BERNOULLI-GAUSSIAN DECONVOLUTION IN NON-GAUSSIAN NOISE, CONTRIBUTION OF WAVELET DECOMPOSITION

H.Rousseau and P.Duvaut
E.T.I.S. - E.N.S.E.A., 6, avenue du Ponceau, 95014 CERGY Cedex
Tel : (33 1) 30 73 66 45 - Fax : (33 1) 30 73 66 27
e-mail : rousseau@ensea.fr

ABSTRACT

We introduce a method to restore Bernoulli-Gaussian processes immerged in a non-gaussian noise. It uses wavelet decomposition to "gaussianize" the noise. The convergence, after wavelet projection, of some non-gaussian noise to a gaussian noise quantifies the quality of the "gaussianization" effect of the wavelet. This property is used to apply a Bernoulli-Gaussian algorithm at each scale of wavelet decomposition. After, we use a fusion strategy to merge all results. We obtain also a new deconvolution algorithm which is very performant, for all satistical noises, when the noise variance is not well estimated. When the noise variance is correctly estimated, it improves the classical Bernoulli-Gaussian algorithm for strongly non-Gaussian noises.

1 INTRODUCTION

We use the gaussianization property of wavelet multiresolution in a deconvolution problem like Bernoulli-Gaussian deconvolution. In this classical problem (often encountered specially for stripes extraction of the periodogram ([7]) or, in seismology, the detection-estimation of pulses ([5])), the noise is considered as gaussian noise. The model of signal $y(t)$ to be analysed is composed of Bernoulli-Gaussian processes $x(t)$ filtered by a known $a(t)$ convolution nucleus which is embedded in a Non-Gaussian noise $b(t)$:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$$

where $\mathbf{A}$ is also the convoluting matrix of vector $\mathbf{a}$.

Tanks to the property of the limit central theorem, filtering changes a non-gaussian noise into a "more" gaussian noise. Therefore, we can quantify this quality for any filter. Wavelet projection can be viewed as filter's bank with the main advantage of its reconstruction property which ensure the filtering action does not reduce signal's informations.

Therefore, wavelet decomposition plus deconvolution provides several estimated signals which can be mixed together to build a new deconvoluted signal by fusion strategy. We have implemented a fusion algorithm of impulse detection and tracking from scale to scale (see [8]). As the deconvolution algorithm uses the hyperparmeter of noise variance, we test our method in two cases: first, when the noise variance estimation is correctly estimated in the deconvolution algorithm and second, when the noise variance is not well estimated.

2 GAUSSIANIZATION BY WAVELET DECOMPOSITION

2.1 Wavelet decomposition

For our multiresolution analysis, we use the algorithm with holes (or à trous Algorithm) ([3]). This algorithm induces no decimation, so, wavelet decomposition conserves, at each scale, the stripes localisation. It is very useful for tracking stripes fromscale to scale (scale number i will correspond to the scale 2^i). The signal transformation from scale i to scale $(i+1)$ corresponds to a filtering action. If $\mathbf{s_i}$ and $\mathbf{d_i}$ are respectively the average decomposition and the wavelet decomposition, the filtering action is summed up by:

$$\mathbf{s_{i+1}} = \mathbf{F_{(i)}}\mathbf{s_i} \text{ and } \mathbf{d_{i+1}} = \mathbf{G_{(i)}}\mathbf{s_i}$$

where $\mathbf{F_{(i)}}$ and $\mathbf{G_{(i)}}$ are convoluting matrix of the average and wavelet coefficients ($\mathbf{f_i}$ and $\mathbf{g_i}$) at scale i (In filter $\mathbf{f_i}$, there are $2^{i-1}-1$ zeros between each coefficients of $\mathbf{f_1}$). Finally, each scale decomposition can be summarized by the followig equation:

$$\mathbf{d_i} = \mathbf{G_{(i-1)}}\mathbf{F_{(i-2)}}\mathbf{F_{(i-3)}}...\mathbf{F_{(1)}}\mathbf{s}$$

We note $\mathbf{H_i}$ the final matrix convolution of the decomposition at scale i.

