International Workshop on Bayesian Inference and
Maximum Entropy Methods in Science and Engineering

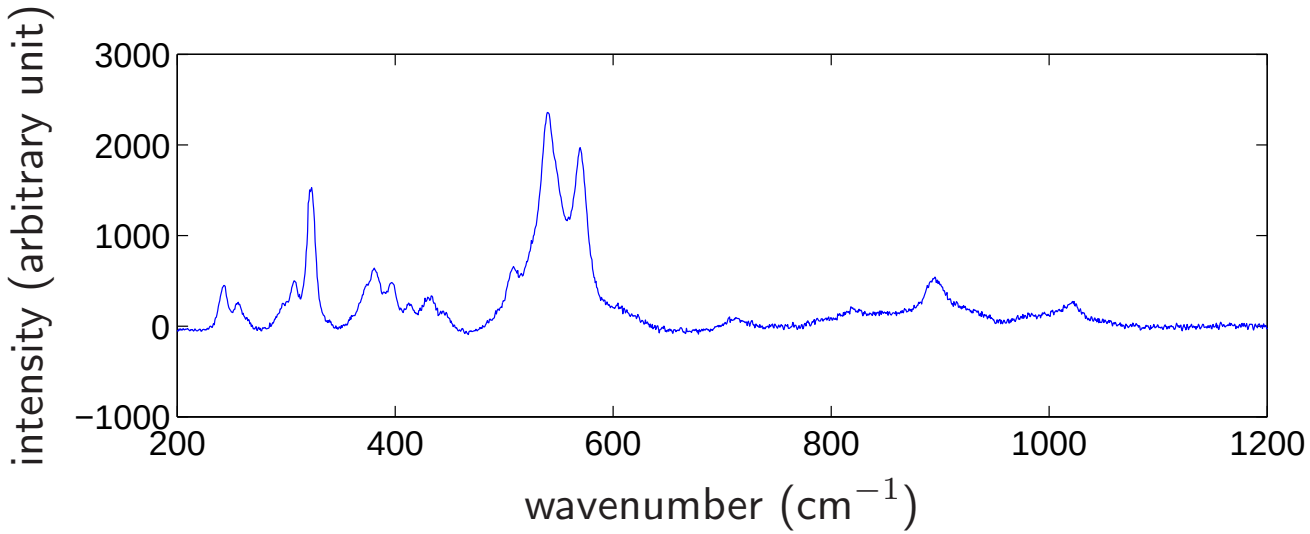
DECOMPOSITION OF A CHEMICAL SPECTRUM USING A MARKED POINT PROCESS AND A CONSTANT DIMENSION MODEL

V. Mazet, D. Brie, J. Idier

CRAN UMR 7039, Nancy University, CNRS,
BP 239, 54506 Vandœuvre-lès-Nancy Cedex, France
firstname.lastname@cran.uhp-nancy.fr

IRCCyN UMR 6597, École Centrale de Nantes, CNRS,
1 rue de la Noë, BP 92101, 44321 Nantes Cedex 3, France
jerome.idier@irccyn.ec-nantes.fr

Introduction



Goal: estimating the peak parameters (locations, amplitudes and widths) in a spectrum.

➜ Provide an interpretation for physico-chemists.

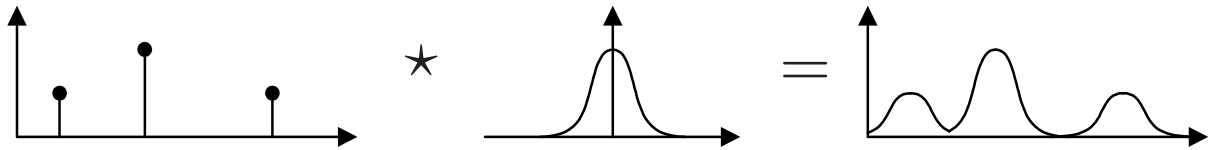
➜ Bayesian approach + MCMC method.

Summary

1. Problem Formulation

Marked point process: finite set of objects lying in a bounded space and characterized by their locations and some marks.

