Poisson Intensity Estimation Based on Wavelet Domain Hypothesis Testing

Bo Zhang, Jalal M. Fadili, Jean-Luc Starck

Poisson Intensity Estimation Based on Wavelet Domain Hypothesis Testing

Bo Zhang*, Jalal M. Fadili and J.-L. Starck

Running head: Poisson Intensity Estimation

Bo Zhang is with the Quantitative Image Analysis Unit URA CNRS 2582 of Pasteur Institute, 75015 Paris France
E-mail: bzhang@pasteur.fr

Dr. M.J. Fadili is with the Image Processing Group GREYC CNRS UMR 6072 14050 Caen Cedex France.
E-mail: Jalal.Fadili@greyc.ensicaen.fr

J.-L.Starck is with the CEA-Saclay, DAPNIA/SEDI-SAP, Service d'Astrophysique, F-91191 Gif sur Yvette, France.
E-mail: jstarck@cea.fr

Abstract

In this paper, we present the estimation of Poisson intensity based on hypothesis testing in the wavelet domain for any dimensional data. The testing framework for wavelet-based Poisson intensity estimation was first introduced by Kolaczyk, where a thresholding estimator, which realizes the hypothesis testing, is derived for Haar wavelet coefficients. Here we propose for the same wavelet a new thresholding estimator which is based on Fisher's normal approximation. Furthermore, we have demonstrated that non-normalized biorthogonal Haar coefficients converge in distribution to non-normalized Haar coefficients as the scale increases. This allows us to directly apply the threshold in the biorthogonal Haar domain. Therefore we gain, by using this more regular wavelet, a reconstruction with less artifacts. Simulations show that on a wide range of intensity types, the proposed threshold combined with undecimated biorthogonal Haar transform gives one of the best estimation result compared with existing estimators of various kinds. Finally, potential applicability of our approach is illustrated on astronomical data.

Index Terms

Poisson intensity estimation, wavelet hypothesis testing, Fisher approximation, biorthogonal Haar wavelets

I. Introduction

The ability to restore the underlying intensity from an inhomogeneous Poisson process is crucial for many applications. We observe a discrete signal of counts $\mathbf{v} = (v_i)_{i \in I}$ where I is the index set. Each count v_i can be thought of as coming independently from a Poisson distribution with a mean λ_i

$$v_i \sim \mathcal{P}(\lambda_i)$$

We can then decompose the observed count v_i into two parts, i.e.

$$v_i = \lambda_i + b_i, \quad i \in I$$

where b_i is the Poisson noise, and the mean vector $\Lambda = (\lambda_i)_{i \in I}$ is the source intensity which needs to be estimated from $\mathbf{v}$.

A host of estimation methods have been proposed in the literature. A common solution is to use a variance stabilizing transform (VST), which "Gaussianizes" the Poisson noise before applying the standard wavelet thresholding denoising [1][2] on the transformed signal. For example, Donoho [3] and Fryźlewicz and Nason [4] proposed respectively the Anscombe transform [5] and the Fisz transform [6]. Rather than using a nonlinear VST, direct wavelet filtering has been studied by Nowak and Baraniuk [7] and Antoniadis and Sapatinas [8]. The latter concerns actually a wider distribution family, but it is essentially equivalent to the former in the Poisson case. The resulting filter can be considered as a data-adaptive wavelet domain Wiener filter. The state-of-the-art methods are Bayesian approaches [9][10][11][12] as reported by Besbeas et al. [13]. In effect a great part of above approaches are based on Haar wavelet transform until recently, Jansen [14] introduced conditional variance stabilization, which generalizes the idea of [4] and is applicable to any family of wavelet transforms. A Bayesian scheme was also derived within this framework, which can be deemed as an extension of [9][10] and [11]. Other technique as maximum penalized

likelihood estimation [15][16] can also be considered Bayesian, since the penalization term implicitly introduces a certain prior on the underlying intensity. In general, Bayesian methods outperform (cf. [13]) those of direct wavelet filtering [7][8]. However Bayesian approaches require having enough "useful signals" in the observations in order to well fit the prior (usually by estimating prior parameters). For wavelet-based Bayesian estimation, "useful signals" are those significant wavelet coefficients. In the low-intensity case where we generally lack such "useful signals", the final estimation can suffer from a large bias produced by a improperly fitted model.

Another previous important contribution to this field is the wavelet-domain hypothesis testing framework first introduced by Kolaczyk [17][18]. Thresholds based on user-specified false detection rate are derived for Haar coefficients both in the constant and model-based background situations. It is then found out that universal thresholds are also available in the above two cases [18][19]. Always in the same framework, Kolaczyk [20] proposed the Poisson-corrected version of the Gaussian-based threshold for any wavelet. But the asymptotic approximation used in [20] may not allow reasonable threshold solution in a very low-intensity setting.

In this paper, we adopt the same hypothesis testing framework as used in [17][20][18]. This approach is non-parametric (model-free) where a binary hypothesis testing procedure, which is based on the calculation of the p-value of the wavelet coefficient, is applied on each coefficient. This testing procedure is actually realized by a hard thresholding estimator. The coefficients with values bounded by the upper and lower thresholds are considered insignificant and are thresholded to zero, while significant ones are kept intact. The main advantage of this method is that the user can control the false positive rate in a coefficient-by-coefficient manner. Global statistical error rates could also be controlled if multi-test scheme is applied, such as Bonferroni correction [21] to control the Family-Wise Error Rate (FWER) and Benjamini-Hochberg's procedure [22][23] to control the False Discovery Rate (FDR).

