Modeling Infrared Spectra : an Algorithm for an Automatic and Simultaneous Analysis

C. Butucea[1] & J.-F. Delmas[2] & A. Dutfoy[3] & C. Hardy[2,3]

[1]*CREST, ENSAE, IP Paris, France. E-mail: cristina.butucea@ensae.fr*
[2]*CERMICS, Ecole des Ponts, France. E-mail: jean-francois.delmas@enpc.fr, clement.hardy@enpc.fr*
[3]*EDF R&D, Palaiseau, France. E-mail: anne.dutfoy@edf.fr*

Infrared spectroscopy is a widely used technology for nondestructive testing of materials. We propose a novel approach to automatically and simultaneously analyze a dataset of infrared spectra. They are modeled by linear combinations of peaks whose shape and position are parametrized. The observed data consist of linear combinations of the time-discretized peaks with an additive noise. In order to recover the peak parameters, common to all the dataset, and the associated amplitudes, which are specific to each spectrum, we formulate a penalized non-linear optimization problem. In this context, the penalization ensures that the spectra are recovered using a sparse set of common peaks.

Due to the non-convex nature of the problem and the continuous nature of the parameters, a resolution via standard procedures is out of reach. Therefore, we propose an off-the-grid algorithm with alternating convex optimization updates (to estimate the amplitudes of the peaks) and non-convex steps (to estimate the location and the scale of the peaks). In practice, this gives satisfactory results and provides sparse solutions.

We also study the numerical performances of the algorithm on simulated data and on real infrared spectra. The latter come from polychloroprene rubbers used in a marine environment at different aging levels. Eventually, we use a clustering algorithm in order to identify the peaks corresponding to the chemical components involved in the aging process of this material.

Keywords: IR spectroscopy, Convex optimization, Non-convex optimization, Group-Lasso, Sparsity

1. Introduction

Infrared spectra measure the interaction of infrared radiations with the matter. They reveal the presence of chemical substances or functional groups in solid, liquid or gaseous forms. This information on the composition of the matter is essential to prevent failures of materials. Therefore, the use of spectroscopy has become widespread in the industry for nondestructive testing. When a large number of spectra are to be analyzed, an automatic procedure is required. In this paper, we propose a procedure to automatically identify anomalous aging processes of polychloroprene rubbers in contact with sea water. Principal component analysis or its variants such as the partial least square analysis are often performed on a large dataset of spectra but produce results that are difficult to analyze physically. The spectra have many peaks, each peak corresponding to the absorption of an infrared radiation by a chemical compound. Each peak is characterized by its width, its location and its amplitude. The larger the amplitude of the peak, the more concentrated the chemical compound in the material. Several physical phenomenons imply that these peaks have a shape and a width that depend on the chemical substance (Hollas (2004)). When it comes to complex materials, the analysis of a spectrum may require some expertise. It involves determining the location and the width of overlapping peaks that are difficult to distinguish. In this case, the use of a numerical method is necessary. We model the peaks using Gaussian functions (the use of Lorentz functions is also common) and then compute the parameters from the observed spectra. In order to estimate the parameters of the model, curve fitting algorithms are commonly used: it amounts to solve a non linear least square problem as in Aragoni et al. (1995) or Antonov and Nedeltcheva (2000). However, the optimization techniques involved for such ill-conditioned problems require an initialization close to the real values of the parameters of the model. A first guess on the location of peaks and the number of peaks in the model is usually necessary. It would be preferable to have automated procedures allowing to get rid of a prior knowledge of the studied material such as in Alsmeyer and Marquardt (2004) or Kriesten et al. (2008). In order to give a physical sense to each parameter, it is essential not to over-parametrize the model by adding too many peaks to fit the spectra, particularly in the presence of noise. For this reason, we introduce in this

Proceedings of the 31st European Safety and Reliability Conference
Edited by Bruno Castanier, Marko Cepin, David Bigaud, and Christophe Berenguer
Published by Research Publishing, Singapore.
ISBN: 978-981-18-2016-8; doi:10.3850/978-981-18-2016-8_732-cd

paper estimators that are solutions of a penalized optimization problem similar to the problem that can be found in Golbabaee and Poon (2020). The penalization favors sparse solutions. Our choice of penalization and the fact that our analysis is run simultaneously on spectra ensures that the spectra are recovered using a sparse set of common peaks.

This work uses recent advances in optimization and statistics to extract physically motivated features from a dataset of infrared spectra observed with an additive noise. Note that our approach does not rely on any prior information on the material being studied. Finally, let us stress that the method presented here can be extended to all peak-shaped models and in particular to numerous branches of spectroscopy.

