Maximum Likelihood Models for Peak Shifts
BMBF HYPERMATH Project Workshop, Heidelberg

Ulrike Mayer¹  Pia Heins²  Alexandra Koulouri²
Henryk Zaehle¹  Martin Burger²  Frank Wübbeling²

¹Saarland University
²University of Muenster

March 04, 2015
Outline

Introduction
   Raman Spectroscopy
   Application & Experimental Set-up

Modeling of Different Estimators
   Regression Model
   Estimators

Deconvolution of Sparse Spikes
   From the Estimator to the Variational Model
   Reconstruction of a Single Spike
   Numerical Deconvolution of Three δ-Spikes

Summary & Outlook
Reminder: Raman Spectroscopy

- Small portion of back-scattered light activates bond oscillation in molecules

- **Measure** scattered light → **Full spectrum** in range 624 to 1834 nm wave shift

- Typical **spectral fingerprint** for simple, pure materials is generated

- Complex structures (e.g. biological tissue) will generate **complex spectra**
**Figure:** Raw Raman spectrum containing phase shifts, baseline and high frequency noise
Mathematical Problems

• We do not have a model for Raman spectroscopy of biological material (i.e. a calibration model does not exist)

• Data is polluted by unknown background, which may vary from sample to sample
  ~→ No evaluation of absolute numbers possible

• Data may contain random shifts in wave length

• Data contains high frequency noise, which is filtered by the instrument (smoothing)
Mathematical Problems

- We do not have a model for Raman spectroscopy of biological material (i.e. a calibration model does not exist)

- Data is polluted by unknown background, which may vary from sample to sample
  - No evaluation of absolute numbers possible

- Data may contain random shifts in wave length

- Data contains high frequency noise, which is filtered by the instrument (smoothing)
Peak-Shifts

Peak shift can happen randomly for different measurements

Figure: Comparison of the data sets
Outline

Introduction
  Raman Spectroscopy
  Application & Experimental Set-up

Modeling of Different Estimators
  Regression Model
  Estimators

Deconvolution of Sparse Spikes
  From the Estimator to the Variational Model
  Reconstruction of a Single Spike
  Numerical Deconvolution of Three $\delta$-Spikes

Summary & Outlook
Application

Measured data:
Raman spectrum of seminal plasma

Goal:
Determination of the concentration of a certain protein called Hemolysin

⇒ Knowledge about a particular type of bacterial infection

Hemolysin is present in case there is an infection
Experimental Set-up

• Commercial hemolysin was diluted in a medium called RPMI

• RPMI including hemolysin was then diluted in seminal plasma

• Samples with different levels of hemolysin concentration

• Measurement of Raman spectra of these samples

• The same experiment was performed a second time using a new set of samples
Gaussian Peak-fitting

Idea:

Describe characteristic parts of the spectrum with a linear combination of Gaussians.

Advantages:

- Obtain information about shape and structure of the spectra
  ▸ Samples are better comparable

- Compression of the data possible
  ▸ less storage space needed
Gaussian Peak-fitting

Hly concentration 27.5 (High concentration)

Hly concentration 6.9 (low concentration)

Hly concentration very low (almost zero)
Task

- State inverse problem / regression model
- Express / approximate data as linear combination of Gaussians
- Model the noise and consider different estimators
- Formulate as variational problem
- Consider specific regularization
Outline

Introduction

Raman Spectroscopy
Application & Experimental Set-up

Modeling of Different Estimators

Regression Model
Estimators

Deconvolution of Sparse Spikes

From the Estimator to the Variational Model
Reconstruction of a Single Spike
Numerical Deconvolution of Three δ-Spikes

Summary & Outlook
Consider the regression model

\[ Y_{n,i}(\omega) = f(t_{n,i} + \eta_{n,i}(\omega); \theta) + \xi_{n,i}(\omega), \quad 1 \leq i \leq n, \omega \in \Omega, \]

where the parameterized regression function is defined by

\[ f(x; \theta) := \sum_{k=1}^{K} h_k e^{-\frac{(x-c_k)^2}{\nu}} \]

for some given \( \nu > 0 \).

