Inférence Bayésienne et problémes inverses

Quentin Legros

Laboratoire Jean Kuntzmann
Université Grenoble Alpes

March 13, 2023

Example 2 - the presumption of innocence

The case of Sally Clark

- In 1999 (UK), a mother is accused of the murder of her two infant sons.
- The prosecutor defense speech, based on the expert pediatrician statistics, was overwhelming.
- It convinced all of the audience, and the mother was imprisoned.

Example 2 - the presumption of innocence

Prosecutor defense

- The prosecution case relied on statistical evidence, based on Sudden infant death syndrome (SIDS) probability.
- Prosecution expert argued $p(\text{SIDS}) = 1/8543$
- $p(\text{both SIDS}) = p(\text{SIDS}) \times p(\text{SIDS}) \simeq 1/73$ millions.
- Concluded $p(\text{innocence}) \simeq 1/73$ millions

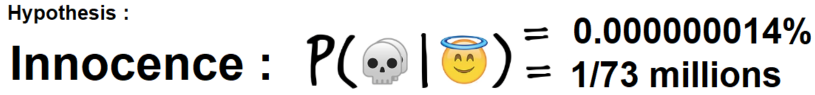
Royal Statistical Society reaction

- In October 2001, the RSS expressed its concern at the "misuse of statistics in the courts".
- In January 2002 : "The calculation leading to $1/73$ millions is false"

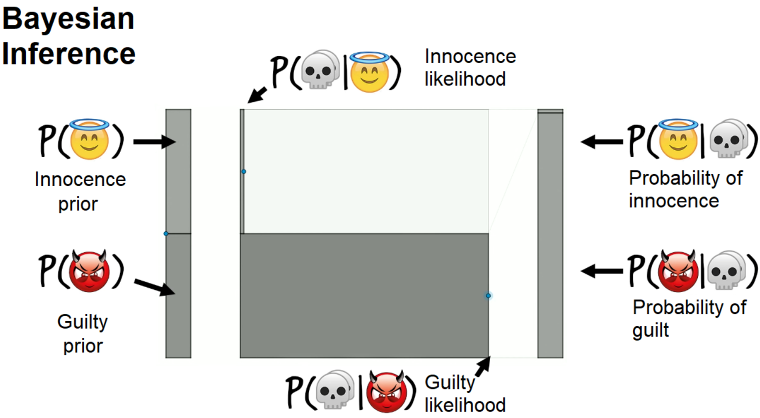
Example 2 - the presumption of innocence

What went wrong

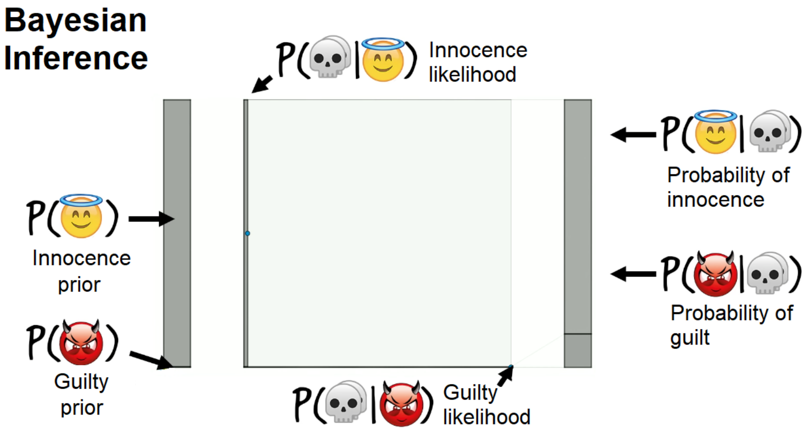
- Two SIDS in the same family are not independent.
- Genetic or environmental factors.
- However, the reasoning is also wrong.



Example 2 - the presumption of innocence



Example 2 - the presumption of innocence



Example 2 - the presumption of innocence

Bayesian analysis

- Two SIDS in the same family is rare.
- But infanticidal mothers are also very rare (1/500 millions).
- All hypothesis should be compared.
- Presumption of innocence $\Rightarrow$ a priori, peoples are statistically innocents.
- Guilt : evidence should be higher than the innocence prior belief.
- Sally Clark was convicted after evidence of the statistics misrepresentation.