➜ Blind sparse spike train deconvolution

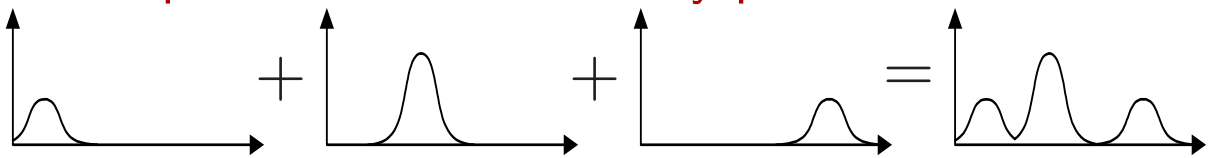


→ Bernoulli-Gaussian process (widespread model for sparse spike trains)

Drawbacks:

- common implementation with MCMC methods not efficient
- peaks located on discrete positions
- one peak shape

➜ Decomposition into elementary patterns



→ $\mathbf{y} = \sum_{k=1}^{K} \mathbf{f}(\mathbf{n}_k, \mathbf{w}_k, \mathbf{s}_k) + \mathbf{e}$

2. Proposed Model
2.1 A Constant Dimension Model

$$\mathbf{y} = \sum_{k=1}^{K} \mathbf{f}(\mathbf{n}_k, \mathbf{w}_k, \mathbf{s}_k) + \mathbf{e}$$

Problem: peak number unknown

$\Rightarrow$ system order likely to change!

➜ MCMC techniques for model uncertainty (RJMCMC algorithm, ...)

➜ Constant Dimension Model

peak number equals to constant $K_{\max}$ (upper bound fixed by the user).

Bernoulli-Gaussian model $\rightarrow \mathbf{q} \sim \mathcal{B}er(\lambda)$ codes the peak occurrences:

- $\mathbf{q}_k = 1$: the kth peak is present (at $\mathbf{n}_k$ with amplitude $\mathbf{w}_k$ and width $\mathbf{s}_k$)
- $\mathbf{q}_k = 0$: the kth peak is not present

$$\Rightarrow \qquad \mathbf{y} = \sum_{k=1}^{K_{\max}} \mathbf{f}(\mathbf{n}_k, \mathbf{w}_k, \mathbf{s}_k) + \mathbf{e}$$

➜ variable number smaller than a common BG implementation ($3K_{\max}$ vs. N).

➜ allows to use Gibbs sampler

Noise:

white, Gaussian and i.i.d.

$$\mathbf{e} \sim \mathcal{N}(\mathbf{0}, r_{\mathbf{e}}\mathbf{I})$$

Peak Location:

uniformly distributed on $[1, N]$

$$\mathbf{n}_k \sim \mathcal{U}_{[1,N]}$$

Peak Amplitude:

BG process + positive amplitudes

$$\mathbf{q}_k \sim \mathcal{B}er(\lambda)$$

$$\mathbf{w}_k|\mathbf{q}_k \sim \begin{cases} \delta_0(\mathbf{w}_k) & \text{if } \mathbf{q}_k = 0 \\ \mathcal{N}^+(0, r_{\mathbf{w}}) & \text{if } \mathbf{q}_k = 1 \end{cases}$$

Peak Width:

inverse gamma with mean 6 cm^{-1}
and variance 2.5 cm^{-1}

$$\mathbf{s}_k \sim \mathcal{IG}(\alpha_{\mathbf{s}}, \beta_{\mathbf{s}})$$

➜ Hyperparameters:

Bernoulli parameter:

conjugate prior to penalize high values $\lambda \sim \mathcal{B}e(1, K_{\max} + 1)$

Peak Amplitude Variance:

conjugate prior less informative as possible $r_{\mathbf{w}} \sim \mathcal{IG}(\alpha_{\mathbf{w}}, \beta_{\mathbf{w}})$

Noise variance:

Jeffreys prior $r_{\mathbf{e}} \sim 1/r_{\mathbf{e}}$

Peak Location: $$\mathbf{n}_k|\ldots \sim \exp\left(-\frac{1}{2r_\mathbf{e}}\left|\left|\mathbf{y}-\sum_{l=1}^{K_{\max}}\mathbf{f}(\mathbf{n}_l,\mathbf{w}_l,\mathbf{s}_l)\right|\right|^2\right)\mathbb{1}_{[1,N]}(\mathbf{n}_k)$$

Peak Amplitude: $$\mathbf{q}_k|\ldots \sim \mathcal{B}er(\lambda_k)$$

$$\mathbf{w}_k|\ldots \sim \begin{cases}\delta_0(\mathbf{w}_k) & \text{if } \mathbf{q}_k=0\\ \mathcal{N}^+(\mu_k,\rho_k) & \text{if } \mathbf{q}_k=1\end{cases}$$

Peak Width: $$\mathbf{s}_k|\ldots \sim \exp\left(-\frac{1}{2r_\mathbf{e}}\left|\left|\mathbf{y}-\sum_{l=1}^{K_{\max}}\mathbf{f}(\mathbf{n}_l,\mathbf{w}_l,\mathbf{s}_l)\right|\right|^2-\frac{\beta_\mathbf{s}}{\mathbf{s}_k}\right)\frac{1}{\mathbf{s}_k^{\alpha_\mathbf{s}+1}}\mathbb{1}_{\mathbb{R}^+}(\mathbf{s}_k)$$