**Notation**

- \( t_{n,1}, \ldots, t_{n,n} \) deterministic design points
- \( \theta = (c_1, h_1, \ldots, c_K, h_K) \) unknown true parameter
- \( \eta_{n,1}, \ldots, \eta_{n,n} \) and \( \xi_{n,1}, \ldots, \xi_{n,n} \) i.i.d. error random variables
Regression Model

Consider the regression model

\[ Y_{n,i}(\omega) = f(t_{n,i} + \eta_{n,i}(\omega); \theta) + \xi_{n,i}(\omega), \quad 1 \leq i \leq n, \ \omega \in \Omega, \]

where the parameterized regression function is defined by

\[ f(x; \theta) := \sum_{k=1}^{K} h_k e^{-\frac{(x-c_k)^2}{\nu}} \]

for some given \( \nu > 0 \).

In simulation study below:

- \( K = 1 \) (single peak)
- \( \theta = (c, h) \) unknown true parameter
- \( \eta_{n,i} \sim \mathcal{N}_0, s^2 \) for unknown \( s > 0 \)
- \( \xi_{n,i} \sim \mathcal{N}_0, \sigma^2 \) for unknown \( \sigma > 0 \)
Least-squares estimator

Consider the regression model

$$Y_{n,i}(\omega) = f(t_{n,i} + \eta_{n,i}(\omega); \theta) + \xi_{n,i}(\omega), \quad 1 \leq i \leq n, \ \omega \in \Omega,$$

where the parameterized regression function is defined by

$$f(x; \theta) := \sum_{k=1}^{K} h_k e^{-\frac{(x-c_k)^2}{\nu}}$$

for some given $\nu > 0$.

Least-squares estimator

The classical (deterministic) approach to estimate $\theta$ is

$$\hat{\theta}_{n}^{LS}(\omega) := \arg\min_{\theta} \sum_{i=1}^{n} (Y_{n,i}(\omega) - f(t_{n,i}; \theta))^2, \quad \omega \in \Omega.$$ 

**Question:** Good choice?
Consider the regression model

$$Y_{n,i}(\omega) = f(t_{n,i} + \eta_{n,i}(\omega); \theta) + \xi_{n,i}(\omega), \quad 1 \leq i \leq n, \ \omega \in \Omega,$$

where the parameterized regression function is defined by

$$f(x; \theta) := \sum_{k=1}^{K} h_k e^{-\frac{(x-c_k)^2}{\nu}} \text{ for some given } \nu > 0.$$

**Least-squares estimator**

The classical (deterministic) approach to estimate $\theta$ is

$$\hat{\theta}_{n}^{LS}(\omega) := \arg\min_{\theta} \sum_{i=1}^{n} \left( Y_{n,i}(\omega) - f(t_{n,i}; \theta) \right)^2, \quad \omega \in \Omega.$$

**Remark:** The noise $\eta$ is not taken into account explicitly.
Probabilistic Approach: Maximum likelihood method

The Likelihood function is defined by

\[ L_n(y_1, \ldots, y_n, t_{n,1}, \ldots, t_{n,n}; \theta, s, \sigma) := p(Y_{n,1}, \ldots, Y_{n,n};(\theta, s, \sigma)(y_1, \ldots, y_n), \]

where \( p(Y_{n,1}, \ldots, Y_{n,n};(\theta, s, \sigma) \) is the Lebesgue density of the probability distribution of the random vector \((Y_{n,1}, \ldots, Y_{n,n})\) under \((\theta, s, \sigma)\).