Plan

1. Introduction
2. In practice
3. Lidar analysis
4. Pseudo-Bayesian approach
5. Conclusion

Plan

1. Introduction
2. In practice
3. Lidar analysis
4. Pseudo-Bayesian approach
5. Conclusion

Frequentist vs Bayesian statistics

Probability of events

Knowing a given theory

Frequentist statistics

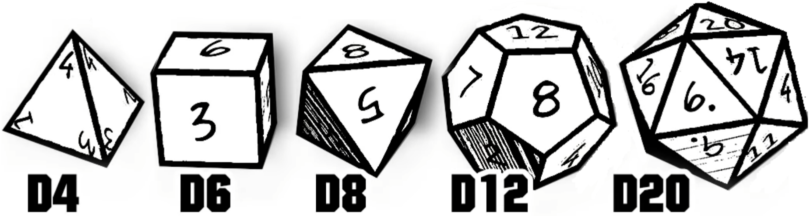
Probability of theories

Knowing given events

Bayesian statistics

Frequentist vs Bayesian statistics

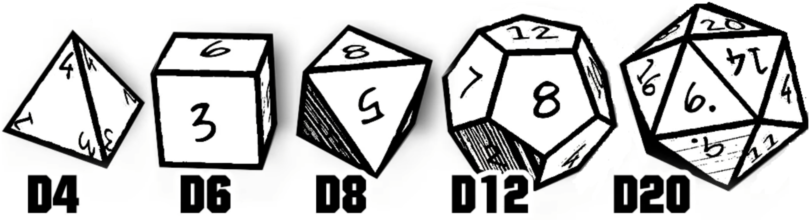
- Let's consider 5 different dices.



First questions

- Given dice and a result (i.e. D12 and 7), is this result extraordinary?
- Frequentist approach: depends of what is extraordinary (p value).
- Here $p \simeq 8\% \Rightarrow$ not extraordinary.
- Different conclusion if we obtain a 10 times successive 7

Frequentist vs Bayesian statistics



Second questions

- We obtain a 7, what dice has been thrown?
- We can discard D4 and D6.
- Hard question using frequentist inference.
- However the answer is simple when adopting a Bayesian approach.

Frequentist vs Bayesian statistics

Bayes' theorem

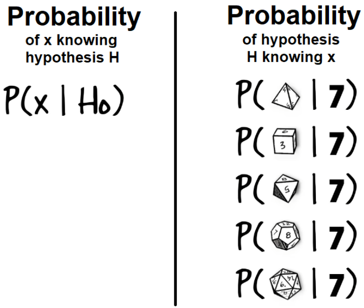
$$P(A)P(B|A) = P(B)P(A|B)$$

- with A and B two events.
- Allows to express the union of two events.
- Laplace rewrites it several years later:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

- Before we could only compute the probability of an event if knowing the cause.
- Laplace wrote "Mémoire sur la probabilité des causes par les événements", 1771

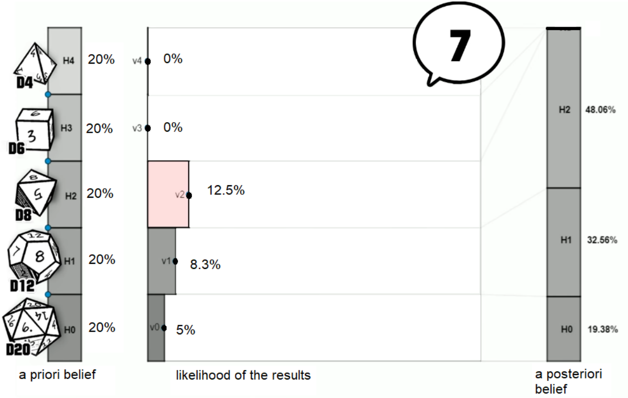
Frequentist vs Bayesian statistics



Second questions

- Allows for the evaluation of each hypothesis
- We are not rejecting hypothesis anymore, we compare them.
- Require an a prior belief on hypothesis before testing them.

Frequentist vs Bayesian statistics



Second questions

- Posterior belief : D8 is the most probable
- If another dice is thrown, we can update the prior belief.

Plan

1 Introduction

2 In practice

3 Lidar analysis

4 Pseudo-Bayesian approach

5 Conclusion

Summary

Bayes' theorem

$$P(variable|data) = \frac{P(data|variable)P(variable)}{P(data)}$$

- We want to estimate a set of variable from a posterior distribution.
- Need a likelihood function and prior models.
- We can work on
  - The observation model $\Rightarrow$ likelihood.
  - The priors model $\Rightarrow$ regularization.
  - The estimation strategy $\Rightarrow$ computational tractability, performances.

Observation model

Likelihood

- Likelihood is a function of parameters of a statistical model.
- Computed from the observations.
- $\Rightarrow$ Observation model.
  - What we (almost) always do.
  - Traducing a real-life problem into a mathematical equation.
  - Maths are useful to solve problem.
- Note: maximum likelihood estimation can be enough.

Prior model

Meaning

- Prior : probability knowledge before observation.
- Can inform on the level of knowledge (Uniform, Spike and Slab).
- Regularization (Markov chain, physical constraints)
- Can be adapted to estimation strategies (assumed density filtering)

Posterior probability

Bayes' theorem

$$P(var1, var2, var3|data) \propto P(data|var1, var2, var3)P(var1)P(var2)P(var3)$$

Meaning

- Posterior proportional to the product of likelihood and priors.
- Different type of estimator (MMSE, MAP)
- Convexity is preferred.
- Alternative : Markov chain Monte Carlo algorithms.
- Link with optimization techniques.