2.2 Generalisation of limit central theorem

Applying a FIR filter on a process "gaussianizes" it, thanks to the Central Limit Theorem ([9]). For any filter $\mathbf{h}$, we can find two values which are characteristic of the gaussianization effect of filtering. Let us suppose that $\mathbf{x}$ is white noise vector the cumulants of which are $0, \sigma^2, \kappa_3, \kappa_4$ and so on, its second characteristic function is also,

$$\Psi(t) = log(E(exp(itx))) = 1 - \frac{1}{2}\sigma^2 t^2 - i\frac{1}{6}\kappa_3 t^3 + \frac{1}{24}\kappa_4 t^4 + ...$$

If $\mathbf{p}$ is the filtered noise, $p(n) = \sum_{k=1}^{N} h(n-k)x(k)$, its second characteristic function $\Phi_h(t)$ verifies

$$\begin{aligned} log(\Phi_h(t)) &= -\frac{1}{2}\sigma^2\mathcal{M}_2^{(h)}t^2 - i\frac{1}{6}\kappa_3\mathcal{M}_3^{(h)}t^3 \\ &+ \frac{1}{24}\kappa_4\mathcal{M}_4^{(h)}t^4 + \frac{1}{120}\kappa_5\mathcal{M}_5^{(h)}t^4 + ... \end{aligned}$$

where $\mathcal{M}_j^{(h)} = \sum_{k=1}^{N} h^j(k)$ will be called j^{th} moment of the filter. To simplify results, we take a normalised filter $\mathcal{M}_2^{(h)} = 1$ and a normalised noise ($m_x = 0, \sigma_x = 1$). $\mathbf{p}$ is close to a gaussian noise $\mathcal{N}(0,1)$ only if the two following values are close to zero. The third and four moments of the filter can be, respectively, considered like asymetrical and symetrical "gaussianization" values of the filter $\mathbf{h}$ and can quantify the quality of the "gaussianization" effect. Moreover, these two values are quite sufficient because if there are close to zero, other moments have to be closer to zero. The third moment can reach zero perfectly while the minimum value of $\mathcal{M}_4^{(h)}$ is 1/N coresponding to the best achievable gaussianization with a N taps filter. It appears that the best filters for "gaussianisation" (N fixed) are mean average filters with all coefficients equal to $+/-\frac{1}{\sqrt{N}}$.

As, from the lower scale to the upper scale of multiresolution analysis, taps filter increases and filter coefficients draw near to best filters conditions, non-gaussian noise is getting closer and closer to a gaussian noise. Remark: for symetrical probability density functions (pdf) ($\kappa_3 = 0$), asymetrical "gaussianisation" value has no effect of noise' "gaussianization".

2.3 Choice of Wavelet-basis

We calculate, for different wavelet basis as the Daubechies' ([1]) (For the record, taps wavelet and average filters of the order N Daubechies' wavelet are equal to 2*N) and the quadratique spline used by Mallat in ([2]) (taps wavelet and average filters are equal to 2 and 4), the "gaussianization" values.

	$\mathcal{M}_3$			
Wavelet	1st proj	2nd proj	3rd proj	4th proj
Daub's(N=4)	.3690	0.2152	0.1558	0.1112
Daub's(N=8)	0.0982	0.0452	0.0272	0.0188
quad. spline	0	0	0	0
	$\mathcal{M}_4$			
Wavelet	1st proj	2nd proj	3rd proj	4th proj
Daub's(N=4)	0.4866	0.3032	0.1467	0.0737
Daub's(N=8)	0.3421	0.1433	0.0762	0.0381
quad. spline	0.5000	0.2500	0.1250	0.0625

Table 1: third and four moments of 3 wavelet basis

As the asymetrical "gaussianisation" value of quadratique spline is null and the symetrical "gaussianisation" value has some comparable results to other studied wavelet-basis (see Table 1), we choice the quadratique spline to our multiresolution analysis. For a wide class of pdf, it appears to be the best trade off between "gaussianization" effect and length of wavelet coefficients.