2. Definitions and Notations

Let $d \geq 1$ be the dimension of the space $\Theta \subset \mathbb{R}^d$ of peak parameters. Let φ be a positive smooth function defined on $\Theta \times \mathbb{R}$ modeling the shape of peaks. We denote by $T \in \mathbb{N}$ the size of the discretization grid over a wavenumber interval. For a discretization scheme on the real line $(\sigma_j)_{1\leq j\leq T}$, we set $\Delta_T = (\sigma_T - \sigma_1)/T$. For f, g measurable functions defined on $\mathbb{R}$, we set:

$$\langle f, g\rangle_T = \Delta_T \sum_{j=1}^{T} f(\sigma_j)g(\sigma_j),$$

and also $\|f\|_T = \langle f, f\rangle_T^{1/2}$. We assume that $\|\varphi(\theta)\|_T$ is non zero for all $T \in \mathbb{N}$ and $\theta \in \Theta$. We also define the normalized function ϕ_T on Θ taking values in $\mathbb{R}^T$ by:

$$\phi_T(\theta) = \|\varphi(\theta)\|_T^{-1} \left(\varphi(\theta, \sigma_1) \cdots \varphi(\theta, \sigma_T)\right).$$

Let $K \in \mathbb{N}^*$. For $\vartheta = (\theta_1, \ldots, \theta_K) \in \Theta^K$, we define the function Φ_T on Θ^K taking its values in $\mathbb{R}^{K\times T}$ by $\Phi_T(\vartheta) = \left(\phi_T(\theta_1)^\top, \cdots, \phi_T(\theta_K)^\top\right)^\top$. When there is no risk of confusion, we write Φ and ϕ instead of Φ_T and ϕ_T, respectively.

3. The Model

In this paper, we model infrared spectra using linear combinations of parametric functions φ (called peaks) with an additive noise. The parametric functions can be Gaussian or Lorentz functions, as usually done in the literature, see Hollas (2004). The location and the width of a peak are specific to a chemical group whereas its amplitude which encodes the concentration of the group depends on the material. Here we lead our study with Gaussian peaks:

$$\begin{aligned} \varphi\colon \Theta \subset \mathbb{R}^2 &\to L^2(\mathbb{R}) \\ (\mu, \nu) &\mapsto \mathrm{e}^{-\frac{(\cdot-\mu)^2}{2\nu^2}}. \end{aligned} \qquad \text{(Gauss)}$$

Consider a data set of n spectra $(y_i)_{1\leq i\leq n}$ disctretized on T wavenumbers $(\sigma_j)_{1\leq j\leq T}$, then write for all $1 \leq i \leq n, 1 \leq j \leq T$:

$$y_i(\sigma_j) = \sum_{k=1}^{K} B^\star_{i,k} \frac{\varphi(\theta^\star_k, \sigma_j)}{\|\varphi(\theta^\star_k)\|_T} + W_{ij}. \qquad (1)$$

In model (1), $(W_{ij})_{1\leq i\leq n, 1\leq j\leq T}$ denote the random variables modeling the noise, assumed to be independent Gaussian variables with zero-mean and variance s^2. The row vectors $(B^\star_{i,\cdot})_{1\leq i\leq n} \in \mathbb{R}_+^K$ of the matrix $B^\star$ have their k^{th}coordinate encoding the amplitude of the k^{th} peak involved in the linear combination. The positivity of their entries is physically motivated by the fact that spectra can only take positive values. The peaks are shared by all the spectra in the dataset but their amplitudes are specific to each spectrum. Note that each spectrum individually may have only a few peaks with non zero amplitudes. We denote by K an upper bound of the number of peaks which can be arbitrarily large. Since the family of Gaussian functions $(\varphi(\theta))_{\theta\in\Theta}$ is linearly independent, the spectra decomposition is unique.

The model (1) can be written in matrix form:

$$Y = B^\star \Phi(\vartheta^\star) + W,$$

where $Y \in \mathbb{R}^{n\times T}, Y_{ij} = y_i(\sigma_j)$, $W \in \mathbb{R}^{n\times T}$, $B^\star \in \mathbb{R}_+^{n\times K}$, $\vartheta^\star = (\theta^\star_1, \ldots, \theta^\star_K)$. One can notice that applying the same permutation on the columns of $B^\star$ and the coordinates of $\vartheta^\star$ gives the same model. It amounts to change the order of the peaks in the linear combinations of (1). The matrix $B^\star$ and the K-uplet $\vartheta^\star$ are defined up to such a joint permutation.

The spectra are expected to be decomposed in a small number of active peaks, that is why the matrix $B^\star$ is sparse and have numerous zero entries. Moreover, they are expected to have similarities so that only a few peaks are used to model the whole dataset. Hence, the matrix $B^\star$ have many columns set to zero. We denote by $S^\star$ the indices of the non zero columns of $B^\star$ which feature the active peaks in the dataset:

$$S^\star = \{k, \text{ there exists } 1 \leq i \leq n, B^\star_{ik} \neq 0\}.$$

Thus, the peaks $\varphi(\theta^\star_k)$ whose index k does not belong to the set $S^\star$ play no role in the model. We denote by $\vartheta^\star_{S^\star}$ the restriction of $\vartheta^\star$ to coordinates whose indices belong to $S^\star$.