Bernoulli parameter: $$\lambda|\ldots \sim \mathcal{B}e\big(K+1,2K_{\max}-K+1\big)$$

Peak Amplitude Variance: $$r_\mathbf{w}|\ldots \sim \mathcal{IG}\left(\frac{K}{2}+\alpha_\mathbf{w},\frac{\mathbf{w}^T\mathbf{w}}{2}+\beta_\mathbf{w}\right)$$

Noise variance: $$r_\mathbf{e}|\ldots \sim \mathcal{IG}\left(\frac{N}{2},\frac{1}{2}\left|\left|\mathbf{y}-\sum_{l=1}^{K_{\max}}\mathbf{f}(\mathbf{n}_l,\mathbf{w}_l,\mathbf{s}_l)\right|\right|^2\right)$$

$$\mathbf{n}_k|\ldots \sim \exp\left(-\frac{1}{2r_{\mathbf{e}}}\left|\left|\mathbf{y}-\sum_{l=1}^{K_{\max}}\mathbf{f}(\mathbf{n}_l,\mathbf{w}_l,\mathbf{s}_l)\right|\right|^2\right)\mathbb{1}_{[1,N]}(\mathbf{n}_k)$$

Metropolis-Hastings algorithm $\rightarrow$ which proposal distribution?

➜ If the peak is present ($\mathbf{q}_k = 1$)
define precisely its location: $\mathcal{N}^{[1,N]}(\mathbf{n}_k^{(i-1)}, r_{\mathbf{n}})$

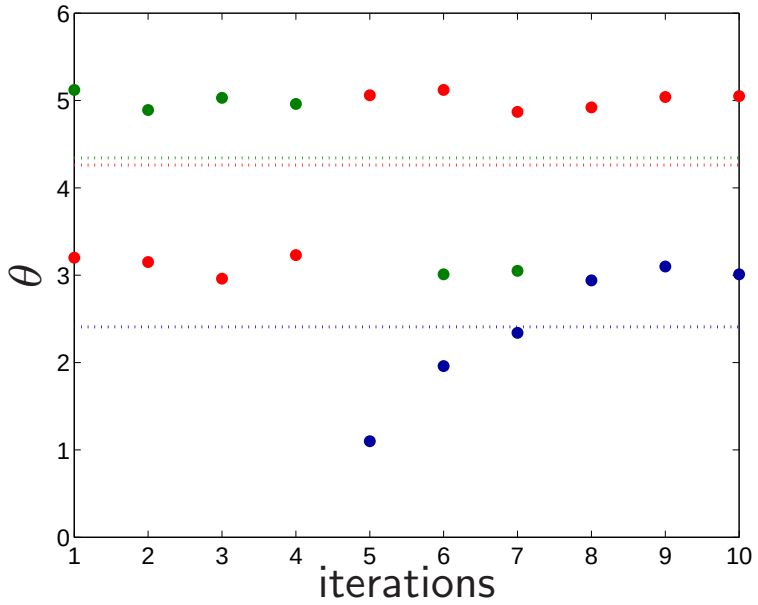
➜ If the peak is absent ($\mathbf{q}_k = 0$)
explore the entire space: $\mathcal{U}_{[1,N]}$

$$\Rightarrow \qquad q(\widetilde{\mathbf{n}}_k) = \delta_0(\mathbf{q}_k)\mathcal{U}_{[1,N]} + \delta_1(\mathbf{q}_k)\mathcal{N}^{[1,N]}(\mathbf{n}_k^{(i-1)}, r_{\mathbf{n}}).$$

3. Label Switching

The label switching problem is due to 2 phenomena:

- same posterior for all permutation of k: $p(\theta_1, \theta_2, \theta_3|\mathbf{y}) = p(\theta_2, \theta_3, \theta_1|\mathbf{y})$
- Gibbs sampler able to explore the $k!$ permutation possibilities



$$\widehat{\theta_1} = 4.26, \quad \widehat{\theta_2} = 4.34, \quad \widehat{\theta_3} = 2.41$$

3. Label Switching

Proposed Method

Minimizing the following cost function (see [Stephens 1997]):