Maximum likelihood estimator

\[
(\hat{\theta}_n^{ML}, \hat{s}_n^{ML}, \hat{\sigma}_n^{ML}) = \arg\max_{(\theta, s, \sigma)} L_n(Y_{n,1}, \ldots, Y_{n,n}, t_{n,1}, \ldots, t_{n,n}; \theta, s, \sigma) 
\]
Probabilistic Approach: Maximum likelihood method

By the independence of $Y_{n,1}, \ldots, Y_{n,n}$, we obtain

$$L_n(y_1, \ldots, y_n, t_{n,1}, \ldots, t_{n,n}; \theta, s, \sigma) = \prod_{i=1}^{n} p_{Y_{n,i};(\theta,s,\sigma)}(y_i),$$

where $p_{Y_{n,i};(\theta,s,\sigma)}$ is the Lebesgue density of the probability distribution of $Y_{n,i}$ under $(\theta, s, \sigma)$. Moreover

$$p_{Y_{n,i};(\theta,s,\sigma)}(y) = p_{f}(t_{n,i}+\eta_{n,i};\theta) + \xi_{n,i}(y)$$

$$= \int p_{f}(t_{n,i}+\eta_{n,i};\theta) + \xi_{n,i}(y) |_{\eta_{n,i}=u} p_{\eta_{n,i}}(u) \, du$$

$$= \int p_{f}(t_{n,i}+u;\theta) + \xi_{n,i}(y) p_{\eta_{n,i}}(u) \, du$$

$$= \int \phi_{f}(t_{n,i}+u;\theta),\sigma^2(y) \phi_{0,s^2}(u) \, du.$$
Probabilistic Approach: Maximum likelihood method

By the independence of $Y_{n,1}, \ldots, Y_{n,n}$, we obtain

$$L_n(y_1, \ldots, y_n, t_{n,1}, \ldots, t_{n,n}; \theta, s, \sigma) = \prod_{i=1}^{n} p_{Y_{n,i};(\theta,s,\sigma)}(y_i),$$

where $p_{Y_{n,i};(\theta,s,\sigma)}$ is the Lebesgue density of the probability distribution of $Y_{n,i}$ under $(\theta, s, \sigma)$. Moreover

$$p_{Y_{n,i};(\theta,s,\sigma)}(y) = p_{f(t_{n,i}+\eta_{n,i};\theta)+\xi_{n,i}}(y)$$

$$= \int p_{f(t_{n,i}+\eta_{n,i};\theta)+\xi_{n,i}|\eta_{n,i}=u}(y) \ p_{\eta_{n,i}}(u) \ du$$

$$= \int p_{f(t_{n,i}+u;\theta)+\xi_{n,i}}(y) \ p_{\eta_{n,i}}(u) \ du$$

$$= \frac{1}{\sqrt{2\pi \sigma^2}} \int e^{-\frac{(y-f(t_{n,i}+u,\theta))^2}{2\sigma^2}} \ \phi_{0,s^2}(u) \ du.$$
Probabilistic Approach: Maximum likelihood method

Maximum likelihood estimator

\[
(\hat{\theta}_n^{ML}, \hat{s}_n^{ML}, \hat{\sigma}_n^{ML}) = \arg\max_{(\theta, s, \sigma)} \prod_{i=1}^{n} \frac{1}{\sqrt{2 \pi \sigma^2}} \int e^{-\frac{(Y_{n,i} - f(t_{n,i} + u, \theta))^2}{2\sigma^2}} \phi_{0,\sigma^2}(u) \, du
\]

\[
= \arg\max_{(\theta, s, \sigma)} \sum_{i=1}^{n} \log \left( \frac{1}{\sigma} \int e^{-\frac{(Y_{n,i} - f(t_{n,i} + u, \theta))^2}{2\sigma^2}} \phi_{0,\sigma^2}(u) \, du \right)
\]
Probabilistic Approach: Maximum likelihood method

Note that for $s = 0$ (so that $\phi_{0,s^2} = \delta_0$) we have

$$\hat{\theta}_{ML}^\text{n} = \arg\max_{\theta} \sum_{i=1}^{n} \log \left( \frac{1}{\sigma} \cdot e^{-\frac{(Y_n,i-f(t_n,i,\theta))^2}{2\sigma^2}} \right)$$

$$= \arg\min_{\theta} \sum_{i=1}^{n} (Y_n,i - f(t_n,i,\theta))^2$$

$$= \hat{\theta}_{LS}^\text{n}$$

for every $\sigma > 0$. That is, ML estimator for $\theta = LS$ estimator for $\theta$.