2.4 Evaluation with gaussianity test

We verify the above property with two standard gaussian tests: the Skewness and Kurtosis Statistical tests ([4]). Theses two tests are null for gaussian noise. For the wide class of pdf studied and after projections on wavelet basis, the convergence to the gaussian pdf is fast (see table 2) which confirms the theorical calculus and the linearity of normalised skewness and kurtosis with, respectively, regard to $\mathcal{M}_3$ and $\mathcal{M}_4$.

	normalised skewness (initial = 3)			
Wavelet	1st proj	2nd proj	3rd proj	4th proj
Daub's(N=4)	1.0116	0.6797	0.4784	0.3210
Daub's(N=8)	0.2633	0.1018	0.0650	0.0067
quad. spline	0.0012	-0.0077	0.0079	0.0252
	normalised kurtosis (initial = 12)			
Wavelet	1st proj	2nd proj	3rd proj	4th proj
Daub's(N=4)	5.7624	2.8335	1.4181	0.7475
Daub's(N=8)	4.0951	2.0201	1.0082	0.4689
quad. spline	5.5995	2.8986	1.3879	0.7441

Table 2: skewness and kurtosis tests of χ_2 noise after wavelet projection

3 BERNOULLI-GAUSSIAN DECONVOLUTION

3.1 Bernoulli-Gaussian deconvolution at different scales

The classical Bernoulli-Gaussian deconvolution algorithm, BGG ([5],[6]), assumes a white additive gaussian noise (the covariance matrix $\mathbf{\Gamma_b}$ is taken as an hyperparameter and considered as a diagonal matrix in the original algorithm (when we do not estimate the noise variance, this hyperparameter is taken equal to 0dB)). The signal $\mathbf{x}$ is modeled as the Hadamard product of a Bernoulli vector $\mathbf{q}$ (the density parameter of which, λ, is considered as an hyperparameter) and of a gaussian vector $\mathbf{a}$. BGG is splitted into 2 steps. First, we find the Bernoulli vector which maximizes the a posteriori marginal density of the vector $\mathbf{q}$. Second, using the precedent estimation $\hat{\mathbf{q}}$, we estimate $\hat{\mathbf{x}}$ by maximising the joint likelihood probability of $\mathbf{x}$.

Assuming $\mathbf{R} = \sigma^2\mathbf{diag(q)}$, $\mathbf{\Gamma} = \mathbf{O^T R O} + \mathbf{\Gamma_b}$ and Q the number of stripes counted in $\mathbf{q}$, these two steps are summarized by the two expressions below:

$$\mathbf{\hat{q} = Arg_q\{Max\{-y^T\Gamma^{-1}y - 2Qln(\frac{1-\lambda}{\lambda}) - ln(det(\Gamma))\}\}}$$

$$\mathbf{\hat{x} = \hat{R}O^T\hat{\Gamma}^{-1}y}$$

Because of the limit central argument, we may assume a colored gaussian noise, $\mathbf{\Gamma_b}$ can not be considered not as a diagonal matrix, but as $\mathbf{\Gamma_b = \sigma^2 H_i^T H_i}$.

Unfortunately, the filtering matrix $\mathbf{H_i}$ is very ill-conditionned, so we forced a Toeplitz structure on matrix borders to improve condition number. Another problem stems from the nature of the $\mathbf{\Gamma_b}$ matrix which smooths too much at upper scales (the BGG performances are also extremely bad). We solve this problem as the BGG algorithm is quite robust to colorness of noise by not considering colorness from scale 5th to upper scales, so for these scales $\mathbf{\Gamma_b} = \sigma^2\mathbf{Id}$.

Figures 1 and 2, show an example of BGG deconvolution at each scale. In the example, we estimate variance noise before deconvolution (the variance error estimation do not pass 4 dB).

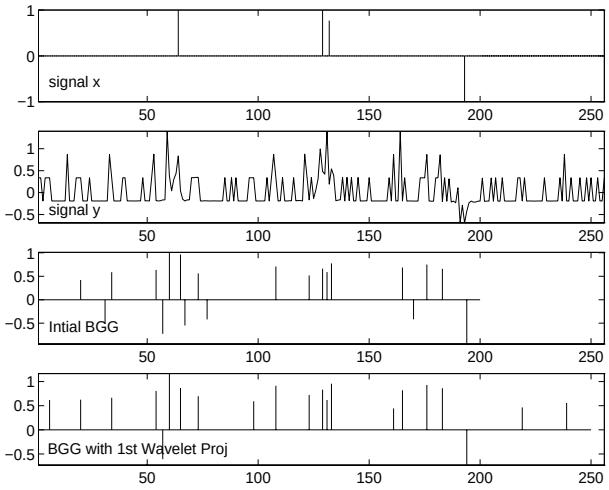

Figure 1: Deconvolution (poisson noise, SNR=-10dB)

