



Abstract — Nonlinearity in Raman spectral mixtures caused by wavenumber shifts, has been investigated in this paper. The spectral shifts are mainly caused by the existence of multiple chemicals in the mixtures, with complex molecular interactions, which can change the spectral features of each constituent. While such non-linear behaviour may be negligible in some mixtures, it may lead to incorrect identification of chemicals in some instances. We investigate some real spectra and demonstrate the nature of such nonlinearity in Raman spectra. We then mathematically formulate such spectral behaviour and present an approach to compensate the nonlinearity artifacts. The nonlinearity has been modelled as a smooth transition in a parametric space, which can be locally modelled using first order approximation. Such a first order approximation can be translated to some augmented spectral libraries to be used with a linear generative model. A convex sparse approximation program, with nonlinearity considerations, has finally been introduced to decompose the spectral mixtures. Such decomposition has been used for chemical fingerprinting and quantification. The effect of new approach has been demonstrated with some real and synthetic spectra.

Raman Spectral Mixture Analysis

Motivations:

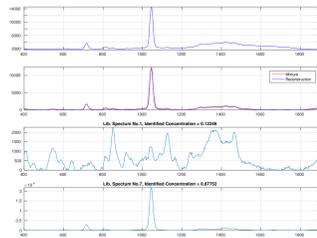
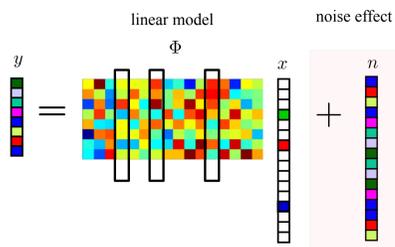
1. Raman spectra are feature-rich signatures for unknown material detection.
2. Detection of the trace of chemicals and chemical mixtures from a single snapshot is difficult using current similarity based methods.
3. Most techniques are computationally intensive and not suitable for a real-time implementation.

Aim: Efficiently decompose the spectra, fingerprint and quantify the mixtures.

How:

1. **Baseline correction** of the spectra to remove the Florescent background artefact.
2. **Nonnegative Sparse Decomposition** of spectra using a reference library of the pure spectra.
3. **Fingerprinting and Quantification** based on the nonnegative contribution, i.e. coefficients.

Our approach [Yaghoobi16]: using a novel greedy sparse approximation method called the fast non-negative orthogonal matching pursuit.



Wavenumber Shifts and Nonlinear Mixtures

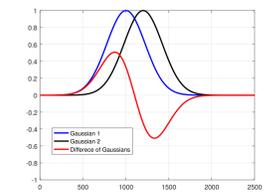
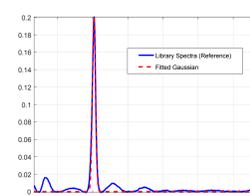
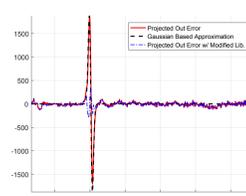
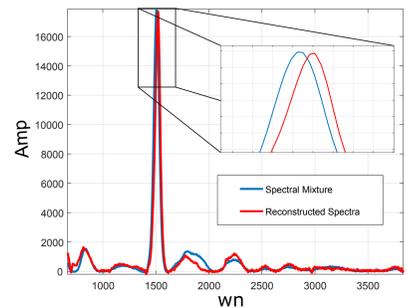
Observation:

- The peaks of spectra in the mixtures can shift, due to molecular interactions between constituents.
- The amount of shift is a **nonlinear function** of components concentrations.

Issues:

- **Misdetection:** When the spectral peaks, which have most energy of the spectra, are shifted, the similarity measure significantly drops. → Algorithm will then misdetect a chemical and possibly **select another incorrect component** or count it as the **measurement error**.
- **Misquantification:** the correct quantification highly depends on the peaks on the original positions. The quantification error is introduced as the result of local peak shifts.

Proposed solution: Approximation of the shift function with the first order Taylor's series.



Robust Raman Nonnegative Sparse Decomposition with Taylor's First Order Modelling

Locally Perturbed Spectra and First Order Approximations:

- Let the nonlinear spectral generative model be as follows:

$$y = \phi(\mathbf{M}, \alpha) + \omega = \sum_{\mu_i \in \mathcal{M}} \alpha_i \mu_i + \omega$$

where μ_i is a signal in \mathcal{M}_i and $\mathcal{M} = \cup_{i=1}^N \mathcal{M}_i$.

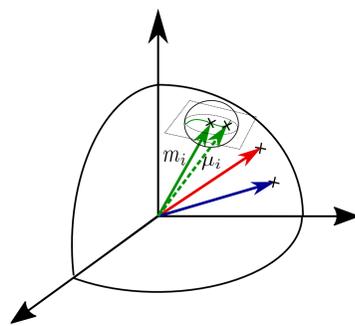
- Let $f_i(\Delta) : \mathbb{R}^r \rightarrow \mathcal{M}_i$ be the function mapping a Euclidean space dimension r to \mathcal{M}_i , with the following properties for each $i, 1 \leq i \leq N$,

$$\mu_i = f_i(\Delta), \quad \mathbf{m}_i = f_i(0).$$

- Taylor's first order approximation of $f_i(\Delta)$ can be an estimate for μ_i , i.e. $\mu_i \approx f_i(\Delta_0) + (\Delta - \Delta_0) \frac{\partial}{\partial \Delta} f_i(\Delta_0)$.