4. Optimization Problem

In this section, we retrieve the non linear parameters $\vartheta^\star_{S^\star}$ (encoding the active peak), as well as the linear parameters $B^\star$ (encoding the amplitudes of peaks) that fully describe the model. We formulate

a program similar to the group-Lasso problem for sparse linear models introduced in Yuan and Lin (2006) and discussed in many papers since then (Lounici et al. (2011), Obozinski et al. (2011)). We generalize it to a larger range of sparse models in the same way as Boyer et al. (2017) and more recently as Golbabaee and Poon (2020). This leads to a non-linear least square problem with a penalization term weighted by a real parameter $\lambda > 0$:

$$\min_{\substack{B \in \mathbb{R}_+^{n\times K} \\ \vartheta \in \Theta_{K,T}(h)}} \frac{1}{nT} \|Y - B\Phi(\vartheta)\|_2^2 + \lambda \|B\|_{1,2}, \quad (2)$$

where,

- $\|\cdot\|_2$ is the usual Euclidean norm and $\|B\|_{1,2} = \sum_{k=1}^{K} \|B_{\cdot,k}\|_2$ is the mixed (1,2)-norm,
- $\Theta_{K,T}(h) \subset \Theta^K$ with $h > 0$, is the set of parameters $\vartheta = (\theta_1, \cdots, \theta_K) \in \Theta^K$ such that for all $1 \leq \ell, k \leq K, \ell \neq k$:

$$\mathcal{K}_T(\theta_\ell, \theta_k) := \frac{|\langle \varphi(\theta_\ell), \varphi(\theta_k)\rangle_T|}{\|\varphi(\theta_\ell)\|_T \, \|\varphi(\theta_k)\|_T} < h.$$

Let $(\hat{B}(\lambda), \hat{\vartheta}(\lambda))$ be solution of the problem (2) (or simply $(\hat{B}, \hat{\vartheta})$ when there is no ambiguity). We denote by $\hat{\vartheta}_{\hat{S}}$ the restriction of $\hat{\vartheta}$ to coordinates whose indices belong to:

$$\hat{S} = \{k : \text{ there exists } 1 \leq i \leq n, \hat{B}_{ik} \neq 0\}.$$

The set $\hat{S}$ gathers the indices of the active peaks used to fit the spectra. The penalization used in model (2) promotes group sparsity in the sense that it favors a matrix B that has columns with zero entries. This leads to solutions that use fewer peaks while correctly approximating the data. We refer to Obozinski et al. (2011) and Lounici et al. (2011) to better understand the penalization procedure in the case where the parameters $\vartheta^\star$ are known (but $S^\star$ is unknown). We also enforce the positivity of the linear parameters with constraints on the entries of the matrix $\hat{B}$.

The set $\Theta_{K,T}(h)$ introduced in this paper corresponds to a separation criterion on the peaks used to fit the data, the separation being measured by h. Provided that h is small enough, the matrix $\Phi(\hat{\vartheta})\Phi(\hat{\vartheta})^\top$ is of full rank. This is for example the case if $h < 1/(K-1)$, thanks to Gershgorin's theorem. This implies then that $\hat{B}$ is the unique solution of the problem:

$$\min_{B \in \mathbb{R}_+^{n\times K}} \frac{1}{nT} \left\| Y - B\Phi(\hat{\vartheta}) \right\|_2^2 + \lambda \|B\|_{1,2},$$

which amounts to minimize a strictly convex function over a convex set. The value chosen for h is a compromise: it must be large enough to estimate overlapping peaks in the dataset but sufficiently small to make the model identifiable in terms of the linear parameters. The identifiability for linear parameters is crucial to give them a physical sense.

5. Algorithm

5.1. *Presentation of the algorithm*

The resolution of the optimization problem (2) is not an easy task at first glance since the optimization problem is non-convex. Indeed, the peaks are not even convex in terms of their parameters. A brutal gradient-descent would be hopeless without a very good initialization. In this paper, we want to proceed without any prior knowledge on the parameters to be estimated. It might be tempting to use a grid on the space of non-linear parameters describing the peaks and use sparse methods to retrieve the amplitudes as suggested in Tang et al. (2013). But, the approximation of the spectra would depend on the chosen grid. Typically, it would be impossible to recover exactly the parameters of a peak without an infinitely thin grid. That is why an off-the-grid algorithm which does not discretize the parameter space must be preferred. Thus, it will be possible to recover exactly the parameters of the peaks provided their overlap is low (characterized by the parameter h). Recent progress in optimization have shown the efficiency of off-the-grid procedures such as the sliding Franck-Wolfe iterations (see Denoyelle et al. (2019)) or the alternating descent conditional gradient method (see Boyd et al. (2017)). Both algorithms are based on the addition of a new peak at each iteration to approximate one spectrum. During an iteration, a new peak is placed, then all the parameters are re-estimated with an improved initialization. The work of Golbabaee and Poon (2020) extended the Franck-Wolfe algorithm to fit several spectra simultaneously. We propose a variant of the Sliding Franck-Wolfe algorithm, see Algorithm 1 below, that separates the optimization of linear and non-linear parameters and which merges peaks that are highly overlapping. This allows the use of classical algorithms to solve a standard group-Lasso problem for linear parameters. Hence, this approach takes advantage of the fact that linear parameters are often more numerous than non linear parameters ($n \times K$ v.s $d \times K$) and always much easier to compute.