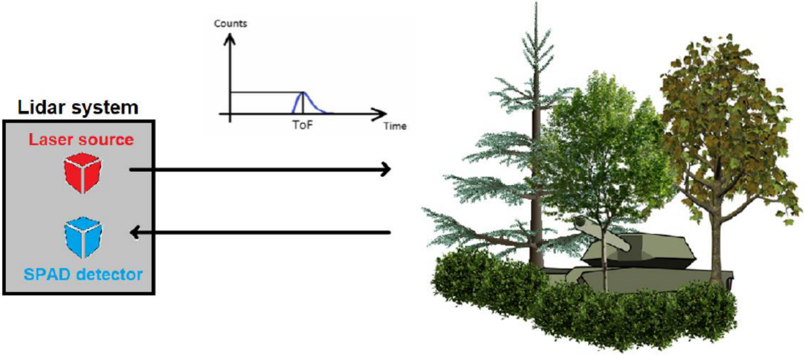
Summary

- Define an observation model.
- Compute the likelihood using the model and the observations.
- Assign a prior model to each model parameter.
- Compute the posterior distribution.
- Infer estimates from the posterior distribution.

Plan

1. Introduction
2. In practice
3. Lidar analysis
4. Pseudo-Bayesian approach
5. Conclusion

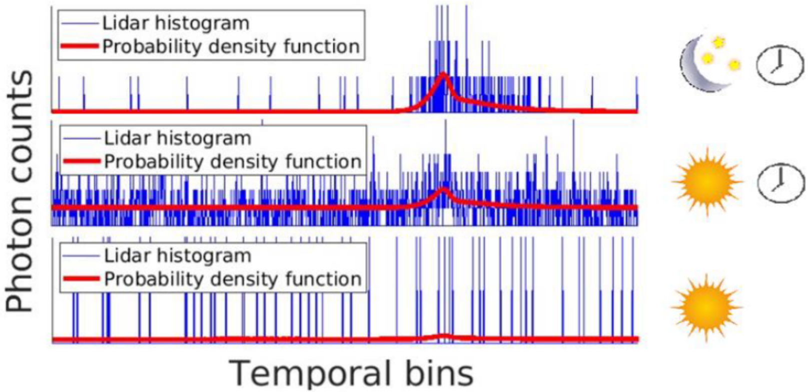
Single-photon Lidar



Motivation

- Registering of first incident photon.
- Excellent temporal resolution (few ps).
- Limitation to ambient illumination.

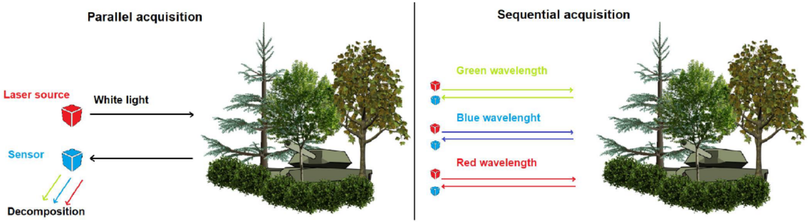
Single-photon Lidar



Motivation

- The imaging conditions determine the reconstruction problem difficulty.

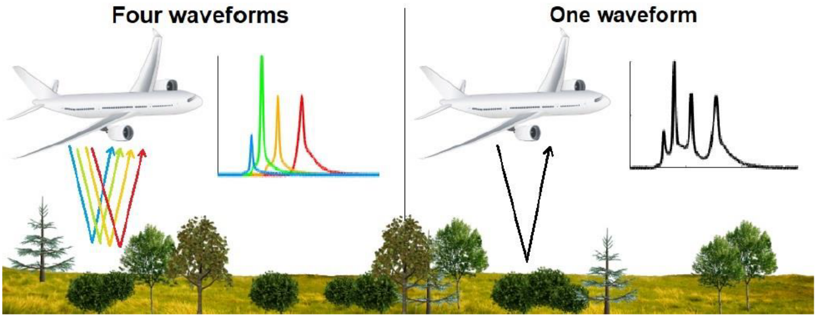
Multispectral analysis



Motivation

- Complex imaging system required.
- Longer acquisition time.
- Data size dependent to the number of wavelengths.

Single-waveform multispectral



Motivation

- All colors acquired with a single detector [1].
- Acquisition time / data size ⇒ independent to spectrum size

[1] X. Ren, Y. Altmann, R. Tobin, A. McCarthy, S. McLaughlin and G. S. Buller, "Wavelength-time coding for multispectral 3D imaging using single-photon Lidar", Optics Express 2018

How and what to do

Objective

- 3D imaging from single waveform multispectral Lidar data.
- Estimation of depth, reflectivity and background profiles.
- Fast estimation process.
- Uncertainty quantification.

Outline

- Observation model.
- Prior models.
- Estimation strategy.
- Results.
- Conclusion and future work.

Observation model

Standard observation model

$$y_t | (r, b, t_0) \sim \mathcal{P}\left(rg(t - t_0) + b\right), \quad (1)$$

- $\mathcal{P}$: Poisson distribution.
- y_t : photon counts in time bin t.
- r : surface reflectivity in pixel.
- b : background level in pixel.
- t_0 : surface depth in pixel.
- g : instrumental response function.