The contribution of this paper is threefold. First, we propose a new estimator for the testing procedure in the Haar domain, which is based on Fisher's normal approximation [24] and results in a higher sensitivity level than Kolaczyk's threshold [17][18]. The universal threshold is also available, serving as a default choice. Second, unlike Kolaczyk's threshold, which requires prior background intensity and is only scale-dependent, our threshold estimates local background from the observations and it depends on both wavelet scale and location. Third, toward the goal of preserving the regularity of reconstructed data, which is of paramount importance for many applications, we propose to apply the same threshold in the biorthogonal Haar (Bi-Haar) domain instead of in the classical Haar domain. To justify this, we formally prove that non-normalized Bi-Haar coefficients converge in distribution to non-normalized Haar coefficients as the scale increases for any dimensional data. We will see, from the experiments, that the modified threshold combined with biorthogonal transform offers one of the best results compared with other estimators.

The paper is organized as follows. We first review in Section II the hypothesis testing framework in the wavelet domain. After presenting the original Kolaczyk's threshold in Section II-A.1, our threshold is derived in Section II-A.2. In Section III-A, we reveal the approximation property between non-normalized Bi-Haar coefficients and those of Haar. The entire estimation procedure is summarized in Section III-B and the results are shown in Section IV.

Finally, a discussion is presented and some concluding remarks are drawn in Section V. The proofs of theoretical results are deferred to the appendices.

II. Hypothesis testing in the wavelet domain

The thresholding estimator in the wavelet domain will be derived in a binary hypothesis testing framework. To accomplish this denoising task, we need the information of the background emission rate in the data. In our case, we will assume this background intensity to be locally constant (at least in the wavelet support). We will see that this local intensity can be estimated from the approximation coefficients. It can also be obtained if the user knows *a priori* the background rate. In any case, we can define respectively the null hypothesis H_0 and the alternative hypothesis H_1 for each wavelet coefficient w

H_0 : w is consistent with the background;

H_1 : w is inconsistent with the background.

This means that if w is judged to be consistent with the background (H_0 accepted), w should be replaced by the value obtained by computing a wavelet transform of the background without noise. As any wavelet has a zero mean and due to the local-constancy hypothesis of the background rate, w should be set to zero if H_0 is true. In other words, a non-zero value of w when the null hypothesis is true means that this value must uniquely due to the fluctuations of the counting statistics. Reversely, an observed coefficient verifying H_1 will be kept intact, since it contains the information that reflects the variation of the underlying intensity.

We will use the conventional notation for a wavelet coefficient, i.e. the capital letter W denotes the random variable while the lower-case w stands for the observed value.

A. Individual hypothesis testing (IHT)

After the wavelet transform, the IHT is carried out in a coefficient-by-coefficient manner. The user is supposed to specify a false detection rate (false positive or type-I error rate) in the wavelet domain, i.e. p, meaning that the probability that an H_0 coefficient is misclassified to H_1 will be upper bounded by p

$$\mathbb{E}\left(\frac{N_{H_0}^{reject}}{N_{H_0}}\right) \leq p \tag{1}$$

where $N_{H_0}^{reject}$ is the number of coefficients verifying H_0 but rejected (false positives) and N_{H_0} is the total number of H_0 coefficients. The lower and upper decision thresholds, t_{min} and t_{max}, are derived respectively from the tail probabilities $\Pr(W \leq t_{min}|H_0) \leq p/2$ and $\Pr(W \geq t_{max}|H_0) \leq p/2$. Then for any observed w verifying $w \geq t_{max}$ or $w \leq t_{min}$, w is deemed significant and will be classified to H_1. Otherwise it is decided insignificant and will be classified to H_0.

For a general wavelet, the analytical form of the distribution of the wavelet coefficient conditioning on H_0 of a Poisson process has been derived in [25]. This implies that theoretically, the hypothesis testing can be applied on any wavelet transform. However, this model requires the computation of the auto-convolution of the normalized wavelet histogram, which has to be evaluated numerically in practice and this calculation can be very heavy. So

we will seek a simple wavelet function that can result in a distribution having wieldy analytical form, and that motivates us to use the Haar wavelet. In this case, the distribution has a closed-form expression [26][27]

$$p_W(W = n; \lambda) = e^{-\lambda} I_n(\lambda)$$

where the Haar coefficient W is the difference of two i.i.d. Poisson variables $W = X_1 - X_2$, $X_1, X_2 \sim \mathcal{P}(\lambda/2)$, λ is the constant average counts in the dilated wavelet support and $I_n(\cdot)$ is the modified Bessel function of the first kind. For negative values of W, p_W can be obtained by symmetry since the distribution of Haar coefficient under H_0 is non-skewed. The tail probability is given by [28]

$$\Pr(W \geq n; \lambda) = \Pr\left(\chi^2_{(2n)}(\lambda) < \lambda\right), \qquad n \geq 1 \tag{2}$$

where $\chi^2_{(f)}(\Delta)$ is the non-central chi-square distribution with f degrees of freedom and Δ as non-centrality parameter.

1) Kolaczyk's CLT-based threshold [17][20][18]: We first denote the low-pass filter of the discrete Haar transform as $h = 2^{-c}[1, 1]$ and the corresponding high-pass filter as $g = 2^{-c}[-1, 1]$, where $2^{-c}(c \geq 0)$ corresponds to a certain normalization. The background intensity is λ_B, which is assumed constant and known in Kolaczyk's constant background model. The data dimension is denoted as d and the current scale index is j $(j \geq 1)$. We can see that every Haar coefficient (of any band) at the scale j is in the form of $W_j = 2^{-cjd}(X_1 - X_2)$, where X_1 and X_2 are independent Poisson random variables. More specifically, X_1 and X_2 are both the sum of 2^{jd-1} independent Poisson random variables. If W_j verifies H_0, we will have $\mathbb{E}(X_1) = \mathbb{E}(X_2) = \lambda_j/2$ where $\lambda_j = 2^{jd}\lambda_B$. We want to determine a threshold $t_j > 0$ such that $\Pr(W_j \geq t_j | H_0) = p/2$, or $\Pr(|W_j| \geq t_j | H_0) = p$ since the distribution of W_j is symmetric about zero. From (2), which corresponds to the case $c = 0$, we can easily deduce that

$$\Pr(W_j \geq t_j | H_0) = \Pr\left(\chi^2_{(2m_j)}(\lambda_j) < \lambda_j\right) \tag{3}$$

where $m_j = 2^{cjd} t_j$.