Let us write for any matrix $B \in \mathbb{R}^{n\times m}$ and $\vartheta \in \Theta^m$,

$$\mathcal{F}_{\lambda,\varphi}(B, \vartheta) = \frac{1}{nT} \|Y - B\Phi(\vartheta)\|_2^2 + \lambda \|B\|_{1,2}.$$

Remark 5.1. In the Sliding Franck-Wolfe algorithm as introduced in Denoyelle et al. (2019),

Algorithm 1:

Data: Y
Input: φ, λ, h
Output: ϑ, B
Initialize: $i := 0$, $R^{(0)} := Y$, $\vartheta^{(0)} := \emptyset$
while $i < K$ **do**

$\theta^{(i+\frac{1}{2})} \in \underset{\theta \in \Theta}{\operatorname{argmax}} \left\| R^{(i)} \phi(\theta)^\top \right\|_2^2$

$\vartheta^{(i+\frac{1}{2})} = \left(\vartheta^{(i)}, \theta^{(i+\frac{1}{2})} \right)$ `// Adding new peak`

$B^{(i+\frac{1}{2})} \in \underset{B \in \mathbb{R}_+^{n \times (i+1)}}{\operatorname{argmin}} \mathcal{F}_{\lambda,\varphi}(B, \vartheta^{(i+\frac{1}{2})})$

`// Linear step`

$\vartheta^{(i+1)} \in \underset{\vartheta \in \Theta^{i+1}}{\operatorname{argmin}} \mathcal{F}_{\lambda,\varphi}(B^{(i+\frac{1}{2})}, \vartheta)$
initialized in $\vartheta^{(i+\frac{1}{2})}$
`// Non-linear step`

Merging routine $(\vartheta^{(i+1)}, h)$
`// Merging overlapping peaks and adding peaks with parameters chosen at random`

$B^{(i+1)} \in \underset{B \in \mathbb{R}_+^{n \times (i+1)}}{\operatorname{argmin}} \mathcal{F}_{\lambda,\varphi}(B, \vartheta^{(i+1)})$

`// Re-estimation of linear parameters`

$R^{(i+1)} = Y - B^{(i+1)} \Phi(\vartheta^{(i+1)})$
$i = i + 1$
end

Algorithm 2: Merging routine

Input: $(\theta_1, \ldots, \theta_m), h$
Output: $(\theta_1, \ldots, \theta_m)$
while $(\theta_1, \ldots, \theta_m) \notin \Theta_{m,T}(h)$ **do**
  for $1 \leq \ell < k \leq m$ **do**
    if $\mathcal{K}_T(\theta_\ell, \theta_k) > h$ **then**
      θ_k is chosen at random in the parameter space
    end
  end
end

the peaks are never merged and therefore the re-estimation of linear and non-linear parameters in Algorithm 1 can be done simultaneously. Splitting into two steps avoids a gradient descent on all the parameters simultaneously.

5.2. *Implementation details*

We used a L-BFGS-B algorithm for the linear and the non linear steps. This algorithm allows the addition of constraints. Typically, peaks that are too thin to appear between two discretization points or wide peaks covering the whole range of observation should not be taken into account in the optimization. One can take $2\nu > \Delta_T$ so that a peak has a significant contribution on the discretization points. As for an upper bound on ν, one can take for the Gaussian model $6\nu < \sigma_T - \sigma_1$ so that a Gaussian function at the center of the observation range puts at least 99% of its mass between σ_1 and σ_T. It is also legitimate to require that the location parameter μ belongs to the range of observations *i.e.*: $\sigma_1 \leq \mu \leq \sigma_T$.

Without any prior information on the overlapping of the peaks, it may be necessary to re-run the algorithm and decrease h until one gets $\Phi(\hat{\vartheta})$ of full rank.

6. Numerical Applications

6.1. *Simultated data*

We tested the Algorithm 1 on noisy spectra composed of at most 15 Gaussian peaks. We generated a set of $n = 10$ spectra within the range $\sigma_{\min} = 0$ to $\sigma_{\max} = 20$. The parameters location μ and scale ν of the 15 Gaussian peaks were chosen at random according to a uniform distribution on the parameter space ($5 \leq \mu \leq 15$ and $10 \cdot \Delta_T \leq \nu \leq \frac{T \cdot \Delta_T}{6}$). We considered a Gaussian noise on the spectra by adding at each point of the discretization independent and identically distributed Gaussian random variables of mean 0 and standard deviation $s \in \{0, 0.01, 0.1, 0.5\}$ (see Figure 1). In order to show the consistency of the method, we computed for the different values of s, the mean square error between the data reconstructed with the estimated parameters and the data without noise,

$$MSE^* = \frac{1}{nT} \| B^\star \Phi(\vartheta^\star) - \hat{B} \Phi(\hat{\vartheta}) \|_2^2.$$

The estimated parameter $(\hat{B}, \hat{\vartheta})$ depends on the penalization parameter λ taken in the optimization problem (2). We took for λ the orders of magnitude that lead to the optimal convergence rates in the case of linear models as shown in Lounici et al. (2011) ($\lambda \sim s/\sqrt{nT}$) and we took $\lambda = 0.01/\sqrt{nT}$ for $s = 0$. The values of MSE^* from Figure 2 show that we managed to reconstruct almost exaclty the spectra in less than

100 iterations of the algorithm when $s \leq 0.01$.