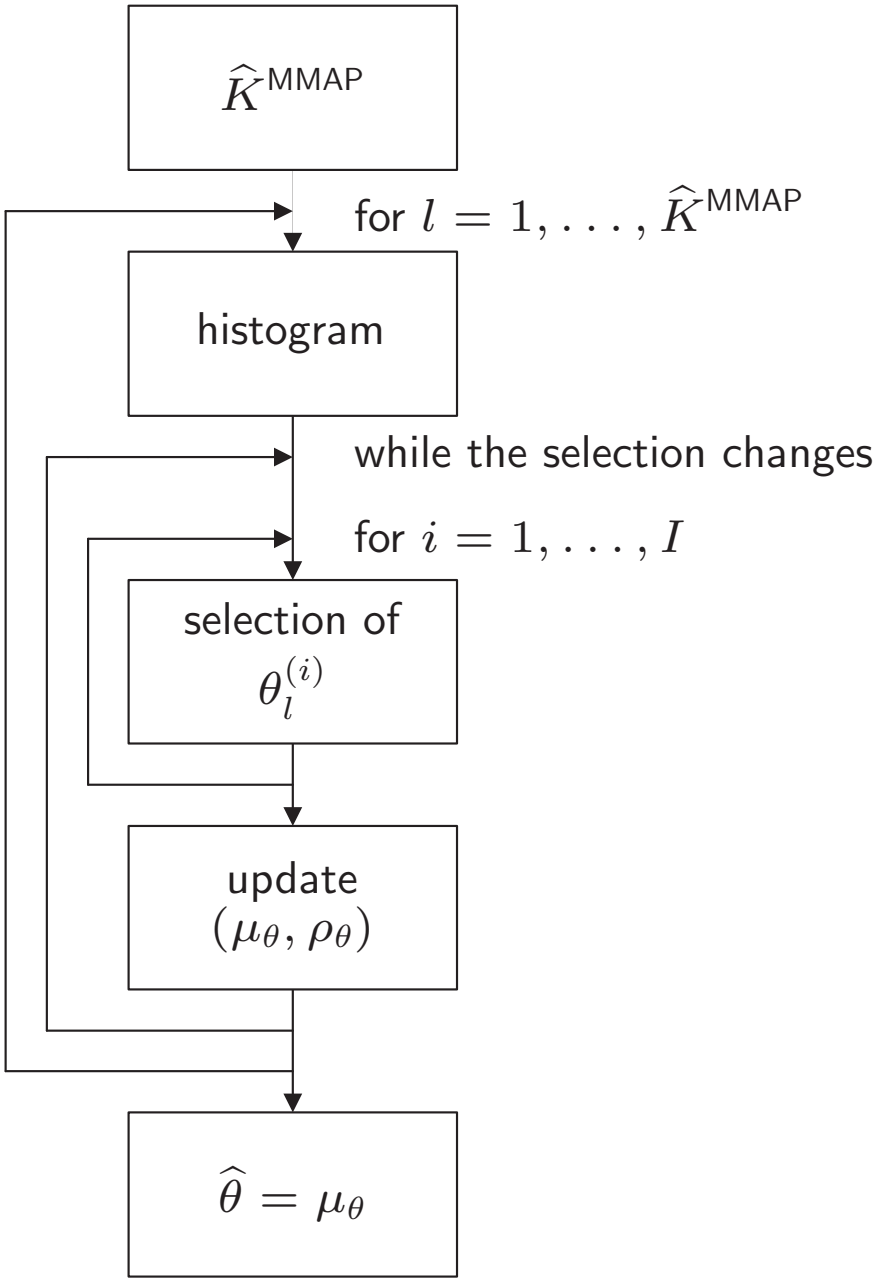
$$\mathcal{L}_0(\mathbf{n}, \mathbf{w}, \mathbf{s}, \mu_{\mathbf{n}}, \rho_{\mathbf{n}}, \mu_{\mathbf{w}}, \rho_{\mathbf{w}}, \mu_{\mathbf{s}}, \rho_{\mathbf{s}}) = -\ln \left[\prod_{k=1}^{K_{\max}} \mathcal{N}(\mathbf{n}_k | \mu_{\mathbf{n}k}, \rho_{\mathbf{n}k}) \mathcal{N}(\mathbf{w}_k | \mu_{\mathbf{w}k}, \rho_{\mathbf{w}k}) \mathcal{N}(\mathbf{s}_k | \mu_{\mathbf{s}k}, \rho_{\mathbf{s}k}) \right]$$

Major differences to general relabelling algorithms:

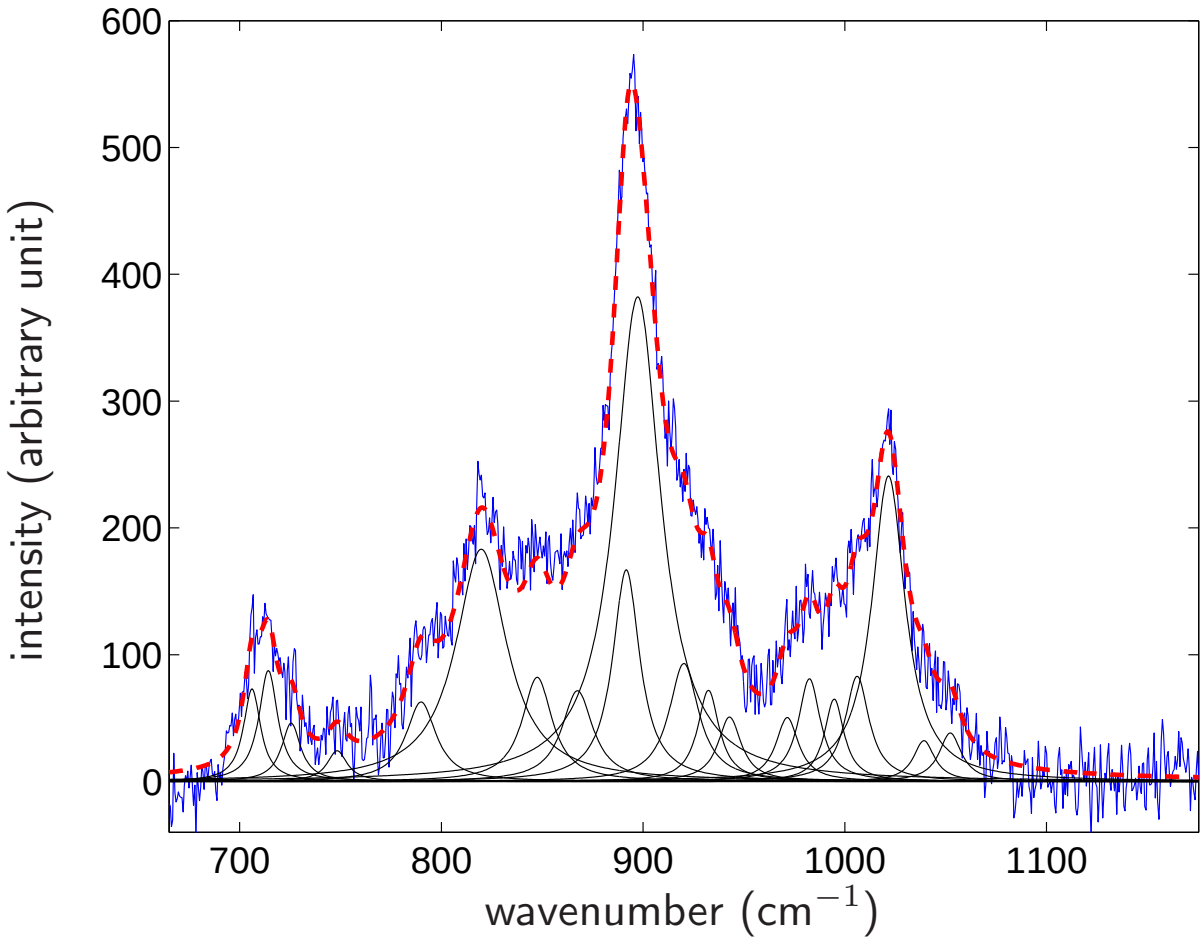
- initialization obtained by selecting the maximum in the histogram of $(\mu_{\mathbf{n}}, \mu_{\mathbf{w}}, \mu_{\mathbf{s}})$ (closer to the global optimum than a simple identity permutation)
- relabelling $(\mathbf{n}_l, \mathbf{w}_l, \mathbf{s}_l)$ one after the other (no permutation)
- taking into account the fact that the peak number is expected to change

3. Label Switching

$\theta = (\mu_{\mathbf{n}}, \mu_{\mathbf{w}}, \mu_{\mathbf{s}})$



4. Application



Raman spectrum of gibbsite $Al(OH)_3$

➜ 10,000 iterations (burn-in period of 5,000 iterations).

➜ Initialization: spectrum with no peak, $\lambda^{(0)} = 0.5$, $r_{\mathbf{w}}^{(0)} = 10$, $r_{\mathbf{e}}^{(0)} = 0.1$.

Conclusion

➜ Signal decomposition into elementary patterns (marked point process)

Alternative to blind sparse spike train deconvolution

- more efficient than a common implementation with BG model
- peaks located on a continuous space
- peak with different shapes

➜ Constant dimension model

Alternative to RJMCMC

➜ New method for label switching

- initialization close to the global optimum using an histogram
- relabelling with no permutation
- the variable number may change