Then two stages of approximation are used

$$\begin{aligned} \Pr\left(\chi^2_{(2m_j)}(\lambda_j) < \lambda_j\right) &\approx \Pr(\alpha\chi^2_{(f)} < \lambda_j) && (4) \\ &\approx \Pr\left(Z > \frac{f - \lambda_j/\alpha}{\sqrt{2f}}\right) && (5) \end{aligned}$$

where $\alpha = (2m_j + 2\lambda_j)/(2m_j + \lambda_j)$, $f = (2m_j + \lambda_j)^2/(2m_j + 2\lambda_j)$, $\chi^2_{(v)}$ is a central chi-square random variable and $Z \sim \mathcal{N}(0, 1)$ is the standard normal distribution. The first approximation (4) is due to Patnaik [29] where the non-central chi-square distribution in (3) is approximated by a central one. The second (5) comes from the central limit theorem (CLT). Therefore in the sequel, the corresponding threshold will be called CLT-based (CLTB) threshold.

A simple calculation from the equation $\frac{f - \lambda_j/\alpha}{\sqrt{2f}} = z_{p/2}$ shows that the unique threshold solution is

$$t_j = 2^{-cjd-1}\left(z^2_{p/2} + \sqrt{z^4_{p/2} + 4 \cdot \lambda_j z^2_{p/2}}\right) \tag{6}$$

where $z_{p/2}$ is the critical value of the standard normal distribution at the significance level $p/2$, i.e. $z_{p/2} = \Phi^{-1}(1-p/2)$, where Φ is the standard normal cumulative distribution function. Now the universal threshold can be obtained by setting $z_{p/2} = \sqrt{2\ln N_j}$ in (6) where N_j is the number of Haar coefficients of one band at the scale j. This is a direct consequence of the well known fact in extreme value theory (cf. [30]) that for n independent standard normal variables $(X_i)_{i=1,\ldots,n}$, we have $\Pr(\max_{1\leq i\leq n}|X_i| > \sqrt{2\ln n}) \to 0$ as $n \to \infty$. Unlike the universal threshold in the Gaussian case [1], only band-dependent version can be expected in the Poisson case, since the inter-band wavelet coefficients are not independent.

2) Proposed threshold: Kolaczyk's method assume a global constant background intensity λ_B. Therefore this method is only well adapted to burst-like intensities characterized by occasional peaks on an almost flat region [17][19]. In addition, prior information on λ_B is not always available in practice. One straightforward way to generalize the above threshold to non-constant background situation without using a prior background model, is to estimate the local background from the observations and derive a threshold depending both on the scale j and on the location k. Charles and Rasson [19] introduced a preprocessing step to do this. The observation interval is first segmented in a way that the counts are almost stationary in each subinterval. Then follows the subinterval background estimation. This preprocessing is actually based on the tests of synthetic data. In our case, due to the local constancy hypothesis, we must estimate the local intensity $\lambda_{j,k}$ in the support of the wavelet generating the coefficient $W_{j,k} = 2^{-cjd}(X_1 - X_2), X_1, X_2 \sim \mathcal{P}(\lambda_{j,k}/2)$. Instead of using a preprocessing, $\lambda_{j,k}$ can be easily estimated from approximation coefficients (cf. Section II-B). Thus, the background estimation and the thresholding are essentially combined in a seamless way in our entire algorithm (cf. Section III-B). In the following, $\lambda_{j,k}$ is assumed to be already estimated.

In the two approximations (4) and (5), the former has been shown to be very accurate even in a low-intensity setting [31]. But the convergence rate of the latter (CLT) is known to be low (also true for the method of Charles and Rasson [19] since the same approximation is used). A widely-known faster version is given by Fisher [24]

$$\sqrt{2\chi^2_{(f)}} \to \mathcal{N}(\sqrt{2f-1}, 1), \quad f \to \infty \tag{7}$$

From the same derivation as above, we should change the equation (5) to

$$\Pr\left(\alpha\chi^2_{(f)} < \lambda_{j,k}\right) \approx \Pr\left(Z > \sqrt{2f-1} - \sqrt{\frac{2\lambda_{j,k}}{\alpha}}\right) \tag{8}$$

where $\alpha = (2m_{j,k} + 2\lambda_{j,k})/(2m_{j,k} + \lambda_{j,k})$, $f = (2m_{j,k} + \lambda_{j,k})^2/(2m_{j,k} + 2\lambda_{j,k})$ and $m_{j,k} = 2^{cjd}t_{j,k}$.

Let's denote

$$\begin{aligned} G(m_{j,k}) &:= \sqrt{2f-1} - \sqrt{\frac{2\lambda_{j,k}}{\alpha}} \\ &= \sqrt{\frac{(2m_{j,k}+\lambda_{j,k})^2}{m_{j,k}+\lambda_{j,k}} - 1} - \sqrt{\frac{\lambda_{j,k}(2m_{j,k}+\lambda_{j,k})}{m_{j,k}+\lambda_{j,k}}} \end{aligned} \tag{9}$$

Therefore, it remains to solve the equation $G(m_{j,k}) = z_{p/2}$, to obtain the new Fisher approximation-based (FAB) threshold. A direct algebraic rearrangement shows that $m_{j,k}$ necessarily satisfies the following quartic polynomial

equation

$$\begin{aligned}
&16m_{j,k}^4 + \left[16\lambda_{j,k} - 8(z_{p/2}^2+1)\right] m_{j,k}^3 + \\
&\left[(z_{p/2}^2+1)^2 - (20z_{p/2}^2+12)\lambda_{j,k} + 4\lambda_{j,k}^2\right] m_{j,k}^2 + \\
&\left[2(z_{p/2}^2+1)^2\lambda_{j,k} - 16z_{p/2}^2\lambda_{j,k}^2 - 4\lambda_{j,k}^2\right] m_{j,k} + \\
&(z_{p/2}^2+1)^2\lambda_{j,k}^2 - 4z_{p/2}^2\lambda_{j,k}^3 = 0
\end{aligned} \tag{10}$$

The final FAB threshold $t_{j,k}$ is obtained from the solution of $m_{j,k}$. A closed-form expression for $m_{j,k}$ is certainly too complicated to write out from (10). Fortunately, according to the following results, we don't even need such an expression.