But for $s > 0$ the optimization problem seems to be nontrivial.
Simpler approach?

Use the estimator

\[ \hat{\theta}_n^{\text{Huber}}(\omega) := \arg \min_{\theta} \sum_{i=1}^{n} \rho(Y_n,i(\omega) - f(t_n,i; \theta)), \quad \omega \in \Omega \]

for Huber’s loss function

\[ \rho(x) := \frac{1}{2} x^2 \mathbb{1}_{\{|x| \leq \delta\}} + \delta (|x| - \delta/2) \mathbb{1}_{\{|x| > \delta\}} \quad \text{for some } \delta > 0. \]

Makes the estimation more robust against outliers (and model misspecification?) compared to \( \hat{\theta}_n^{\text{LS}} \).
Comparison of estimators

Common criteria for the comparison of estimators are

\[
\text{Bias}(\hat{\theta}_n; \theta, s, \sigma) := \mathbb{E}^{(\theta, s, \sigma)}[\hat{\theta}_n] - \theta
\]

\[
\text{MSE}(\hat{\theta}_n; \theta, s, \sigma) := \mathbb{E}^{(\theta, s, \sigma)}[(\hat{\theta}_n - \theta)^2]
\]

Simulation

interval [0, 4] with n=30 design points
\(\theta = (c, h)\) with \(c = 2, h = 2\)
\(s = 0, 3, \sigma = 0, 1, \delta = 0, 1\)

<table>
<thead>
<tr>
<th></th>
<th>Bias((\hat{c}))</th>
<th>Bias((\hat{h}))</th>
<th>MSE((\hat{c}))^{1/2}</th>
<th>MSE((\hat{h}))^{1/2}</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS</td>
<td>-0.000014</td>
<td>-0.0796</td>
<td>0.0762</td>
<td>0.1393</td>
</tr>
<tr>
<td>ML</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Huber</td>
<td>-0.000024</td>
<td>-0.0632</td>
<td>0.0837</td>
<td>0.1273</td>
</tr>
</tbody>
</table>
Comparison of estimators

Common criteria for the comparison of estimators are

$$\text{Bias}(\hat{\theta}_n; \theta, s, \sigma) := \mathbb{E}(\theta, s, \sigma)[\hat{\theta}_n] - \theta$$
$$\text{MSE}(\hat{\theta}_n; \theta, s, \sigma) := \mathbb{E}(\theta, s, \sigma)[(\hat{\theta}_n - \theta)^2]$$

Simulation

interval [0, 4] with n=30 design points
$$\theta = (c, h) \text{ with } c = 1, h = 2$$
$$s = 0, 3, \sigma = 0, 1, \delta = 0, 1$$

<table>
<thead>
<tr>
<th></th>
<th>Bias((\hat{c}))</th>
<th>Bias((\hat{h}))</th>
<th>MSE((\hat{c}))^{1/2}</th>
<th>MSE((\hat{h}))^{1/2}</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS</td>
<td>0,0104</td>
<td>−0,0870</td>
<td>0,0810</td>
<td>0,1448</td>
</tr>
<tr>
<td>ML</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Huber</td>
<td>0,0049</td>
<td>−0,0657</td>
<td>0,0919</td>
<td>0,1303</td>
</tr>
</tbody>
</table>
Comparison of estimators

Common criteria for the comparison of estimators are

\[
\text{Bias}(\hat{\theta}_n; \theta, s, \sigma) := \mathbb{E}(\theta, s, \sigma)[\hat{\theta}_n] - \theta
\]

\[
\text{MSE}(\hat{\theta}_n; \theta, s, \sigma) := \mathbb{E}(\theta, s, \sigma)[(\hat{\theta}_n - \theta)^2]
\]

Simulation

interval [1, 3] with n=15 design points

\(\theta = (c, h)\) with \(c = 2, h = 2\)

\(s = 0, 3, \sigma = 0, 1, \delta = 0, 1\)