Observation model

Standard observation model

$$y_{n,t}|(\boldsymbol{r}_n, b_n, t_n) \sim \mathcal{P}\left(b_n + \sum_{\ell=1}^{L} r_{n,\ell} g_\ell(t - t_n)\right), \quad (2)$$

- $\mathcal{P}$: Poisson distribution.
- $y_{n,t}$: photon counts of the nth photon in the tth temporal bin
- $\boldsymbol{r}_n$: spectral reflectivity of the nth pixel
- b_n :background level of the nth pixel
- L : number of wavelengths
- t_n : depth profile in the nth pixel
- g_ℓ : impulse response function (IRF) associated to the ℓth wavelength
- T : histograms length

Bayesian model

Joint likelihood

$$p(\boldsymbol{Y}|\boldsymbol{R}, \boldsymbol{b}, \boldsymbol{t}) = \prod_n p(\boldsymbol{y}_n|\boldsymbol{r}_n, t_n, b_n). \tag{3}$$

Prior models

- Depth : Total variation [2]

$$p(\boldsymbol{t}|\epsilon) = \exp\left[-\epsilon \mathrm{TV}(\boldsymbol{t})\right], \tag{4}$$

- Background level : Gamma

$$f(\boldsymbol{b}|\alpha, \beta) = \prod_n \mathcal{G}(b_n|\alpha, \beta), \tag{5}$$

[2] Y. Altmann, A. Maccarone, A. McCarthy, G. Newstadt, G. S. Buller, S. McLaughlin and A. Hero, "Robust spectral unmixing of sparse multispectral Lidar waveforms using gamma Markov random fields", IEEE TCI 2018

Bayesian model

Prior models

- Reflectivity : Cluster-Gamma [3]

$$f(\boldsymbol{R}|k_{c,\ell},\theta_{c,\ell}) = \prod_{c=1}^{C}\prod_{n\in I_c}\prod_{\ell=1}^{L}\mathcal{G}(r_{n,\ell};k_{c,\ell},\theta_{c,\ell}), \tag{6}$$

- Hyperparameters, hierarchical model: gamma and inverse gamma

$$f(k_{c,\ell},\theta_{c,\ell}) = \prod_{c,l}\mathcal{G}(k_{c,\ell};2,0.5)\mathbf{1}_{(1,\infty)}(k_{c,\ell})\mathcal{IG}(\theta_{n,\ell};1.01,0.5). \tag{7}$$

[3] Q. Legros, S. Meignen , S. McLaughlin, and Y. Altmann , "Expectation-Maximization based approach to 3D reconstruction from single-waveform multispectral Lidar data", IEEE TCI 2020

Completing Bayesian model and Stochastic EM algorithm

Prior models

- Joint posterior distribution (with $\boldsymbol{\Phi} = (\{k_{c,\ell}, \theta_{c,\ell}\}_{c,\ell}), \alpha, \beta)$:

$$f(\boldsymbol{R}, \boldsymbol{b}, \boldsymbol{t}, \boldsymbol{\Phi}|\boldsymbol{Y}) \propto p(\boldsymbol{Y}|\boldsymbol{R}, \boldsymbol{b}, \boldsymbol{t})p(\boldsymbol{t})f(\boldsymbol{R}, \boldsymbol{b}|\boldsymbol{\Phi})f(\boldsymbol{\Phi}), \tag{8}$$

- Challenging joint estimation: multimodal likelihood function
- Sequential estimation $\Rightarrow$ stochastic EM algorithm

$$(\hat{\boldsymbol{R}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{\Phi}}) = \underset{\boldsymbol{R},\boldsymbol{b},\boldsymbol{\Phi}}{\operatorname{argmax}} \sum_{\boldsymbol{t}} f(\boldsymbol{R}, \boldsymbol{b}, \boldsymbol{t}, \boldsymbol{\Phi}|\boldsymbol{Y})$$

- Marginal maximum a posteriori depth estimation

$$\hat{t}_n = \underset{t_n}{\operatorname{argmax}} \sum_{\boldsymbol{t}_{\setminus n}} p(\boldsymbol{t}|\boldsymbol{Y}, \hat{\boldsymbol{R}}, \hat{\boldsymbol{b}}, \hat{\boldsymbol{\Phi}}), \quad \text{with } \boldsymbol{t}_{\setminus n} \text{ is } \boldsymbol{t} \text{ expect } t_n. \tag{9}$$

Estimation strategy

Computational bottlenecks

- Computation of the likelihood function.
- Depth sampling in the stochastic step.

Likelihood function computed on a coarser grid

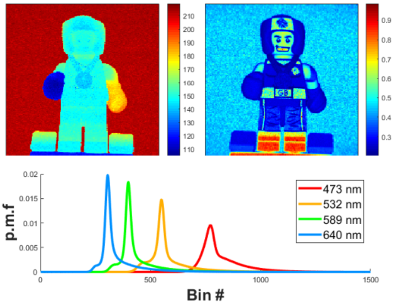
- Discrete range grid $\Rightarrow$ used to compute marginal probability.
- Depth grid subsampled by a factor T_s.
- Allows satisfactory depth sampling if T_s remains lower than the delays between each IRF.