Fact 1: The feasible condition is

$$m_{j,k} \geq \frac{1}{8}\left[z_{p/2}^2 - 2\lambda_{j,k} + 1 + \sqrt{z_{p/2}^4 + (12\lambda_{j,k}+2)z_{p/2}^2 + 4\lambda_{j,k}^2 + 12\lambda_{j,k} + 1}\right] \tag{11}$$

Proposition 1: The feasible solution for FAB threshold exists and is unique.

The feasible condition (11) and the uniqueness of the feasible solution imply that we can use any numerical routine that solves quartic polynomial equations, e.g. Hacke's method [32]. Then the four solutions of (10) will be tested by the condition (11) and there will be one and only one winner. The universal threshold can also be obtained by solving (10) where $z_{p/2} = \sqrt{2\ln N_j}$. We refer the reader to Appendix II for the proofs of Fact 1 and of Proposition 1.

To have an idea of the refinement achieved by the modified FAB threshold, let's give the values of the two thresholds in a toy example $d = 1$, $c = 1/2$, $z_{p/2} = 3$ and $\lambda_B = 1$. This corresponds to use the classical Haar transform on the $1D$ data of a constant background rate ($= 1$) with a 3σ detection level. For $j = 1, 2, 3$, the values of CLTB threshold are given by $t_1 = 7.56$, $t_2 = 6.00$ and $t_3 = 4.99$, while those of the proposed version are $t_1 = 4.28$, $t_2 = 3.84$ and $t_3 = 3.54$. In the extreme case where the underlying intensity is zero, a threshold as low as possible is expected since we effectively observe zero-count data and no noise is present. In this situation, CLTB threshold gives $t_j = 9 \cdot 2^{-j/2}$ while the FAB threshold yields $t_j = 2.5 \cdot 2^{-j/2}$. The former is almost three times higher than the latter. The lower values imply that a potential more sensitive decision may be achieved by the proposed threshold. This will be verified in Section IV-A.

B. Background emission rate estimation

We will now determine $\lambda_{j,k}$ as required in the modified threshold. Since we use $h = 2^{-c}[1, 1]$ as the low-pass Haar filter, the approximation coefficient at the scale j is in the form of $A_{j,k} = 2^{-cjd}(X_1 + X_2)$, where X_1 and X_2 are both the sum of 2^{jd-1} independent Poisson variables. When dealing with constant background case with known λ_B, $\lambda_{j,k}$ is obviously $2^{jd}\lambda_B$. In other cases, $\lambda_{j,k}$ must be estimated. Notice that under H_0, $\hat{\lambda}_B := A_{j,k}/2^{jd(1-c)}$ can be treated as an estimate of the background rate per data point in the current wavelet support, since this is in effect the mean of 2^{jd} i.i.d. Poisson variables. Then $\lambda_{j,k}$ is estimated by $\hat{\lambda}_{j,k} := 2^{jd}\hat{\lambda}_B = A_{j,k}/2^{-cjd}$, which is just the total counts in the wavelet support. However these coefficients may contain noise, so the denoising must

be carried out from coarser scales to finer ones. More specifically, after we have denoised the wavelet coefficients at the scale $j+1$, the approximation coefficients at the scale j are reconstructed. These reconstructed coefficients normalized by 2^{-cjd} and submitted to a positivity projection (since $\lambda_{j,k} \geq 0$) serve as the estimates of $\lambda_{j,k}$ in the denoising process at the scale j.

III. Regularity preservation

A. Biorthogonal Haar transform

The main disadvantage of using the Haar transform is that the reconstructed data can have "staircase" artifacts. To achieve a more regular reconstruction, we propose to use the Bi-Haar wavelet instead of the classical Haar wavelet. It appears that we must redesign the threshold since the Bi-Haar coefficients are no more differences of i.i.d. Poisson variables. However we can apply straightforwardly the above proposed threshold in the Bi-Haar domain owing to the following result.

Theorem 1: Suppose $h = 2^{-c}[1,1]$ and $g = 2^{-c}[-1,1]$ to be respectively the low and high-pass decomposition filters of the Haar transform; those of the Bi-Haar transform are respectively denoted as $h = 2^{-c}[1,1]$ and $g = 2^{-c}[1/8, 1/8, -1, 1, -1/8, -1/8]$. Then for any dimensional data, with $c > 1$, the Bi-Haar coefficients converge in distribution to those of Haar as the scale increases.

This theorem is proved in Appendix I. It justifies that we can treat non-normalized Bi-Haar coefficients as if they were coming from a non-normalized Haar transform. Therefore, the thresholds derived in the case of Haar can be used in the Bi-Haar transform without modification.

B. Summary of the estimation algorithm

The proposed intensity estimation procedure is outlined as follows. We use j as the scale index where $0 \leq j \leq J$. $j = 0$ corresponds to the initial scale before wavelet decomposition. $a_{j,k}$ and $w_{j,k}$ stand for respectively the observed approximation coefficient and wavelet coefficient at the scale j and location k.