<table>
<thead>
<tr>
<th></th>
<th>Bias((\hat{c}))</th>
<th>Bias((\hat{h}))</th>
<th>MSE((\hat{c}))^{1/2}</th>
<th>MSE((\hat{h}))^{1/2}</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS</td>
<td>-0,00085</td>
<td>-0,0913</td>
<td>0,0879</td>
<td>0,1496</td>
</tr>
<tr>
<td>ML</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Huber</td>
<td>-0,00087</td>
<td>-0,0662</td>
<td>0,1045</td>
<td>0,1339</td>
</tr>
</tbody>
</table>
Questions

1. Consistency: $\hat{\theta}_n \xrightarrow{\text{a.s.}} \theta$ ?

2. Asymptotics: $\text{law}\{\sqrt{n}(\hat{\theta}_n - \theta)\} \xrightarrow{\text{w}} \mathcal{N}_0,\Gamma(\theta,s,\sigma)$ ?

3. Qualitative robustness:

   \[ s \leq \delta \varepsilon \quad \Longrightarrow \quad d(\text{law}\{\hat{\theta}_n^{\text{Huber}}\}, \text{law}\{\hat{\theta}_n^{\text{Huber}} \mid s = 0\}) \leq \varepsilon ? \]

4. Quantification of “degree” of robustness of $\hat{\theta}_n^{\text{Huber}}$ ?
Outline

Introduction

Raman Spectroscopy
Application & Experimental Set-up

Modeling of Different Estimators

Regression Model
Estimators

Deconvolution of Sparse Spikes

From the Estimator to the Variational Model
Reconstruction of a Single Spike
Numerical Deconvolution of Three δ-Spikes

Summary & Outlook
General Idea

Goal:
Formulate the problem as variational model, i.e.

$$\min_{\theta} D(\theta) + \alpha R(\theta)$$

- Approximate ML-estimator $\hat{\theta}^{ML}$
- Reformulate as data term $D$
- Add suitable regularization $R$

Example:

Inverse problem: $\mathcal{K} \theta = \hat{Y} + \xi$

Noise: $\xi$ Gaussian

Data term: $D(\theta) = \frac{1}{2} \| \mathcal{K} \theta - Y \|_2^2$
ML-Estimator and its Approximation

Reminder:

\[ \hat{\theta}_{n;\eta\neq 0}^{ML} = \arg\max_{\theta} \sum_{i=1}^{n} \log \left( \frac{1}{\sigma} \int e^{-\frac{(Y_{n,i} - f(t_{n,i} + u, \theta))^2}{2\sigma^2}} \phi_{0,s^2}(u) \, du \right) \]

Taylor approximation of log and exp:

\[ \hat{\theta}_{n;\eta\neq 0} = \arg\min_{\theta} \frac{1}{2} \sum_{i=1}^{n} \int (f(t_{n,i} + u, \theta) - Y_{n,i})^2 \phi_{0,s^2}(u) \, du \]
ML-Estimator and its Approximation

Reminder:

\[ \hat{\theta}_{n; \eta \neq 0}^{\text{ML}} = \arg\max_{\theta} \sum_{i=1}^{n} \log \left( \frac{1}{\sigma} \int e^{-\frac{(Y_{n,i} - f(t_{n,i+u,\theta}))^2}{2\sigma^2}} \phi_{0,s^2}(u) \, du \right) \]

Taylor approximation of log and exp:

\[ \hat{\theta}_{n; \eta \neq 0} = \arg\min_{\theta} \frac{1}{2} \sum_{i=1}^{n} \int \left( f(t_{n,i} + u, \theta) - Y_{n,i} \right)^2 \phi_{0,s^2}(u) \, du \]
Conversion to Variational Framework

Substitution: $\tau := t_{n,i} + u$

$$\min_{\theta} \frac{1}{2} \int \sum_{i=1}^{n} \left( f(\tau, \theta) - Y_{n,i} \right)^2 \phi_{0,s^2}(\tau - t_{n,i}) \, d\tau$$

Let $\mathcal{K}$ be the convolution with $e^{-\tau^2/v}$ and define $\mu := \sum_{k=1}^{K} h_k \delta_{c_k}$

$$\min_{\mu \in \mathcal{M}(\Omega)} \frac{1}{2} \int \sum_{i=1}^{n} (\mathcal{K}\mu(\tau) - Y_{n,i})^2 \phi_{0,s^2}(\tau - t_{n,i}) \, d\tau$$
Conversion to Variational Framework