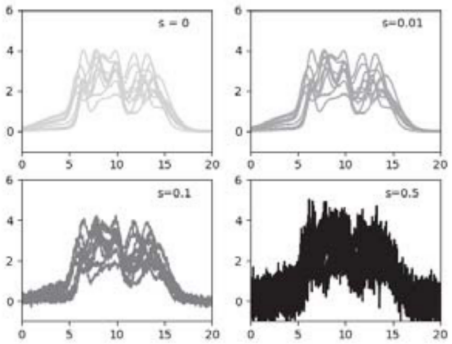

Fig. 1. Representation of the simulated data with different noise levels ($s \in \{0, 0.01, 01, 0.5\}$).

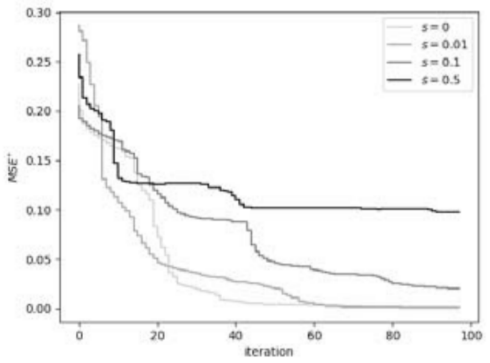

Fig. 2. Evolution of MSE^* over the iterations of the algorithm for different noise levels.

6.2. *Aging of polychloroprene rubbers*

6.2.1. *Presentation of the dataset*

The data used in our study were obtained from spectroscopic analysis of samples of polychloroprene rubbers, one side of which was in contact with seawater and the other was glued to steel. The device to obtain the spectra is a Fourier-transform infrared spectrometer in Attenuated Total Reflectance (ATR) mode. The spectra, visualized in a graph of infrared light absorbance on the vertical axis vs. wavenumbers on the horizontal axis, have to be normalized for the quantitative analysis of peak amplitudes. In addition, a preprocessing is also performed to remove the baselines present on the spectra. The multiplicative normalization is specific to each spectra and such that its peak amplitude for the $C - Cl$ bond situated at 825 cm^{-1} is equal to 1. The $C - Cl$ bond was chosen for the normalization of the spectra because of its stability with respect to the aging process, see Le Gac et al. (2012). It is then possible to compare the peak amplitudes between the spectra in the dataset (see Figure 3).

The 72 spectra composing the dataset are dis-

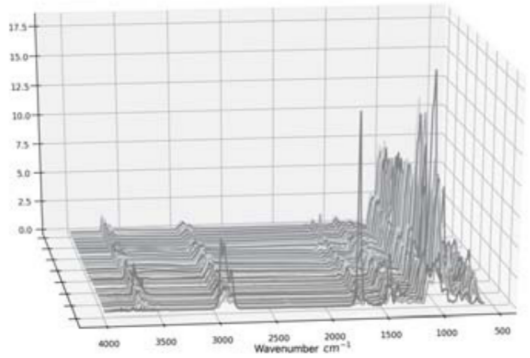

Fig. 3. Representation of all the infrared spectra of polychloroprene samples after normalization and removal of baselines.

cretized between 4000 cm^{-1} and 600 cm^{-1} with measures every 2 cm^{-1}. We focus our analysis on the area between 2000 cm^{-1} and 600 cm^{-1} where are the biggest dissimilarities between the spectra.

6.2.2. *Aging properties of polychloroprene rubbers*

Polychroroprene is often used in marine structure to prevent corrosion. Some works on polychloroprene rubbers in marine environments have brought out some physical phenomenons that occur with aging (see Le Gac et al. (2012), Tchalla et al. (2017)). The sea water diffuses into the material until it is saturated. During the process, several reactions might appear. In Le Gac et al. (2012), the authors have pointed out the hydrolysis of silica fillers. It consists in a formation of silanol from the silica fillers. This reaction is reflected in the spectra by a decrease of the $1160 - 1082$ peaks (attributed to the $Si - O$ bond of the silica filler) and a new peak located around 1009 cm^{-1}(attributed to $Si - OH$) that rises according to the aging duration. In addition, they showed that a carbonyl formation can occur due to an oxidation reaction. This can be seen in the spectra by the appearance of a new peak at 1731 cm^{-1}. In Tchalla (2017), peaks in spectra from polychloroprene samples with aging conditions similar to those in our study are attributed to their corresponding chemical bond (see Table 1).