Reducing the number of depth samples

- Only one sample.
- Multimodal nature of the likelihood will help for robust sampling.

Results

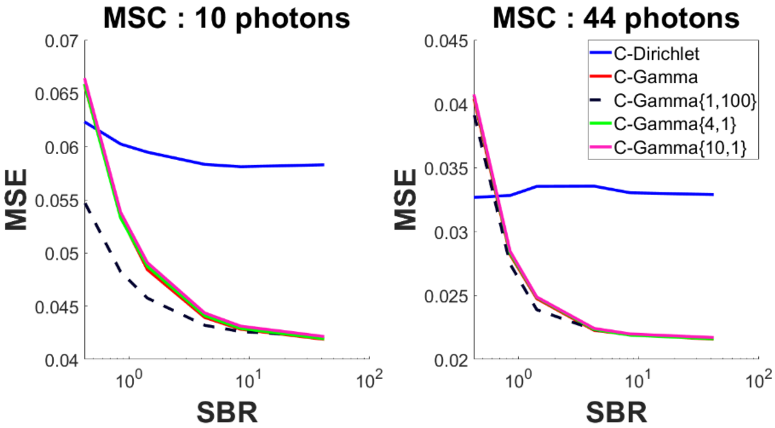
Set $G_\ell = \sum_{t=1}^{T} g_\ell(t - t_n)$.



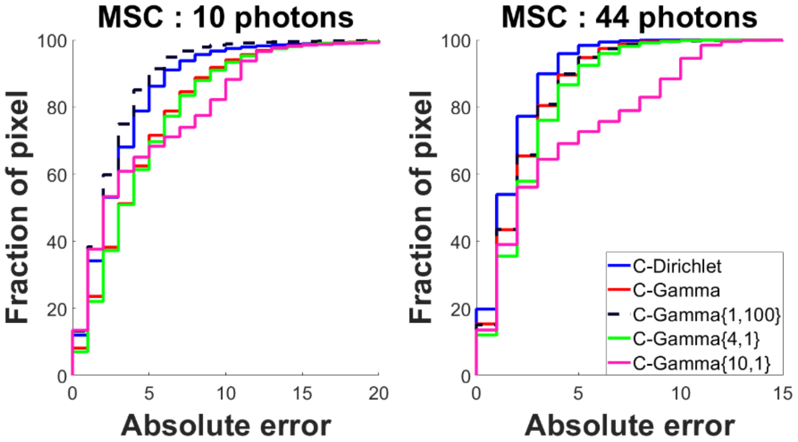
- SBR: Signal-to-background ratio.
- MSC: Mean signal photon counts.
- Comparison with C-Dirichlet [3].
- Different values for T_s and N_t.
- Denoted as C-Gamma $\{T_s, N_t\}$.
- MCS $= \frac{1}{N} \sum_{n,\ell} r_{n,\ell} G_\ell$
- SBR $= \frac{1}{N} \sum_{n,\ell} \frac{r_{n,\ell} G_\ell}{T b_n}$

[3] Q. Legros, S. Meignen , S. McLaughlin, and Y. Altmann , "Expectation-Maximization based approach to 3D reconstruction from single-waveform multispectral Lidar data", IEEE TCI 2020

Results - Reflectivity



Results - Depth



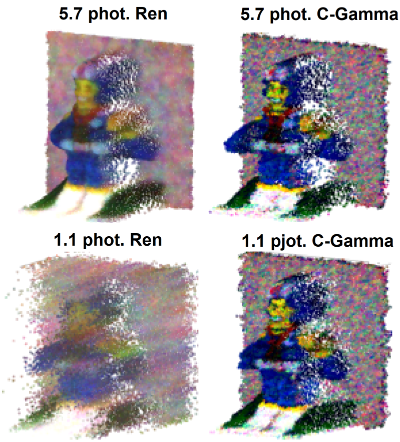
Results - Computational complexity

Computational cost of competing approaches for synthetic SW-MSL data analysis.

MSC/SBR	10 / ∞	10 / 0,046	44 / ∞	44 / 0,426
C-Gamma{1,1}	104s	132s	350s	493s
C-Gamma{4,1}	43s	49s	55s	69s
C-Gamma{10,1}	34s	36s	38s	42s
C-Dirichlet	530s	592s	629s	662s

- Similar estimation performance.
- Significant speed-up.

Results - Example



- Depth/RGB reconstruction with 1.1 (bottom) and 5.7 (top) signal photons per pixel.
- Non-negligible background (SBR= 1.4).
- Comparison with existing approach [4].

[4] X. Ren, Y. Altmann, R. Tobin, A. McCarthy, S. McLaughlin and G. S. Buller, "Wavelength-time coding for multispectral 3D imaging using single-photon Lidar", Optics Express 2018

Conclusion

Conclusion

- Recent Bayesian model for 3D reconstruction from single-waveform multipectral Lidar data.
- Bayesian approaches can be relatively fast.
- Perform well with few information.