Algorithm 1. FAB threshold Poisson denoising

1: Initialize $j = J$

2: Wavelet transform of $\mathbf{v}$ to obtain $a_{J,\cdot}$ and $w_{i,\cdot}$ $(1 \leq i \leq J)$

3: Do the following

- calculate $z_{p/2}$
- $\lambda_{j,k} = 2^{jd}\lambda_B$ if λ_B is known; otherwise $\lambda_{j,k} = \max(a_{j,k}/2^{-cjd}, 0)$
- Solve (10) to obtain the unique feasible solution $t_{j,k}$
- $w_{j,k} = w_{j,k} \cdot \mathbf{1}_{\{|w_{j,k}| \geq t_{j,k}\}}$
- Reconstruct $a_{j-1,k}$ by the inverse wavelet transform

4: $j = j - 1$ and if $j > 0$ goto Step 3

5: $\hat{\Lambda} = \max(a_{0,\cdot}, 0)$

$z_{p/2}$ is obtained either from a user-specified error rate p or from the universal setting, $\mathbf{1}_S$ is the identity function on S and $\hat{\Lambda}$ is the estimated intensity. Notice that the last step corresponds to the positivity constraint on the underlying intensity.

IV. Results

This part is arranged as follows. Section IV-A gives the performance evaluation of the statistical decision procedures employed in CLTB threshold and FAB one. The improvement achieved by using Bi-Haar is discussed in Section IV-B. Then a thorough comparison of our approach with various existing estimators is shown in Section IV-C. Finally in Section IV-D, applicability of our method is explored on astronomical imaging.

A. Sensitivity analysis

Let's assess the statistical power corresponding to the two thresholding policies CLTB and FAB. We will show their true detection rate (or true positive rate) in the wavelet domain for different signals. To achieve a fair comparison, we first adopt the "Bursts" intensity [17](cf. Fig. 1(a)) whose minimum (maximum) intensity is $\lambda_{min} = 0.795$ ($\lambda_{max} = 5.3$) in our experiment. The background level λ_B is thus set to λ_{min}. Notice that the ground truth of the significant coefficients is known, since they are simply those non-zero coefficients of the intensity function. During our experiment, we generated 100 noisy signals from the intensity and for each signal, we measured the true detection rate of the two thresholds at different preset p-values. The mean rate is thus obtained and is shown in Fig. 1(b). The same procedure is applied on another intensity "Spikes", containing spikes and smooth parts [13](cf. Fig. 1(c)) with the result shown in Fig. 1(d). Fig. 1(b)(d) compare also the mean rates given by the universal thresholds. It is clear that the true detection rate given by FAB (universal FAB) significantly exceeds that of CLTB (universal CLTB). Consequently, FAB (universal FAB) results in a more powerful decision than CLTB (universal CLTB).

B. Haar vs. Bi-Haar transform

Since the data reconstructed by Haar transform tend to have "staircase" artifacts, we have proposed to use the Bi-Haar transform instead of Haar. To justify that Bi-Haar really improves the reconstruction regularity, we generate noisy signals from the "Smooth" intensity function [13](cf. Fig. 2(a)) and measure the Normalized Mean Integrated Square Error (NMISE) per bin from the estimates under the two different transforms. NMISE is defined as follows where $\lambda(x)$ is the ground truth and $\hat{\lambda}(x)$ is the estimated intensity

$$\mathbb{E}\left[\int \left(\hat{\lambda}(x) - \lambda(x)\right)^2 / \lambda(x)\, dx\right] \tag{12}$$

Unlike the usual adopted criterion of MISE (Mean Integrated Square Error), NMISE takes the noise variance of the Poisson process ($\lambda(x)$) into account. This variance stabilization is necessary for evaluating inhomogeneous Poisson process estimation because otherwise, large relative errors in a low intensity region can be totally masked by even a small relative error in a high intensity zone.

In using universal FAB threshold, denoising examples with corresponding NMISE (measured from 100 replications) are shown in Fig. 2. It is evident, visually and quantitatively, that Bi-Haar based denoising outperforms that of the classical Haar in every intensity level, which clearly confirms the significant regularity improvement by Bi-Haar. Thus, this transform should be preferred as long as one deals with non piecewise constant intensities.

C. Comparison with other estimators

In Tab. II, we list the methods to compare, which cover most kinds of existing estimators. The first three, i.e. CLTB, FAB and FAB-BH, are hypothesis testing methods; ANS, FISZ and CVS-GCV are based on VST, and the last one BCVS is a Bayesian approach. We do not include the method of direct wavelet filtering [7][8] since it performs poorly in general and are surpassed by most above estimators. We adopt the same six types of intensity as those used in [13], i.e. "Angles", "Bumps", "Bursts", "Clipped Blocks", "Smooth" and "Spikes". A background is added on each one and then, the intensities are multiplied by the scaling factors $s = 0.1$, 1, 10 and 100 to create different intensity levels. Maximum and minimum intensity values before being scaled are shown in Tab. I. The intensities scaled by $s = 100$ with examples of generated noise are shown in Fig. 3. During the experiments, the coarsest scale is set to $J = 5$. Fig. 4 shows the NMISE calculated from 100 replications at each intensity level.

Let's summarize our conclusions in the following take-away messages.

1) For all the intensity levels, FAB-BH offers one of the best performance in all the intensity types except for "Clipped Blocks". It is clear that the performance gain by using FAB-BH instead of CLTB is significant. The same conclusion can be drawn from the comparison of translation-invariant (TI) denoising (not shown here). We refer the reader to Section V for more detailed discussion on TI estimation.
2) The fair performance of FAB-BH on "Clipped Blocks" is due to the conflict between its higher regularity reconstruction property and the piecewise constancy nature of this intensity. As a result, good performance is expected for Haar-based estimators in this case. We observed that FAB gives the best performance in most intensity levels. Although Fisz transform is also based on Haar wavelet, it merely gives average results in our experiments. This is likely due to the "internal" denoising wavelet Daubechies-4, which is less adapted to this piecewise constant case.
3) At the very low intensity level, all the methods are comparable except for CVS-GCV, which generates systematically large errors. The BCVS gives the second largest NMISE, which is due to the bias of the poorly fitted prior as we have discussed in Section I for Bayesian approaches. BCVS performs well on smooth intensities, but is eclipsed by others for less smooth signals such as "Bumps", "Clipped Blocks" and "Spikes".