Table 1. Table of the location of peaks and their corresponding bonds for the polychloroprene samples. The values are taken from Tchalla (2017).

Wavenumbers (cm^{-1})	Peak assignment
3690-3400-3364	-OH
3200-3014	
2952-2920-2850	$\nu - CH_2, CH_3$ Aliphatic
1731	$\nu - C = O$
1647	$\nu - C = C$ of $HC = CH_2$
1540	$\nu - C = C$ of $R - CR = CH - R$ and $\delta - CH_2$ Aliphatic
1419	$\delta - CH_2, \delta - CH$ Aliphatic
1160-1082	$\nu - Si - O$ (SiO_2)
1009-909	$\nu - Si - O$ $(Si - OH)$
825	$C - Cl$
664	CH Aromatic

6.2.3. *Estimation of linear and non linear parameters for the peak-shaped model*

In the optimization problem (2) of Section 4, the penalization parameter λ must be tuned. Intuitively, choosing a large value for λ will make the term of penalization in (2) preponderant and set a lot of entries of the matrix $\hat{B}$ to zero. In this case, one expects the solutions to underestimate the number of peaks in the model. On the contrary, a small value for λ will set very few entries of the matrix $\hat{B}$ to zero and will lead to overestimate the number of peaks in the model. There is no easy way to choose the penalization parameter λ. To achieve a compromise between the number of peaks used and the quality of the spectra approximation,we ran the algorithm on the set of polychloroprene spectra for different values of the tuning parameter λ. It appeared that for the Gaussian model, around $\lambda \approx 5 \cdot 10^{-5}$, the unpenalized mean square error $\mathcal{F}_{0,\varphi_G}(\hat{B}(\lambda), \hat{\vartheta}(\lambda))$ as well as the penalized mean square error $\mathcal{F}_{\lambda,\varphi_G}(\hat{B}(\lambda), \hat{\vartheta}(\lambda))$ increase drastically (see Figure 4). From this point, the number of peaks used to fit the data drops (see Figure 5). Hence, a reasonable choice for λ, is under this critical point. We ran the Algorithm 1 with the Gaussian model ($\varphi := \varphi_G$) for $\lambda = 3 \cdot 10^{-5}$ and $h = 0.9$. We imposed that the location parameter belongs to the range of obervations $[600cm^{-1}, 2000cm^{-1}]$, and that the width parameter ν belongs to $[\frac{\Delta_T}{2} \approx 1, \frac{T\Delta_T}{6} \approx 233]$. Finally, we obtained 66 active peaks for the whole dataset in the range $[600cm^{-1}, 2000cm^{-1}]$ after 100 iterations of the algorithm.

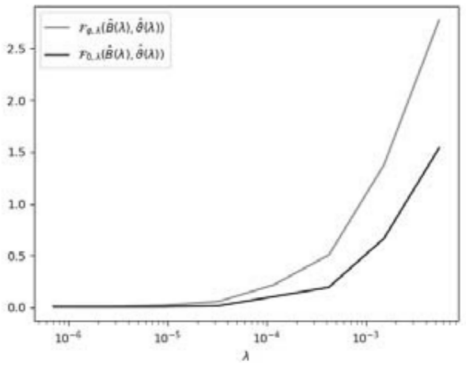

Fig. 4. Mean square error $\mathcal{F}_{0,\varphi}(\hat{B}(\lambda), \hat{\vartheta}(\lambda))$ and penalized mean square error $\mathcal{F}_{\lambda,\varphi}(\hat{B}(\lambda), \hat{\vartheta}(\lambda))$ seen as functions of λ.

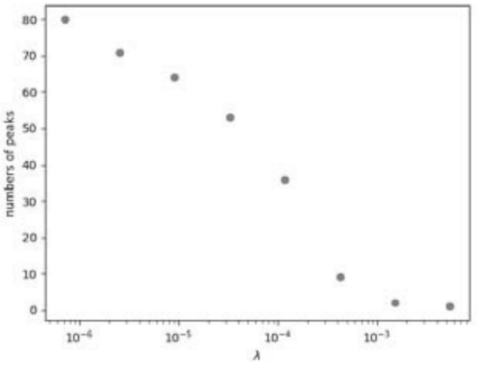

Fig. 5. Number of peaks found by the algorithm to fit the spectra of polychloroprene samples as a function of the tuning parameter λ.

6.2.4. *Boxplots for main peak amplitudes*

To understand the distribution of the peak amplitudes within the dataset, we represented them on Figure 6 with boxplots. The locations of peaks on the x-axis as well as the amplitude values in the boxes, correspond to those estimated by Algorithm 1. For the sake of readability, we have represented only the ten most significant peaks (those with the biggest sum of squared amplitudes). First, one can notice that Algorithm 1 retrieves peaks close to those referenced in Table 1: the carbonyl peaks at 1732 cm^{-1} (vs 1731 cm^{-1} in the Table 1) as well as the silica peaks at $1162 - 1089$ cm^{-1} (versus $1160 - 1082$ cm^{-1} in Table 1) and the silanol peaks at $1012 - 917 - 902$ cm^{-1} (versus $1009 - 909$ cm^{-1} in Table 1). Secondly, it appears that the silanol peak brings the most dissimilarity among the spectra.