Plan

1. Introduction
2. In practice
3. Lidar analysis
4. Pseudo-Bayesian approach
5. Conclusion

Introduction

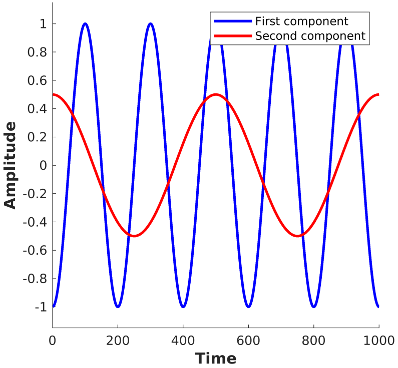
- Focus on MultiComponent Signal (MCS).

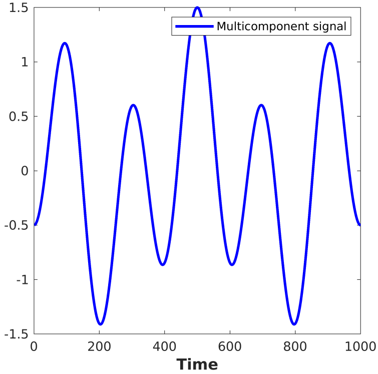
$$x(t) = \sum_{k=1}^{K} x_k(t) \quad \text{, with } x_k(t) = a_k(t)\, e^{j\phi_k(t)},$$

Investigated approaches

- Mixture of K superimposed components.
- $a_k(t)$ and $\phi_k(t)$ the time-varying amplitude and phase of component k.

Introduction

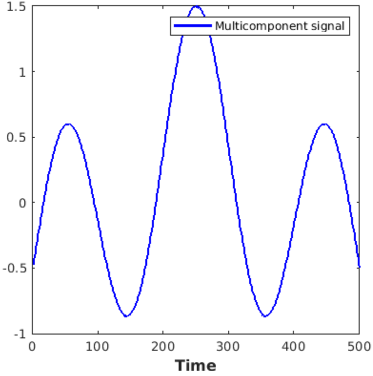


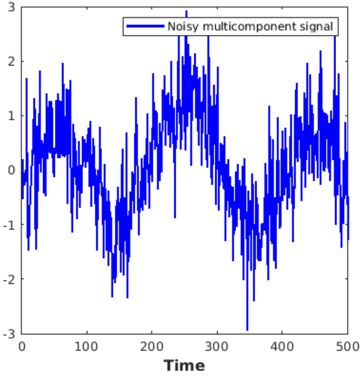


Motivation

- Variety of application.
- Audio, medical, astronomical, echolocation,...
- Purpose: extracting the components.

Introduction





Limitations

- Challenge in the presence of noise.
- Acquisition condition, recording device, presence of outliers,...

Introduction: Time-frequency tools

Frequency analysis

- Classical approach: frequency analysis.
- The Fourier transform of a signal x writes

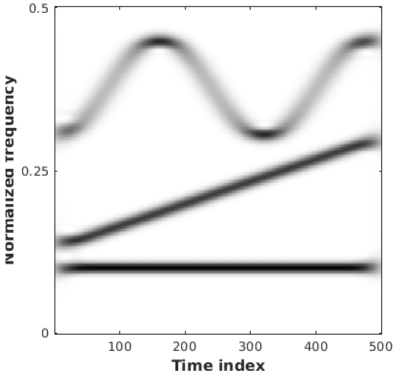
$$\hat{x}(\omega) = \int_{\mathbb{R}} x(u)\, \mathrm{e}^{-j\omega u}\, \mathrm{d}u,$$

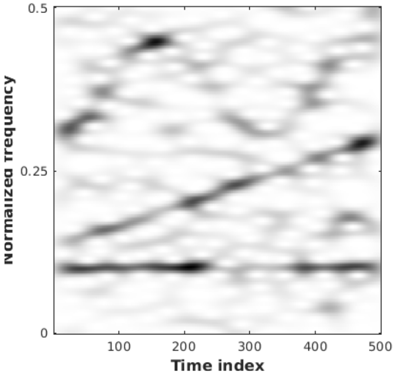
- with $j^2 = -1$ and $\hat{x}$ the Fourier transform of x.
- However: Signal frequency can vary over time.
- $\Rightarrow$ Short-Time Fourier Transform (STFT).

$$F_x^h(t,\omega) = \int_{\mathbb{R}} x(u) h(t-u)^*\, \mathrm{e}^{-j\omega u}\, \mathrm{d}u,$$

- with z^* the complex conjugate of z.
- h the analysis window providing temporal information.

Introduction: Time-frequency tools





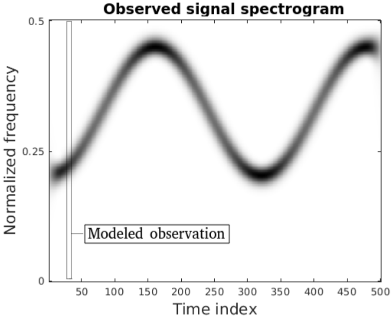
Spectrograms of multi-components signals

- Provide time and frequency content.
- Signals components appear as ridges in the time frequency plane.
- Interest : ridges position.