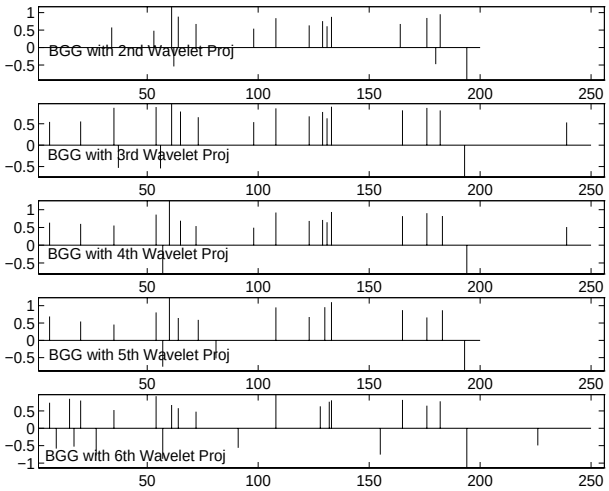

Figure 2: Deconvolution (poisson noise, SNR=-10dB)

3.2 Fusion of Deconvolution Results

Therefore, wavelet decomposition after Bernoulli-Gaussian deconvolution provides several estimated signals which can be mixed together to make another deconvolution algorithm by fusion strategies.

A first fusion strategy considers only the detection-decision aspect. It uses results of BGG deconvolution of a few scales projection. In the fusion result, there is a stripe if that stripe appears in all BGG deconvolution considered, the estimated amplitude is then the mean of all estimated amplitudes. The main problem of this strategy is the stripe delocalisation of the deconvolution result at each scale. The localisation error is all the more important than the scale is high, so, this strategy can not used the upper scale.

Another fusion strategy has been developped. It is based on tracking stripes from upper scale to the lower scale. Hence, we conserve a stripe found in BGG result at scale j if there is a stripe at scale $(j+1)$ the index and the amplitude of which are, respectively, close to the index and the amplitude of the stripe at scale j.

By inverse recurrence, we follow stripes from scale to scale and finally we decide wether a stripe exists or not if we can follow it from the upper scale to the lower one. For concerned lectors, this algorithm is more detailed in [8]. This fusion BGG method will be called Wavelet Bernoulli Gaussian WBGG.

4 RESULTS

4.1 Signal test

To appraise effectively the perfomance of the method, we test our method on a signal of 257 points immerged in three different noises (a gaussian noise and two non-gaussian noise, a $\chi 2$ noise and a poisson noise).

Our results were obtained by the means of 50 trials for each configuration. The signal test is composed of four stripes (see Figure 1) convolved withthe square of a cardinal sinus as convolution nucleus:

$$a(k) = 256 * sinc^2(\frac{\pi}{\eta}(k))$$

where $k \in [-128, 128] \, and \, \eta = 12$).

4.2 Performance

4.2.1 Performance criteria

We choose different performance criteria to value our method. They may be splitted into two classes:
- detection performance criteria
  - false alarm probability,
  - true detection probability (tdp) (in fact, we take 1-tdp),
  - number of detected stripes
- estimation performance criteria
  - amplitude estimation error of good detection,
  - power of false detection stripes,
  - localization error,
  - maximum amplitude of false detection.

All criteria have to be as low as possible except for the number of detected stripes criterium which has to be as close as possible to the number of stripes in signal test.

4.2.2 Results with correct noise estimation

In the first example (Figure 1 and 2), we can see that deconvolution perfomance at each scale are similar. It is due to two factors: the gaussianization of noise (which favours upper scale BGG deconvolution) and the spreading out of convolution nucleus (which disadvantages upper scale BGG deconvolution). But BGG at each scale combined with tracking stripe fusion improves the performance compared to classical BGG as ROC curves (Figure 3) show it. In fact, when SNR is too low, BGG detects a lot of false alarms which can be reduced with using WBBG while good detections are conserved. Other studied fusions are not performant because they reduce too much the best detection performance. On the other side, strategy fusions don't seem to have visual effects for our estimation performance's criteria.