.

TABLE I

INTENSITIES

	Angles	Bumps	Bursts	Clipped Blocks	Smooth	Spikes
λ_{min}	0.882	0.877	0.795	0.918	0.706	0.882
λ_{max}	5.882	5.850	5.299	6.118	4.706	5.882

TABLE II

ESTIMATORS

Name	Brief Description	Wavelet	Threshold	References
CLTB	Kolaczyk's method (λ_B known)	Haar	Universal Hard	[17][18][20]
FAB	Our method (estimated background)	Haar	Universal Hard	Section II-A.2
FAB-BH	Our method (estimated background)	Bi-Haar	Universal Hard	Section II-A.2, III-A
ANS	VST : Anscombe transform	Daubechies-4[a]	Universal Hard	[5][3]
FISZ	VST : Fisz transform	Daubechies-4	Universal Hard	[4]
CVS-GCV	Conditional variance stabilization	Daubechies-4	General Cross Validation Soft	[14][33]
BCVS	Bayesian conditional variance stabilization	Daubechies-4	Bayesian Shrinkage[b]	[14]

[a] Daubechies wavelet of 4 vanishing moments

[b] To be more precise, Bayesian posterior median shrinkage

D. Application to astronomical imaging

In this experiment, we have simulated an image with circle-like X-ray sources (cf. Fig. 5(a)) for XMM-Newton telescope, which is currently the most powerful X-ray satellite. Each source along any radial branch has the same flux and has a more and more extended support as we go farther from the center. The flux reduces as the branches turn in the clockwise direction. In effect, this image can be seen as a model for celestial objects of different sizes and of different flux. The observed image of counts is shown in Fig. 5(b). Fig. 5(c) and (d) present respectively the estimated intensity by CLTB and by FAB-BH, while the results of TI estimations (cf. Section V) are shown in Fig. 6(c)(d).

The local zooms of the estimates presented in Fig. 5(e) and (f) confirm that the reconstruction of FAB-BH has less "staircase" artifacts. When TI denoising is applied, the reconstructions are almost artifact-free (Fig. 6(e)(f)). Fig. 5(g)(h) and Fig. 6(g)(h) show that for those low-flux sources (i.e. those located in the upper-right part of Fig. 5(a)), more significant responses are given by FAB-BH (TI FAB-BH) than by CLTB (TI CLTB) thanks to the higher sensibility of FAB.

V. DISCUSSION AND CONCLUSION

We have presented a Poisson intensity estimation method for d-dimensional data based on hypothesis testing in the wavelet domain. A more sensitive thresholding estimator realizing the hypothesis testing, i.e. FAB threshold, has been derived. In particular, the universal threshold exists and serves as a default value. The IHT controls the error

rate in a coefficient-by-coefficient manner and there is no correction for multiple comparison. To control a global statistical error rate, we could resort to multiple hypothesis testing (MHT). For example, the Bonferroni correction controls the probability of erroneously rejecting even one of the true null hypothesis, i.e. FWER. Alternatively, one can use the Benjamini and Hochberg [22][23] procedure to control the FDR. The control of FDR has many advantages over the control of FWER. For example, usually it has a greater detection power and it can handle correlated data easily [23]. It has also been demonstrated in the Gaussian case (or the Poisson case with a high intensity), that the FDR estimator achieves asymptotic minimaxity, when the total number of tests tends to infinity and that the controlled FDR tends to zero. We refer the reader to [22][23][34][35] for more details.

To achieve a higher reconstruction regularity, we have proposed to use the biorthogonal Haar transform instead of the Haar transform, after proving the approximation property of the biorthogonal coefficients to those of Haar. Our experiments suggest that the FAB universal threshold combined with undecimated biorthogonal Haar transform gives one of the best denoising result among existing estimators for a wide range of intensity types.

Amelioration of a wavelet denoiser can be gained by using TI transforms. In our case, TI denoising helps further eliminate the "staircase" artifacts in the reconstruction. A possible way to achieve TI denoising is by employing the cycle-spinning procedure [36]. However our method allows a more straightforward and computationally efficient way by using the undecimated wavelet transform (UWT). It is important to notice that in our case, cycle-spinning and UWT are not strictly equivalent and offer slightly different results. This is due to our special way of multiscale background estimation at the current scale, which depends on the denoising results in all the coarser scales. Another observation is that the UWT does not change the statement of Theorem 1. From some experiments that we carried out (not shown here), Bi-Haar estimation can even challenge that of TI Haar. Indeed, unlike the orthogonal Haar case whose performance can be considerably enhanced by using the TI version of the transform, translation invariance is less beneficial to the case of Bi-Haar.

One further improvement of this algorithm may be gained by using iterative restoration. In the case of Gaussian denoising, several authors have proposed regularized iterative methods with constraints in multiscale transforms. For instance, a total variation (TV) constraint has been respectively introduced in wavelet denoising [37], in wavelet packet denoising [38] and in curvelet denoising [39]. An L^1 constraint in the wavelet domain is proposed by Starck [40], which has the advantage of free from the creation of staircase-like structures during the TV minimization in the above approaches.

It is also possible to attain additional reconstruction improvement by applying different reconstruction filters in the undecimated transform. Since a redundant transform allows actually an infinite number of inverse operator, we can construct data-adapted reconstruction filter to achieve a more regular reconstruction [41].

However, in the Poisson denoising case, we have not obtained so far outspoken results worth being reported here, either from iterative scheme of [40] or from redesigning the Haar reconstruction filters (we tried B_3-Spline as reconstruction scaling function [41]).