Observation model

Our model

$$s_{n,m} | \bar{m}_n \sim g\left(m - \bar{m}_n\right).$$



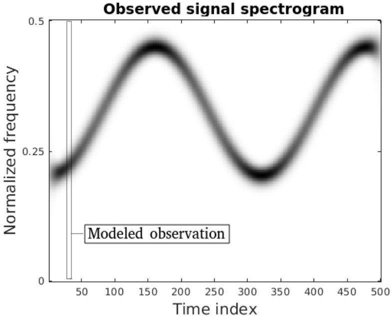
- Model 1D signals $\boldsymbol{s}_n$.
- Known data distribution.

- m : frequency in $[0, M-1]$.
- $\bar{m}_n$: ridge position in the n-th time bin.
- $g(m) = \frac{2\sqrt{\pi}L}{M} \, \mathrm{e}^{-\left(\frac{2\pi mL}{M}\right)^2}$.
- L: time spread of the analysis window.

Observation model

Our model

$$s_{n,m} | \bar{m}_n \sim g\left(m - \bar{m}_n\right).$$



Limitations

- Simple model.
- Computationally attractive.
- Presence of noise neglected.
- Assumes for the presence of a single component.

Observation model

Limitations

- Lack of generality of the postulated model.
- Discrepancies with noisy observations.
- Or in the presence of multiple components.
- Inefficiency of Maximum Likelihood Estimation (MLE).

Proposed approach

- Estimation performance does not only depend on the model quality.
- Modification of the similarity measure[5].

[1]Q. Legros, S. McLaughlin,Y. Altmann, S. Meignen and M. E. Davies. Robust depth imaging in adverse scenarios using single-photon Lidar and beta-divergences, 2020.

Estimation strategy

Note that

- Performing MLE $\Leftrightarrow$ minimizing Kullback-Leibler Divergence (KLD) between model and observations.
- KLD not suitable when the postulated model is inaccurate.
- Implies model mismatch.
  - In the presence of external spurious noise.
  - When observing multicomponent signals.

Alternative variational objective

- KLD replaced by the Rényi and β divergences.
- Allow respectively for mode seeking character and robustness.

Estimation strategy

Alternative inference

- Variational inference based on alternative divergences.
- Need for the divergences cross entropy[2].

Cross entropy

- For the β-divergence (β-d), $\beta > 0$

$$\mathrm{CE}_{\beta}(\bar{m}_n) = -\frac{1+\beta}{\beta}\sum_{m} p(s_{n,m}|\bar{m}_n)^{\beta} + \int p(m|\bar{m}_n)^{1+\beta}\mathrm{d}m. \tag{10}$$

- For the Rényi divergence (R-d), $\alpha > 0, \alpha \neq 1$

$$\mathrm{CE}_{\alpha}(\bar{m}_n) = \frac{1}{\alpha-1}\log\left(\sum_{m} s_{n,m}^{\alpha} p(s_{n,m}|\bar{m}_n)^{1+\alpha}\right). \tag{11}$$

[2] F. Futami, I. Sato, M. Sugiyama. Variational Inference based on Robust Divergences, 2018.

Estimation strategy

Pseudo-Bayesian estimation

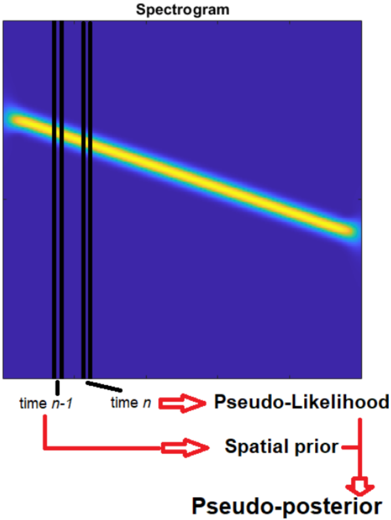
- Approximate posterior distribution obtained by maximizing the evidence lower-bound (ELBO)[6]

$$p(\bar{m}_n|\boldsymbol{s}_n) \propto e^{-\mathrm{MCE}(\bar{m}_n)} p(\bar{m}_n). \qquad (12)$$

- Spatial prior model $p(\bar{m}_n)$ discussed hereafter.
- Plug cross entropy for alternative objectives.
- Ridge position estimated by minimum mean squared error(MMSE).

[2] F. Futami, I. Sato, M. Sugiyama. Variational Inference based on Robust Divergences, 2018.

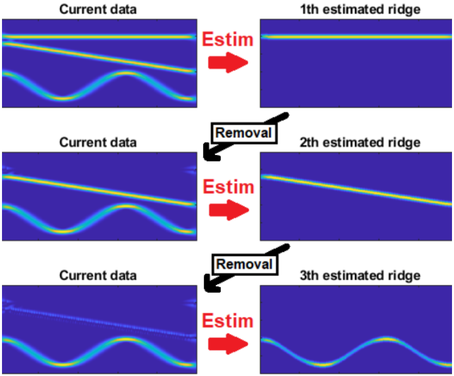
Estimation strategy



Online estimation

- Ridge extraction.
- Iterating on time axis.
- Sequential propagation of the information.
- Spatial prior : Gaussian random walk.
- Complexity: Variational inference.
- Accuracy: Backward correction.