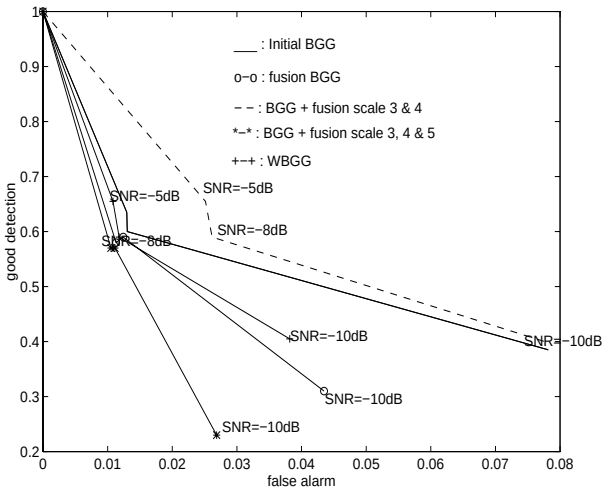

Figure 3: ROC curves (poisson noise)

4.2.3 Results with bad noise estimation

Sometimes, it is quite difficult to have a good approximation of variance noise. In these conditions, as we can see in Figure 4, WBGG algorithm can be viewed as a regularization the BGG deconvolution. In this case, power of false detection stripes and maximum amplitude of false detection criteria reduces, respectively with a factor of 5 and 2 for $\chi 2$ noise (SNR= 10 dB).

5 CONCLUSION

The use of wavelet decomposition, for a Bernoulli-Gaussian deconvolution, when the noise variance is unknown, improves, significantly, the performance compared to the initial BGG algorithm. When noise variance estimation is correctly etimated in the deconvolution algorithm, performance are, of course, improved. In this case, WBGG and BGG performance are nearly the same excepted for strongly non gaussian noises like poisson noise (in this case, WBGG results procures less false alarms than BGG results).

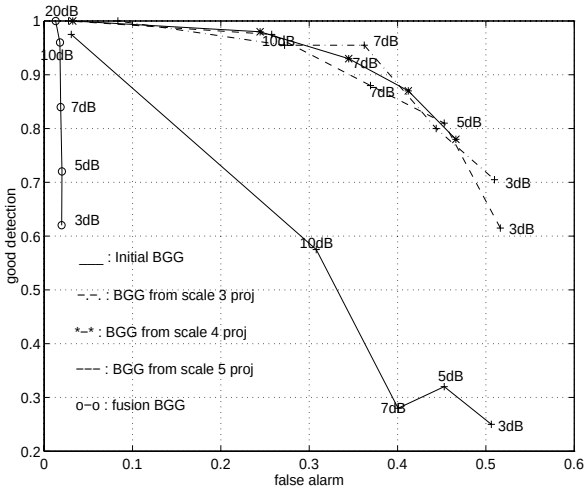

Figure 4: ROC curves (poisson noise)

References

[1] **I.Daubechies (1992)**: "Ten lectures on wavelets",SIAM, Philadelphia.

[2] **S.G.Mallat and S.Zhong**:
"Characterization of Signals from Multiscale Edges",
IEEE Trans on P.A.M.I.Math. Soc., Vol. 14, N^o 7, pp 710-732, 1992.

[3] **M.J.Shensa**:
"The Discrete Wavelet Transform: Wedding the À Trous and Mallat Algorithms"
IEEE Trans on Signal Proc., Vol. 40, N^o 10, pp 2464-2482, 1992.

[4] **C.Coroyer, C.Jorand et P.Duvaut (1994)**: "ROC Curves of Skewness and Kurtosis Statistical Tests : Application to Textures", proceedings of Eusipco, Edimburg.

[5] **J.J.Kormylo and J.M.Mendel**:
"Maximum Likelihood Detection and Estimation of Bernoulli-Gaussian processes",
IEEE Trans on Inform. Theo., vol 28, N^o 3, 1982.

[6] **F. Champagnat and J. Idier**:
"un nouvel algorithme de déconvolution impulsionnelle avec prise en compte de saturations",
Proceeding GRETSI, 1993.

[7] **P.Duvaut and F.Dublanchet**:
"Une methode d'analyse de sinusoides complexes par deconvolution du periodogramme nommée "Expulse" ",
Traitement du Signal, vol 12, N^o 3, 1995.

[8] **H.Rousseau et P.Duvaut**:
"Bernoulli-Gaussian Deconvolution in Non-Gaussian Noise from multiscale edges",
proceeding of TFTS, Paris, 1996.

[9] **M.G.Kendall, A.Stuart and J.Keith Ord (1987)**: "Kendall's Advanced Theory of Statistics", Vol 1, Griffin, London.