One aspect of our future research is to seek for iterative schemes and alternative reconstruction filters adapted to Poisson case. We are also currently investigating the possibility of extending this methodology to the deconvolution

of Poisson noise contaminated data.

APPENDIX I
PROOF OF THEOREM 1

The final proof is based on the cumulant-generating function (CGF).

A. Cumulant-Generating Function

Let's first recall the CGF of a real random variable X.

Definition 1: **(CGF)** *Let $\phi(t)$ be the characteristic function of X. The CGF of (the distribution of) X is $\psi(t) := \ln \phi(t)$ ($t \in \mathbb{R}$).*

Definition 2: **(Cumulants)** *Let ψ be the CGF of X, and let's assume that ψ can be developed into a Maclaurin series*

$$\psi(t) = \sum_{p \geq 1} \kappa_p \frac{(it)^p}{p!}$$

The coefficient κ_p is then called the p-th cumulant of X.

Proposition 2: Let $(X_n)_{n\geq 1}$ be a sequence of random variables. Assuming ϕ_n and ψ_n as respectively the characteristic function and the CGF of X_n. Assume furthermore that the p-th cumulants of X_n is $\kappa_{n,p}$. Let X_0 be a random variable with the characteristic function ϕ_0, the CGF ψ_0 and the p-th cumulant $\kappa_{0,p}$. If $\kappa_{n,p} \to \kappa_{0,p}$ as $n \to \infty$ and if the convergence is uniform with respect to p, then $X_n \xrightarrow{D} X_0$.

Proof: Since

$$\psi_n(t) = \sum_{p \geq 1} \kappa_{n,p} \frac{(it)^p}{p!}$$

$$\psi_0(t) = \sum_{p \geq 1} \kappa_{0,p} \frac{(it)^p}{p!}$$

we have

$$|\psi_n(t) - \psi_0(t)| \leq \sum_{p \geq 1} |\kappa_{n,p} - \kappa_{0,p}| \frac{|t|^p}{p!} \tag{13}$$

As $\kappa_{n,p} \to \kappa_{0,p}$ $(n \to \infty)$ and the convegence is uniform on p, we have when n is sufficient large

$$\forall p, \quad |\kappa_{n,p} - \kappa_{0,p}| < \epsilon \quad \Rightarrow \quad (13) < \epsilon(e^{|t|} - 1)$$

Hence, the sequence ψ_n converges pointwisely to ψ_0. By the continuity of the function $x \mapsto e^x$, we have that ϕ_n converges pointwisely to ϕ_0. Notice that ϕ_0 is always continuous at the origin, so we can apply Lévy's theorem [42] to conclude that $X_n \xrightarrow{D} X_0$. □

The following proposition shows the cumulants of the output of a linear filter.

Proposition 3: Suppose that we have a sequence of independent random variables $(X_n)_{n\geq 1}$. Let's denote the p-th cumulant of X_n as $\kappa_{n,p}^X$ and the impulse response of a filter h as $(h_k)_{k\in\mathbb{Z}}$. Assume that for each i,

$\sum_{k\in\mathbb{Z}} |\kappa^X_{i-k,p}||h_k|^p$ is uniformly bounded with respect to p. If $(Y_i)_{i\in\mathbb{Z}}$ is the output of the filter h by using X_n as input, i.e. $Y_i = \sum_{k\in\mathbb{Z}} h_k X_{i-k}$, then

$$\kappa^Y_{i,p} = \sum_{k\in\mathbb{Z}} \kappa^X_{i-k,p} h_k^p$$

is the p-th cumulant of Y_i.

Proof: By the definition of the CGF of Y_i, we have

$$\psi_{Y_i}(t) = \ln \mathbb{E}(\Pi_{k\in\mathbb{Z}} e^{it\cdot h_k X_{i-k}})$$

From the dominated convergence and the independence of X_n, we have

$$\begin{aligned}
\psi_{Y_i}(t) &= \lim_{(m,n)\to(\infty,\infty)} \ln \Pi_{-m\le k\le n} \mathbb{E}(e^{it\cdot h_k X_{i-k}}) \\
&= \lim_{(m,n)\to(\infty,\infty)} \sum_{-m\le k\le n} \psi_{X_{i-k}}(h_k t) \\
&= \lim_{(m,n)\to(\infty,\infty)} \sum_{-m\le k\le n} \sum_{p\ge 1} \kappa^X_{i-k,p} h_k^p \frac{(it)^p}{p!} \\
&= \lim_{(m,n)\to(\infty,\infty)} \sum_{p\ge 1} \frac{(it)^p}{p!} \left(\sum_{-m\le k\le n} \kappa^X_{i-k,p} h_k^p \right) \qquad (14)
\end{aligned}$$

Since $\sum_{k\in\mathbb{Z}} |\kappa^X_{i-k,p}||h_k|^p$ is uniformly bounded with respect to p, the dominated convergence can be applied to (14) and results in

$$\psi_{Y_i}(t) = \sum_{p\ge 1} \left(\sum_{k\in\mathbb{Z}} \kappa^X_{i-k,p} h_k^p \right) \frac{(it)^p}{p!}$$

Immediately we identify that

$$\kappa^Y_{i,p} = \sum_{k\in\mathbb{Z}} \kappa^X_{i-k,p} h_k^p$$

□

We are at the point to prove Theorem 1 and we first consider one dimensional data.

B. Proof for $1D$ case

Let's consider a sequence of $1D$ Poisson independent random variables $(X_n)_{1\le n\le N}$, where $N = 2^J$. The approximation coefficient (A) of the Haar transform at the scale $1 \le j \le J$ is in the form of

$$A = 2^{-cj} Y$$

where Y is the sum of 2^j independent Poisson random variables (of $(X_n)_{1\le n\le N}$). The wavelet or detail coefficient (W) is in the form of

$$W = 2^{-cj}(Z_1 - Z_2)$$

where Z_1 and Z_2 are independent Poisson random variable and each is the sum of 2^{j-1} independent Poisson random variables (of $(X_n)_{1\le n\le N}$).

