS2.
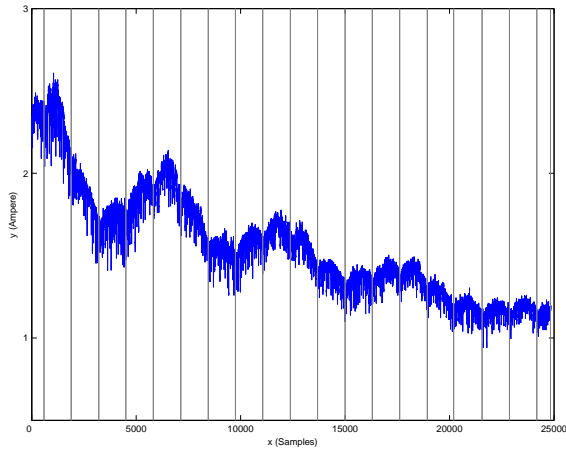
Figure 3: Experimental signal recorded on a track/vehicle transmission system
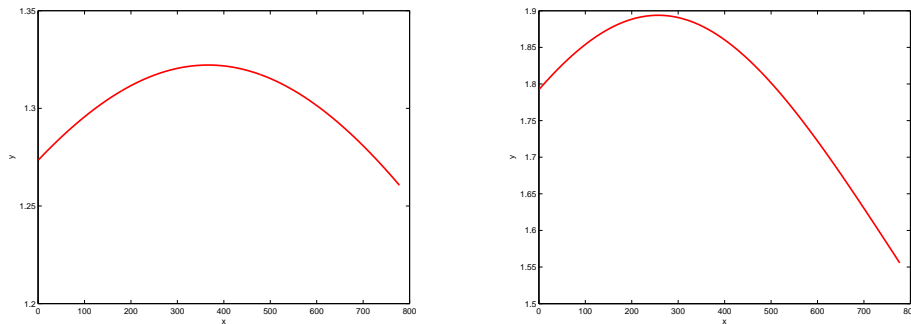

Figure 4: Denoised signal DS1 (left) and DS2 (right) given by an electrical model

Figure 5 shows the denoised signal corresponding to the real signal displayed in figure 3.

6   Conclusion

An original methodology for signal denoising was proposed in this paper. This methodology is based on a polynomial regression model where the noise is supposed to be additive and distributed following a Gaussian mixture distribution. A Generalized EM (GEM) algorithm is developed to estimate both the noise distribution and the polynomial regression coefficients. The experimental study on simulated signals reveals that the GEM algorithm provides reasonably good results for relatively low degree of the regression function, compared to the standard least-squares and wavelets methods. When the noise is distributed according to certain asymmetric distributions like

Table 2: D_2 distance in relation to σ_1^2/σ_2^2, π_1/π_2, Δ and parameter λ of the Weibull distribution; GEM, least-squares and wavelets are applied to noised signals based on DS1.

$\frac{\sigma_1^2}{\sigma_2^2}$	D_2				$\frac{\pi_1}{\pi_2}$	D2			
	GEM (K=2)	GEM (K=3)	LS	Wav.		GEM (K=2)	GEM (K=3)	LS	Wav.
1/5	0.025	0.026	0.040	0.137	1/3	0.024	0.025	0.032	0.127
1/4	0.019	0.019	0.031	0.115	1/2	0.025	0.028	0.031	0.144
1/3	0.019	0.019	0.026	0.091	1	0.027	0.028	0.031	0.165
1/2	0.023	0.024	0.027	0.104	2	0.029	0.031	0.034	0.136
1	0.014	0.015	0.016	0.067	3	0.039	0.039	0.043	0.133
Δ	D_2				λ	D2			
	GEM (K=2)	GEM (K=3)	LS	Wav.		GEM (K=2)	GEM (K=3)	LS	Wav.
0	0.012	0.012	0.015	0.055	4	0.012	0.018	0.029	0.112
5	0.018	0.017	0.031	0.098	6	0.046	0.039	0.069	0.239
10	0.030	0.030	0.054	0.175	8	0.069	0.063	0.118	0.422
16	0.031	0.031	0.091	0.318	10	0.122	0.093	0.151	0.601
21	0.043	0.043	0.125	0.512	12	0.153	0.126	0.207	0.793

Table 3: D_2 distance in relation to σ_1^2/σ_2^2, π_1/π_2, Δ and parameter λ of the Weibull distribution; GEM, least-squares and wavelets are applied to noised signals based on DS2.