Gaussian Peak Modelling and Approximation around 0 and $\Delta_0/2$:

- A spectrum functional with multiple shifting peaks can be written as, $f_i(\Delta) = \bar{\mathbf{m}}_i + \sum_{\tau \in T_i} g_\tau(\nu - \Delta_\tau)$, $\bar{\mathbf{m}}_i \rightarrow$ residual spectrum, i.e. spectrum without peaks, $T_i \rightarrow$ peak locations and $\Delta = \{\Delta_\tau\}_{\tau \in T_i} \rightarrow$ all shifts.
- Each peak can be approximately modelled with a Gaussian kernel, $g_\tau(\nu) := \beta_\sigma \exp(-(\nu - \tau)^2 / 2\sigma_\tau^2)$
- Approximation around 0: $y \approx \sum_i \alpha_i \mathbf{m}_i - \sum_i \sum_{\tau \in T_i} \gamma_{i,\tau} \Delta_0 g'(\nu - \tau, \sigma_\tau) + \omega$
- Approximation around $\Delta_0/2$: $y \approx \sum_i \alpha_i \mathbf{m}_i - \sum_i \sum_{\tau \in T_i} \sum_{j \in \mathcal{J}} \gamma_{i,j,\tau} g'(\nu - \tau - \frac{j\delta}{2}, \sigma_\tau) + \omega$, where $\mathcal{D} = \{j\delta\}_{j \in \mathcal{J}}$ is the set of possible shifts.



Robust Sparse Decomposition With an Augmented Library

- Building up an augmented system of equations to include first order approximations,

$$y = \mathbf{M}\alpha + \mathbf{A}\gamma + \omega = \underbrace{\begin{bmatrix} \mathbf{M} & \mathbf{A} \end{bmatrix}}_{\mathbf{M}^+} \underbrace{\begin{bmatrix} \alpha \\ \gamma \end{bmatrix}}_{\alpha^+} + \omega$$

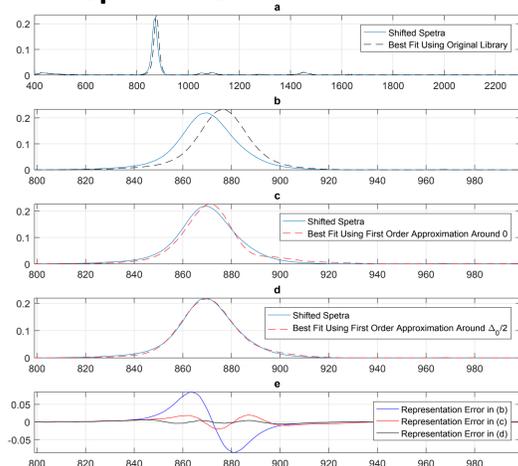
where $\mathbf{A} \in \mathbb{R}^{d \times M}$ is the matrix related to the discretised linear functional $g'(\nu - \tau - \frac{j\delta}{2}, \sigma_\delta)$, and $\alpha^+ \in \mathbb{R}_+^{N+M}$. \mathbf{M}^+ is the augmented library with some extra functions modelling the spectral shifts.

- Finding the non-negative sparse approximation of y to find α^+ ,

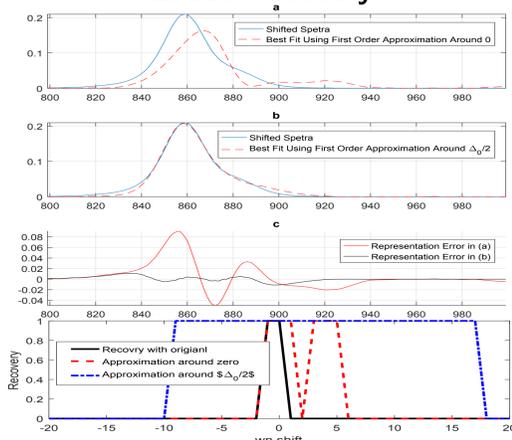
$$\alpha^{+*} = \operatorname{argmin}_{\alpha^+ \in \mathbb{R}_{\geq 0}} \|y - \mathbf{M}^+ \alpha^+\|_2^2 + \lambda \|\alpha^+\|_1,$$

This optimisation problem can be solved using scalable methods like ADMM and Nesterov's optimal first order methods.

Spectral Shifts for 7 Tabs



Spectral Shifts for 17 Tabs and Exact Recovery



Summary

- Raman spectral decomposition can be used to fingerprint and quantify the mixtures.
- The nonlinearity due to spectral shifts can cause misdetection and misquantification.
- A first order approximation of the spectral local shifts was used to model the artefacts.
- Some promising results show that the new model can compensate moderate shifts.
- Future Work:
 - Faster method to solve the optimisation problem, i.e. using greedy methods.
 - A comparison with robust sparse decomposition methods with neighborhood constraints.



Engineering and Physical Sciences Research Council