Substitution: \( \tau := t_{n,i} + u \)

\[
\min_{\theta} \frac{1}{2} \int \sum_{i=1}^{n} \left( f(\tau, \theta) - Y_{n,i} \right)^2 \phi_{0,s^2}(\tau - t_{n,i}) d\tau \\
\sum_{k=1}^{K} h_k e^{-\frac{(\tau - c_k)^2}{\nu}}
\]

Let \( K \) be the convolution with \( e^{-\frac{\tau^2}{\nu}} \) and define \( \mu := \sum_{k=1}^{K} h_k \delta_{c_k} \)

\[
\min_{\mu \in \mathcal{M}(\Omega)} \frac{1}{2} \int \sum_{i=1}^{n} (K\mu(\tau) - Y_{n,i})^2 \phi_{0,s^2}(\tau - t_{n,i}) d\tau
\]
Conversion to Variational Framework

Substitution: \( \tau := t_{n,i} + u \)

\[
\min_{\theta} \frac{1}{2} \int \sum_{i=1}^{n} \left( f(\tau, \theta) - Y_{n,i} \right)^2 \phi_{0,s^2}(\tau - t_{n,i}) \, d\tau
\]

\[
\sum_{k=1}^{K} h_k e^{-\frac{(\tau-c_k)^2}{\nu}}
\]

Let \( \mathcal{K} \) be the convolution with \( e^{-\frac{\tau^2}{\nu}} \) and define \( \mu := \sum_{k=1}^{K} h_k \delta_{c_k} \)

\[
\min_{\mu \in \mathcal{M}(\Omega)} \frac{1}{2} \int \sum_{i=1}^{n} (\mathcal{K}\mu(\tau) - Y_{n,i})^2 \phi_{0,s^2}(\tau - t_{n,i}) \, d\tau
\]
Conversion to Variational Framework

Normalization:
\[
\sum_{i=1}^{n} \phi_{0,s^2}(\tau - t_i) = 1
\]

Additionally define:
\[
w(\tau) := \sum_{i=1}^{n} Y_{n,i} \phi_{0,s^2}(\tau - t_i)
\]

Reformulation:
\[
\min_{\mu \in \mathcal{M}(\Omega)} \frac{1}{2} \int (\mathcal{K}_\mu(\tau) - w(\tau))^2 \, d\tau
\]
Conversion to Variational Framework

Normalization:

\[ \sum_{i=1}^{n} \phi_{0,s^2}(\tau - t_i) = 1 \]

Additionally define:

\[ w(\tau) := \sum_{i=1}^{n} Y_{n,i} \phi_{0,s^2}(\tau - t_i) \]

Reformulation:

\[ \min_{\mu \in \mathcal{M}(\Omega)} \frac{1}{2} \| \mathcal{K}\mu - w \|_{L^2(\Omega)}^2 \]
Conversion to Variational Framework

1. **Pre-process the data**, i.e. convolve it with a Gaussian with standard deviation $s$

2. **Use a common $L^2$-data term** in a variational framework

Which regularization?

Prior knowledge:

Only few Gaussians should characterize the (part of the) spectrum

$\Rightarrow$ Use the **sparsity-promoting $\ell^1$-regularization**
Conversion to Variational Framework

1. **Pre-process the data**, i.e. convolve it with a Gaussian with standard deviation $s$

2. **Use a common $L^2$-data term** in a variational framework

Which regularization?