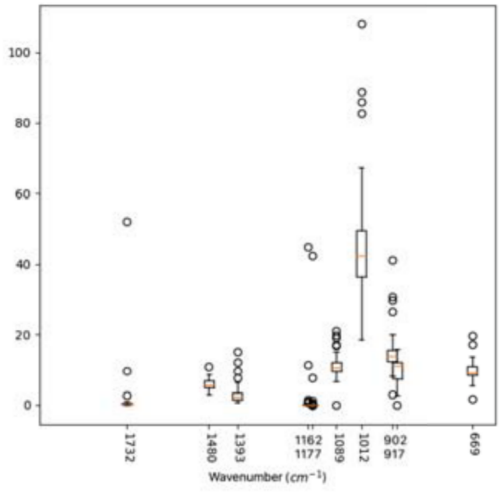

Fig. 6. Boxplot for the amplitudes of the 10 most significant peaks for the 72 polychloroprene spectra in the dataset.

6.2.5. *Clustering on peak amplitudes*

In order to bring out different levels of aging among the spectra, a clustering algorithm such as a k-means algorithm whose inputs are the vectors of estimated peak amplitudes $\hat{B}_{i,.} \in \mathbb{R}^K, 1 \leq i \leq n$ can be used. The k-means algorithm aims to partition the n observations vectors into M sets $\mathcal{A} = \{\mathcal{A}_1, \cdots, \mathcal{A}_M\}$ so as to minimize the within-cluster sum of squares. It amounts to solve

$$\min_{\mathcal{A}} \sum_{\ell=1}^{M} \sum_{i \in A_\ell} \left\| \hat{B}_{i,.} - \beta_\ell \right\|_2^2$$

where the vectors β_ℓ are the centroids of the sets $(\mathcal{A}_\ell)_{1 \leq l \leq M}$. Let us write $(\hat{\mathcal{A}}_1, \cdots, \hat{\mathcal{A}}_M)$ the partition returned by a k-means algorithm and $(\hat{\beta}_1, ...\hat{\beta}_M)$ the associated centroids.

The number of clusters is an input of the algorithm. Therefore, one of the first issue to address is related to the number of clusters used. A compromise must be found between gathering the data in a few groups and not having too much dissimilarity within the group. To tackle this issue, solving the k-means problem for different values of M can be useful. We plotted in Figure 7 the value I(M), as a function of M, of the sum of squared distances of the observations to their closest cluster centroid:

$$I(M) = \sum_{\ell=1}^{M} \sum_{i \in \hat{A}_\ell} \left\| \hat{B}_{i,.} - \hat{\beta}_\ell \right\|_2^2.$$

By taking $M = 4$ where the curve makes an elbow, we separate the data into a number of clusters small enough to be informative while

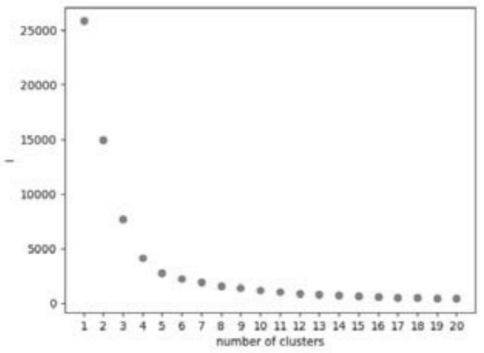

Fig. 7. Sum of squared distances of observations to their closest cluster centroid with respect to the numbers of clusters.

having drastically reduced the sum of squared distances between observations and their associated centroid (see Figure 8). Let us observe that among

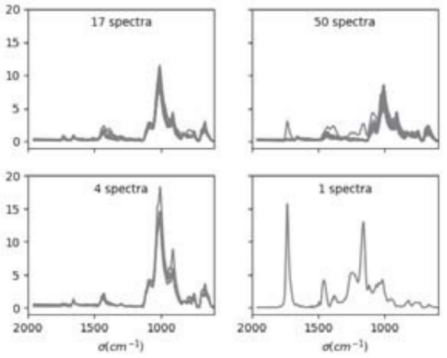

Fig. 8. Representation of the polychloroprene spectra within their cluster after running a k-means algorithm on the row vectors $\hat{B}_{i,.} \in \mathbb{R}^K, 1 \leq i \leq n$.

the four clusters, one gathers about 70% of the data (top right hand graphic in Figure 8). One can also notice that the clusters at the top in Figure 8), gathering more than 90%, are characterized by lower amplitudes for the silanol peak at 1009 cm^{-1}. The spectrum that forms a single-point-cluster presents strong carbonyl peaks centered around 1731 cm^{-1}. Hence, we managed to isolate a spectra that presents a really high level of carbonyl and separate the others with respect to the amplitudes of the silanol and silica peaks without any prior information on the material. Let us recall that the rise of the silanol and carbonyl peaks correspond to reactions involved in the aging process. We also considered the clusters based on the amplitudes of the peaks corresponding only to silica, silanol and carbonyl by selecting the

peaks returned by the algorithm that are located at less than 10 cm^{-1} of the positions referenced in the Table 1. The clusters obtained correspond exactly to those from Figure 8. Therefore, one can conclude that the main differences between the spectra are due to the peaks of carbonyl, silanol and silica which were identified in the literature as involved in the aging process of polychloroprene rubbers in a marine environment. The two clusters at the bottom of Figure 8 gather spectra with chemical characteristics of higher aging levels and represent less than 10% of the data.