Estimation strategy



Sequential demodulation

- Extraction of a ridge.
- Update of the data.
- Removing energy associated with the estimated component.
- Stopping criterion: method of choice[7].

[3] V. Sucic and N. Saulig and B. Boashash. Estimating the number of components of a multicomponent nonstationary signal using the short-term time-frequency Rényi entropy, 2011.

Results

Numerical experiments

- Reconstruction performance of a MCS.
- Reconstruction quality factor: $\mathrm{RQF} = 10\log_{10}\left(\frac{||x||^2}{||x-\hat{x}||^2}\right)$.
- Assessment: Component-wise RQF.

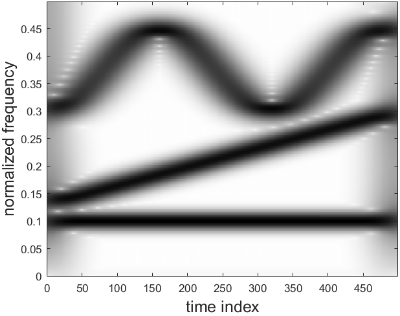


Figure: Spectrogram of the analyzed multicomponent signal.

Three components

- Sinusoidal frequency modulated (FM).
- Linear chirp.
- Sinusoid.

Results

Table: RQF of each components (averaged over 100 realizations) for the different competing approaches for a SNR = 10 dB.

	Sinusoid	Linear chirp	Sin. FM chirp	Average
Brevdo	16.10	**15.46**	2.86	11.47
Brevdo-Synchrosqueezing	16.43	15.34	5.24	12.34
Proposed β-d, $\beta = 0.5$	**16.71**	15.22	9.13	13.69
Proposed β-d, $\beta = 0.8$	16.45	14.92	5.49	12.29
Proposed-KLD	2.46	2.65	1.18	2.10
Proposed R-d, $\alpha = 0.5$	16.59	15.24	**9.57**	**13.80**
Proposed R-d, $\alpha = 0.8$	15.44	15.22	7.84	12.83

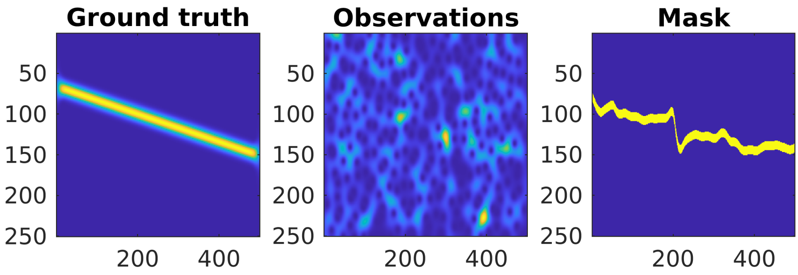
Numerical experiments

- Our method obtains the best averaged RQF using R-d ($\alpha = 0.5$).
- Efficient recovery of the sinusoidally FM chirp.
- Alternative divergences circumvent the lack of generality of our model.

Results - examples 1

Numerical experiments

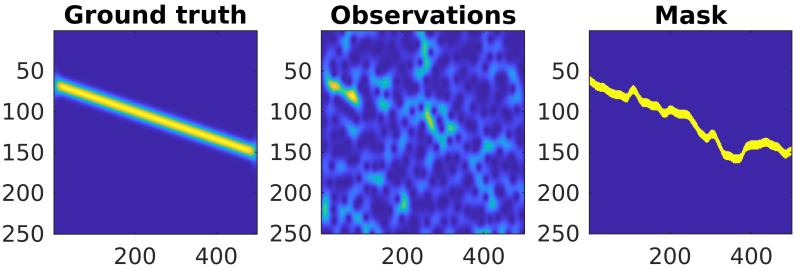
- Single linear chirp component.
- SNR = -15dB.
- Rényi divergence, $\alpha = 0.2$.



Results - examples 2

Numerical experiments

- Single linear chirp component.
- SNR = -15dB.
- Rényi divergence, $\alpha = 0.2$.



Conclusions and perspectives

Conclusions

- A novel pseudo-Bayesian estimation procedure to demodulate MCS.
- Inaccurate model.
- Performances improved through estimation strategy.
- Adaptive prior model.

Plan

1 Introduction

2 In practice

3 Lidar analysis

4 Pseudo-Bayesian approach

5 Conclusion

Conclusions

Conclusions

- Bayesian approaches can be adapted to a wide range of problem.
- Do not require a large amount of data.
- Allow for interpretation and confidence estimation of predictions.
- Not necessarily computationally expensive.
- Different ways to address a problem : models, priors, estimation strategy...

Thanks for your attention !

legros.quentin2@hotmail.fr