école
normale
supérieure
paris—saclay

Blind denoising with random greedy pursuits

Victor Letzelter & Clautrier Mathis

Machine Learning for Time Series
mclautrier98@gmail.com
letzelter.victor@hotmail.fr

Master MVA

April 5, 2022

Introduction

Data analysis

Methods

Results

Conclusion

Introduction

Denoising framework:

$$y = \underbrace{\tilde{y}}_{clean} + \underbrace{w}_{noisy}$$

Given a dictionary $\Phi = (\phi_1, ..., \phi_M) \in \mathbb{R}^{N \times M}$ select an activation vector $\alpha \in \mathbb{R}^M$ that minimizes the quantity

$$||\tilde{y} - \Phi\alpha||$$

under the constraint

$$||\alpha||_0 \ll k$$

where $k \in \mathbb{R}_+^*$

☞ How to prevent overfitting with the noise w ?

Dataset

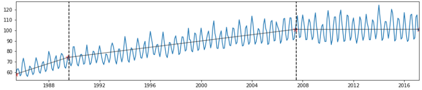

Figure: Initial signal and its (detected) rupture points.

Electrical production (in Australia?) taken from Kaggle.

Two rupture points.

Stationarity? Dicker-Fuller test : 0.094

$\Rightarrow$ Let's remove the trend.

Dataset

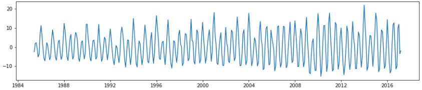

Figure: Obtained signal by removing a fitted piecewise linear trend.

Stationarity? Dicker-Fuller test : $p = 1e - 7$.

The signal seems periodic.

$\Rightarrow$ Let's investigate the periods.

Dataset

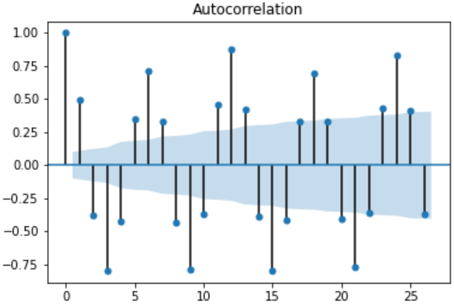

Figure: Autocorrelation of the signal after removing the trend.

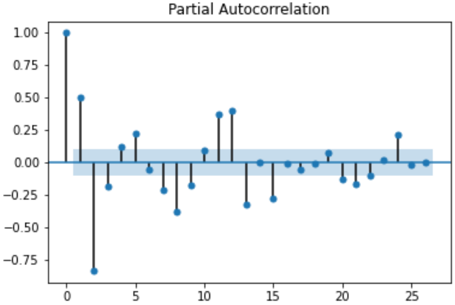

Figure: Partial autocorrelation of the signal after removing the trend.

Dataset

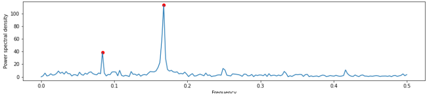

Figure: Periodogram of the signal after removing the trend.

Two frequencies (6 months and 1 year), consistent with the problem.

$\Rightarrow$ Let's obtain the trend by moving average ($T = 12$).

Dataset

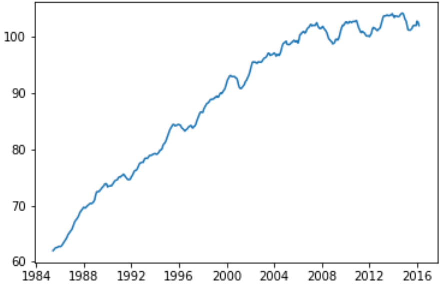

Figure: Obtained signal by moving average.

Similar trend to the piecewise linear one.

$\Longrightarrow$ We'll use this signal for the experiments.

Color noise addition on the data

Definition

A **color noise** is a real valued stationary Gaussian process $\left(Z(t)\right)_{t\in\mathbb{R}}$, whose power spectral density is $S(f) = \frac{1}{f^{\alpha}} \quad (\alpha \in \mathbb{Z})$

Algorithm for color noise generation

Calculation of $\boldsymbol{S} = \left(S(f_1), \ldots S(f_n)\right)$ with $S(f) \triangleq \frac{1}{f^{\alpha}}$

Gaussian vector: $\boldsymbol{Y} = \left(Y_1, \ldots, Y_n\right)^t$

Realisation of the Gaussian Process: $\boldsymbol{Z} = \mathrm{IFT}(\sqrt{\boldsymbol{S}} \cdot \times \mathrm{DFT}(\boldsymbol{Y}))$

Denoising with Matching Pursuit

Dictionnary used : Modified Cosine Transform of window functions with L in [32,64,128] ($\sim$ 500.000 atoms).

Stochastic matching pursuit [4]

Matching pursuit is run J times to average the obtained results.

First run with the whole dictionary then with $\Phi_n \subset \Phi$ a random sub-dictionary.

Denoising with Matching Pursuit

Stochastic orthogonal matching pursuit [1]

Orthogonal matching pursuit is run J times to average the obtained results.

At each run the atoms are not selected by taking the maximum of the scalar product but by sampling from $\exp\left(c^2 D^T r_{t-1} / \sigma^2\right) \Rightarrow$ use of the Gumbel-max trick.

Gumbel-max trick

Let $p = (p_i)_i = \frac{\exp(\alpha_i)}{\sum_i \exp(\alpha_i)}$ a categorical distribution and $g = (g_i)_i \sim G(0)$ a sample from the standard Gumbel distribution.

Then one can sample from p having only access to $\alpha = (\alpha_i)_i$ by taking $i^* = \arg\max \alpha + g$.