7. Conclusion

This paper estimates an arbitrary number of infrared spectra simultaneously without any prior information. The spectra are modeled under the physical constraints by linear combinations of peaks and each peak belongs to a nonlinear parametric family of functions (e.g. Gaussian). The estimation consists in a generalization to nonlinear models of the group-Lasso optimization problem. This formulation allows to limit the number of peaks used to fit the data. A numerical method is proposed with an off-the-grid scheme. The limited resolution problems, intrinsic to the use of a grid on the parameter space, are thus avoided. The method is numerically consistent in the presence of noise and favors sparse solutions. Although the problem is nonlinear, the method works without any special care for the initialization. Moreover, the computation time behaves well with a large number of spectra as long as the number of peaks to fit the data does not increase drastically. Theoretical guarantees for the consistency of the parameter estimators will be the subject of a further study. We apply this approach to real data of polychloroprene rubber spectra, and recover the main peaks associated with its chemical components and identify by clustering those involved in its aging process. The locations of the peaks found by the algorithm are consistent with those established by previous work in the field of chemistry. Next, we plan to develop procedures to detect anomalous spectra and to quantify the uncertainty of an estimated spectrum using confidence bands.

Acknowledgement

This work was partially supported by the ANRT grant N°2019/1260. The authors are grateful to the members of the Corrosion and Electrochemistry team at EDF for the construction of the IR database.

References

Alsmeyer, F. and W. Marquardt (2004). Automatic generation of peak-shaped models. *Applied spectroscopy 58*(8), 986–994.

Antonov, L. and D. Nedeltcheva (2000). Resolution of overlapping uv–vis absorption bands and quantitative analysis. *Chemical Society Reviews 29*(3), 217–227.

Aragoni, M. C., M. Arca, G. Crisponi, and V. M. Nurchi (1995). Simultaneous decomposition of several spectra into the constituent Gaussian peaks. *Analytica chimica acta 316*(2), 195–204.

Boyd, N., G. Schiebinger, and B. Recht (2017). The alternating descent conditional gradient method for sparse inverse problems. *SIAM Journal on Optimization 27*(2), 616–639.

Boyer, C., Y. De Castro, and J. Salmon (2017). Adapting to unknown noise level in sparse deconvolution. *Information and Inference: A Journal of the IMA 6*(3), 310–348.

Denoyelle, Q., V. Duval, G. Peyré, and E. Soubies (2019). The sliding Frank-Wolfe algorithm and its application to super-resolution microscopy. *Inverse Problems 36*(1), 014001.

Golbabaee, M. and C. Poon (2020). An off-the-grid approach to multi-compartment magnetic resonance fingerprinting. *arXiv preprint arXiv:2011.11193*.

Hollas, J. M. (2004). *Modern spectroscopy*. John Wiley & Sons.

Kriesten, E., F. Alsmeyer, A. Bardow, and W. Marquardt (2008). Fully automated indirect hard modeling of mixture spectra. *Chemometrics and Intelligent Laboratory Systems 91*(2), 181–193.

Le Gac, P.-Y., V. Le Saux, M. Paris, and Y. Marco (2012). Ageing mechanism and mechanical degradation behaviour of polychloroprene rubber in a marine environment: Comparison of accelerated ageing and long term exposure. *Polymer degradation and stability 97*(3), 288–296.

Lounici, K., M. Pontil, S. Van De Geer, and A. B. Tsybakov (2011). Oracle inequalities and optimal inference under group sparsity. *Annals of statistics 39*(4), 2164–2204.

Obozinski, G., M. J. Wainwright, and M. I. Jordan (2011). Support union recovery in high-dimensional multivariate regression. *The Annals of Statistics 39*(1), 1–47.

Tang, G., B. N. Bhaskar, and B. Recht (2013). Sparse recovery over continuous dictionaries-just discretize. In *2013 Asilomar Conference on Signals, Systems and Computers*, pp. 1043–1047. IEEE.

Tchalla, S. T., P.-Y. Le Gac, R. Maurin, and R. Creac'Hcadec (2017). Polychloroprene behaviour in a marine environment: Role of silica fillers. *Polymer Degradation and Stability 139*, 28–37.

Tchalla, T. S. (2017). *Durabilité d'assemblages métal/élastomère en milieu marin*. Ph. D. thesis, Université de Bretagne occidentale - Brest, France.

Yuan, M. and Y. Lin (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68*(1), 49–67.