$\frac{\sigma_1^2}{\sigma_2^2}$	D_2				$\frac{\pi_1}{\pi_2}$	D2			
	GEM (K=2)	GEM (K=3)	LS	Wav.		GEM (K=2)	GEM (K=3)	LS	Wav.
1/5	0.021	0.021	0.028	0.109	1/3	0.027	0.028	0.029	0.118
1/4	0.022	0.022	0.035	0.159	1/2	0.029	0.030	0.036	0.166
1/3	0.021	0.024	0.027	0.121	1	0.038	0.045	0.042	0.150
1/2	0.019	0.019	0.022	0.090	2	0.023	0.023	0.025	0.150
1	0.015	0.016	0.016	0.086	3	0.031	0.032	0.039	0.171
Δ	D_2				λ	D2			
	GEM (K=2)	GEM (K=3)	LS	Wav.		GEM (K=2)	GEM (K=3)	LS	Wav.
0	0.014	0.015	0.014	0.083	4	0.022	0.018	0.031	0.115
5	0.017	0.018	0.024	0.086	6	0.040	0.032	0.065	0.285
10	0.026	0.026	0.049	0.207	8	0.053	0.047	0.076	0.458
16	0.025	0.025	0.064	0.339	10	0.109	0.091	0.159	0.645
21	0.044	0.044	0.133	0.591	12	0.171	0.141	0.206	1.043

Weibull, the experimental study shows that the number of component $K = 3$ is more adapted than $K = 2$. This point suggests to adopt a suitable strategy for choosing the optimal number of components of the mixture as regards of the likelihood criterion and the complexity of the model. For this purpose the adaptation of Bayesian Information Criteria (BIC) (Schwarz 1978) could be an interesting direction of work. Final results on physical signals in the context of a real application in the railway domain show that the GEM algorithm can represent an efficient mean for denoising real signals in presence

Table 4: Results obtained for DS2+noise in the case of small values of the mixture component variances : $\sigma_1^2 = 5.10^{-4}$.

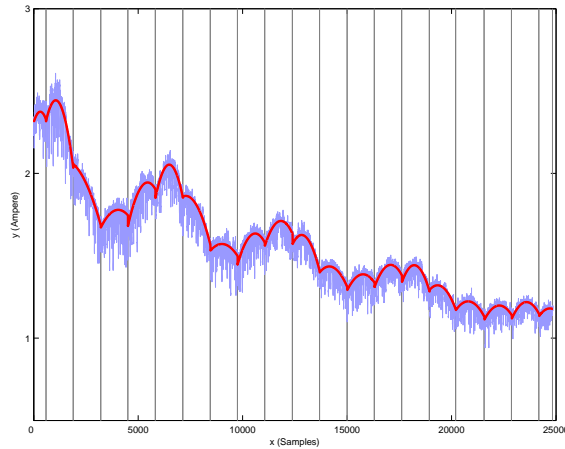| $\frac{\sigma_1^2}{\sigma_2^2}$	D_2 ($\times 10^{-4}$)			
	GEM (K=2)	GEM (K=3)	LS	Wav.
1/5	0.56	0.57	0.57	0.39
1/4	0.56	0.56	0.56	0.30
1/3	0.56	0.56	0.56	0.30
1/2	0.54	0.54	0.54	0.22
1	0.54	0.54	0.54	0.16

Figure 5: Denoised signal corresponding to the experimental signal of figure 3.

of asymmetrical noise.

References

Aitkin M, Wilson G T (1980) Mixture models, outliers and the EM algorithm. Technometrics **22**, pp. 325-331

Aknin P, Oukhellou L, Vilette F (2003) Track circuit diagnosis by automatic analysis of inspection car measurements. 6th World Congress on Railway Research, Edinburgh

Bartolucci F, Scaccia L (2005) The use of mixtures for dealing with non-normal regression errors. Computational Statistics and Data Analysis **48**, pp. 821-834

Dempster A P, Laird N M, Rubin D B (1977) Maximum likelihood from

incomplete data via the EM algorithm. Journal of the Royal Statistical Society Series B **39**, pp. 1-38

DeSarbo W S, Cron W L (1988) A maximum likelihood methodology for clusterwise linear regression. Journal of classification **5**, pp. 249-282

Donoho D L, Johnstone I M, Kerkyacharian G, Picard D (1996) Density estimation by wavelet thresholding. Annals of Statistics **24**, pp. 508-539

Mallat S (1989) A theory for multiresolution signal decomposition: the wavelet representation. IEEE Pattern Analysis and Machine Intelligence **11**, pp. 674-693

Mallat S (1998) A wavelet tour of signal processing. Academic Press

Marron J S, Wand M P (1992) Exact mean integrated squared error. Annals of Statistics **20**, pp. 712-736

McLachlan G, Krishnan T (1997) The EM Algorithm and Extentions. Wiley, New York

Ogden T, Parzen E (1996) Data dependent wavelet thresholding in nonparametric regression with change-point applications. Computational Statistics and Data Analysis **22**, pp. 53-70

Schwarz G (1978) Estimating the number of components in a finite mixture model. Annals of Statistics **6**, pp 461-464