**Prior knowledge:**
Only few Gaussians should characterize the (part of the) spectrum

$\Rightarrow$ Use the **sparsity-promoting $\ell^1$-regularization**
Continuous Variational Model

Minimization in the space of finite Radon measures $\mathcal{M}(\Omega)$

The continuous counterpart to the discrete $\ell^1$-norm is the TV-norm in $\mathcal{M}(\Omega)$:

$$\min_{\mu \in \mathcal{M}(\Omega)} \frac{1}{2} \| \mathcal{K} \mu - w \|_{L^2(\Omega)}^2 + \alpha \| \mu \|_{TV(\Omega)}$$

Exact solution:

$$\hat{\mu} = \sum_{i=1}^{S} \rho_i \delta_{\xi_i}$$

$\mathcal{K}$ convolution with symmetric kernel $\bar{\phi} \in C^3(\Omega)$,

$\bar{\phi}(|x|)$ strong monotonic decreasing and $\max_x \bar{\phi}(x) = \bar{\phi}(0)$. 

Discrete Variational Model

Let

\[ \mu^N = \sum_{i=1}^{N} c_i \delta_{x_i} \]

be a solution of

\[ \min_{\mu^N \in \mathbb{R}^N} \frac{1}{2} \| A \mu^N - w \|_2^2 + \alpha \| \mu^N \|_1 \]

Questions:

- How well can we approximate \( \hat{\mu} \) by \( \mu^N \)?
- What happens when the grid becomes finer?
- When do \( \mu^N \) and \( \hat{\mu} \) consist of the same number of spikes?
Discrete Variational Model

Let

\[ \mu^N = \sum_{i=1}^{N} c_i \delta_{x_i} \]

be a solution of

\[ \min_{\mu^N \in \mathbb{R}^N} \frac{1}{2} \| A\mu^N - w \|_2^2 + \alpha \| \mu^N \|_1 \]

Questions:

- How well can we approximate \( \hat{\mu} \) by \( \mu^N \)?
- What happens when the grid becomes finer?
- When do \( \mu^N \) and \( \hat{\mu} \) consist of the same number of spikes?
Exact Reconstruction of a Single Spike

Exact spike at position $\xi$ between the grid points $x_k$ and $x_{k+1}$

- $\xi$ in green interval $\Rightarrow$ single peak is reconstructed at $x_k$
- $\xi$ in red interval $\Rightarrow$ there are at least peaks at $x_k$ and $x_{k+1}$

$$h$$

$\xi$ $x_k + \frac{\alpha h}{2\phi(0)}$ $x_k + \frac{h}{2}$ $x_{k+1}$

**Figure:** Region where the exact number of spikes can be reconstructed
Numerical Deconvolution of Three $\delta$-Spikes

(a) Spikes at the positions $\xi_1 = -\frac{\pi}{6}$, $\xi_2 = \frac{\pi}{300}$ and $\xi_3 = \frac{e}{5}$. Note that the plot is only for demonstration. The height of the spikes is actually $+\infty$.

(b) Continuous convolution $w = K \hat{\mu}$ of the exact data $\hat{\mu}$ with kernel $G$. 

\[ w = K \hat{\mu} = G * \mu \]
Numerical Deconvolution of Three $\delta$-Spikes

(c) Deconvolved data $\mu^N$ for $N = 50$ and $\alpha = 0.1$

(d) Deconvolved data $\mu^N$ for $N = 100$ and $\alpha = 0.1$

Figure: Deconvolution of three spikes, which were positioned between the grid points
Numerical Deconvolution of Three $\delta$-Spikes

(a) Minimal distances of the reconstructed peaks to the exact spikes

(b) Number of reconstructed peaks

Figure: Results for different discretizations; $\alpha = 0.1$
Summary

✓ Spectra better comparable through description as linear combination of few Gaussians
✓ Modeling of noise and peak shifts possible
× Not easily implementable without approximation
– Huber estimator produces better results than the approximation, however, under slightly different conditions
✓ Regularization could improve the results of the approximated estimator
× Only fixed standard deviation was considered
Outlook & Open Questions

¿? Does regularization improve the results?
   ▼ Implementation

¿? Is the Huber estimator reasonable?

¿? Does the combination of the Huber estimator and a regularization yield better results?

¿? Is it possible to consider the exact ML-estimator?

¿? Can we additionally estimate the standard deviations $s$ and $\sigma$?
   ▼ Far more complicated

¿? How close can the peaks get in order to still be able to separate them?
Thank you!

Questions, remarks, ideas?