Denoising with Matching Pursuit

To properly compare BIRDS to the different algorithms (i.e regular Matching Pursuit, SOMP, SMP) we used an oracle version of these algorithms to have a stopping criterion.

Oracle criterion

Let's $\tilde{f}_t$ be the reconstruction at step t, f be the denoised signal and ϵ be a given threshold.

If $DTW\left(\tilde{f}_t, f\right) < \epsilon$ then stop.

école normale supérieure paris-saclay

Denoising with Shrinkage [2]

DWT - Shrinkage - IDWT

f signal to be transformed. Subsignal reconstructed at level j:

$$f_j(t) = \sum_k W_f(j,k)\psi^*\left(2^{-j}t - k\right)$$

where $W_f(j,k) = \int_{-\infty}^{+\infty} f(t)\psi_{j,k}^*(t)dt$ with $\psi_{j,k}(t) = 2^{-j/2}\psi\left(2^{-j}t - k\right)$

Denoising with Shrinkage [2]

Shrinkage functions (of the DWT coefficients)

Hard shrinkage function:

$$\delta_\lambda^H(x) = \begin{cases} 0 & |x| \leq \lambda \\ x & |x| > \lambda \end{cases}$$

Soft shrinkage function:

$$\delta_\lambda^S(x) = \begin{cases} 0 & |x| \leq \lambda \\ x - \lambda & x > \lambda \\ \lambda - x & x < -\lambda \end{cases}$$

Stopping condition of the BIRD algorithm

Assumptions on the dictionary [3]

1. The projections of the noise on its atoms are distributed according to a zero mean Gaussian distribution (GD).
2. The distribution of the projections of the informative part has a heavier tail than the GD.

Stopping condition of the algorithm

$$\lambda_\Phi(R^k f) \triangleq \sup_{\phi \in \Phi} \frac{|\langle R^k f, \phi \rangle|}{\|R^k f\|_2} \leq \Lambda_W(\Phi, p) \tag{1}$$

where $\Lambda_W(\Phi, p) = \frac{\sqrt{2}}{N}\sqrt{1 - \frac{2}{\pi}}\,\mathrm{erfinv}\left((1-p)^{\frac{1}{M}}\right)$

Stopping condition of the BIRD algorithm

Assumptions on the dictionary [3]

1. The projections of the noise on its atoms are distributed according to a zero mean Gaussian distribution (GD).

Hypothesis 1. checking

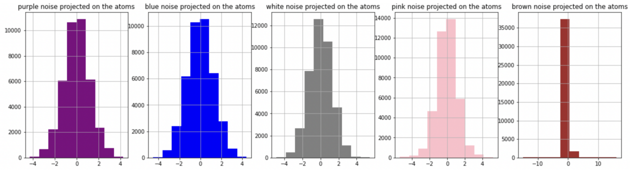

Figure: Projection of the noise generated in the MDCT dictionary école normale supérieure paris-saclay

Stopping condition of the BIRD algorithm

Assumptions on the dictionary [3]

2. The distribution of the projections of the informative part has a heavier tail than the GD.

Hypothesis 2. checking

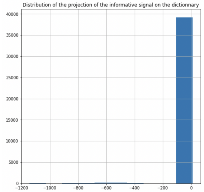

Figure: Distribution of the projection of the informative signal on the dictionary

Denoising results

Noise Color	Purple	Blue	White	Pink	Brown
DTW w. clean signal	322.23	353.09	325.01	311.07	155.42

Figure: DTW distances with the original sigal for each algorithms and noise types

Noise color vs. Algorithm	BIRD	SMP	SOMP	MP	Shrinksoft	Shrinkhard
Purple	158.52	206.15	322.23	196.25	**120.96**	255.66
Blue	182.45	273.24	353.09	203.89	**123.32**	260.58
White	190.17	222.04	325.01	226.91	**110.61**	266.79
Pink	144.95	323.83	311.07	198.25	**120.19**	242.15
Brown	159.92	310.33	157.42	197.52	**121.08**	128.79

Figure: DTW distances with the original clean signal for each algorithms and noise types

About the stopping condition of BIRD

Noise color vs. Algorithm	Avg. Index min DTW	Avg. Index min DTW noisy	Avg. Index stop
Purple ($\alpha = -2$)	877.58 ± 98.79	919.24 ± 76.18	562.44 ± 77.71
Blue ($\alpha = -1$)	673.73 ± 133.42	921.49 ± 76.97	578.97 ± 78.00
White ($\alpha = 0$)	514.08 ± 84.26	950.38 ± 57.56	619.12 ± 88.84
Pink ($\alpha = 1$)	491.95 ± 89.00	948.11 ± 56.52	644.61 ± 92.14
Brown ($\alpha = 2$)	507.41 ± 86.88	943.99 ± 49.43	629.3 ± 84.83

Figure: Index values for different noise types

Conclusion

☞ Choice of the dictionary

☞ Stopping condition not really dependant on the noise type ; an issue when dealing with pink and Brownian noise.

Thanks for your attention

References

A. Ejaz, E. Ollila, and V. Koivunen.
Randomized simultaneous orthogonal matching pursuit.
In *2015 23rd European Signal Processing Conference (EUSIPCO)*, pages 704–708. IEEE, 2015.

B. Kozłowski.
Time series denoising with wavelet transform.
Journal of Telecommunications and Information Technology, pages 91–95, 2005.

M. Moussallam, A. Gramfort, L. Daudet, and G. Richard.
Blind denoising with random greedy pursuits.
IEEE Signal Processing Letters, 21(11):1341–1345, 2014.

T. Peel, V. Emiya, L. Ralaivola, and S. Anthoine.
Matching pursuit with stochastic selection.
In *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pages 879–883. IEEE, 2012.