Now for biorthogonal case, the approximation coefficient has the same form as that of Haar. But its detail coefficient (W_b) is

$$W_b = 2^{-cj}(Z_1 - Z_2 + \frac{1}{8}(Z_3 - Z_4))$$

where Z_1 and Z_2 are respectively the sum of 2^{j-1} independent Poisson random variables (of $(X_n)_{1\leq n\leq N}$), and Z_3 and Z_4 are respectively the sum of 2^j independent Poisson random variables (of $(X_n)_{1\leq n\leq N}$). Moreover, Z_1, Z_2, Z_3 and Z_4 are mutually independent.

By Proposition 3, we can write the p-th cumulant of W_b, i.e. $\kappa_p^{W_b}$

$$\begin{aligned}\kappa_p^{W_b} &= 2^{-cpj}\left(\kappa_p^{Z_1} + (-1)^p\kappa_p^{Z_2} + 8^{-p}(\kappa_p^{Z_3} + (-1)^p\kappa_p^{Z_4})\right)\\ &= \kappa_p^{W} + 2^{-cpj}8^{-p}\left(\kappa_p^{Z_3} + (-1)^p\kappa_p^{Z_4}\right)\end{aligned}$$

Therefore we have,

$$\begin{aligned}|\kappa_p^{W_b} - \kappa_p^{W}| &= 2^{-cpj}8^{-p}|\kappa_p^{Z_3} + (-1)^p\kappa_p^{Z_4}|\\ &\leq 2^{-cpj}8^{-p}2\max(\kappa_p^{Z_3}, \kappa_p^{Z_4})\\ &\leq 2^{-cpj}8^{-p}\cdot 2^j 2\lambda_{max}\\ &= 2^{j(1-cp)-3p+1}\lambda_{max}\\ &=: r\end{aligned}$$

where λ_{max} is the maximum intensity of $(X_n)_{1\leq n\leq N}$. We can see that $r \to 0$ exponentially when p increases. The greatest error is reached when $p = 1$. Notice that if we want $\kappa_p^{W_b} \to \kappa_p^{W}$ for any p, c must be greater than one. In this case, the approximation error decreases exponentially as the scale, i.e. j, increases. Now Proposition 2 can be applied here. Notice that the uniform convergence with respect to p is verified. Immediately we conclude that for $1D$ data, when $c > 1$, the wavelet coefficients of non-normalized biorthogonal Haar transform converge in distribution towards those of non-normalized Haar transform.

C. Proof for dD case

Now let's look at the multi-dimensional case. We begin with the $2D$ case. We have a $2D$ Poisson signal $(X_{m,n})_{1\leq m,n\leq N}$ where $N = 2^J$. Clearly, we can only consider the approximation on detail coefficients, since the approximation coefficients of the two wavelet transform are identical. Let's denote the p-th cumulant of the Haar coefficient of the band hg, gh and gg respectively as κ_p^{HG}, κ_p^{GH} and κ_p^{GG}. Those of biorthogonal coefficient will be denoted respectively as $\kappa_p^{HG_b}$, $\kappa_p^{GH_b}$ and $\kappa_p^{GG_b}$. Similar calculation as that in the $1D$ case shows the following results. At the j-th scale,

for the band hg,

$$|\kappa_p^{HG_b} - \kappa_p^{HG}| \leq 2^{2j(1-cp)-3p+1}\lambda_{max}$$

for the band gh,

$$|\kappa_p^{GH_b} - \kappa_p^{GH}| \quad \leq \quad 2^{2j(1-cp)-3p+1}\lambda_{max}$$

and for the band gg,

$$|\kappa_p^{GG_b} - \kappa_p^{GG}| \quad \leq \quad 2^{2j(1-cp)-3p+2}(1+8^{-p})\lambda_{max}$$

If we repeat the same reasoning as that in the $1D$ case, we will again find that when $c > 1$, the wavelet coefficients of non-normalized biorthogonal Haar transform converge in distribution towards those of non-normalized Haar transform. It is clear that the above calculation and reasoning can be extended to d-dimensional ($d > 2$) case without difficulty, and the conclusion is the same. □

APPENDIX II
PROOF OF PROPOSITION 1

A. Proof of Fact 1

After an algebraic rearrangement of $G(m_{j,k}) = z_{p/2}$ and with the fact $z_{p/2} > 0$, it is found out that feasible solutions necessarily satisfy

$$\frac{2m_{j,k}(2m_{j,k} + \lambda_{j,k})}{m_{j,k} + \lambda_{j,k}} - (z_{p/2}^2 + 1) \geq 0$$

The above inequality combined with $2f - 1 \geq 0$, $m_{j,k} > 0$ and $\lambda_{j,k} \geq 0$ shows Fact 1. □

B. Proof of Proposition 1

First let's denote the right hand side of (11) as β. When $m_{j,k} = \beta$, we have

$$G(m_{j,k}) = \sqrt{z_{p/2}^2 + \frac{\lambda_j}{2m_{j,k}}(z_{p/2}^2 + 1)} - \sqrt{\frac{\lambda_j}{2m_{j,k}}(z_{p/2}^2 + 1)} \leq z_{p/2}$$

Second, we can verify that G is a strictly increasing function under the condition (11) since $G'(m_{j,k}) > 0$, and clearly $G(m_{j,k}) \to \infty$ as $m_{j,k} \to \infty$.

Obviously, the above results guarantee immediately the existence and uniqueness of the feasible solution of the equation $G(m_{j,k}) = z_{p/2}$. This solution is contained in the solutions of (10). □

REFERENCES

[1] D. L. Donoho, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[2] ——, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[3] ——, "Nonlinear wavelet methods for recovery of signals, densities and spectra from indirect and noisy data," *Proc. Symp. Applied Mathematics: Different Perspectives on Wavelets*, vol. 47, pp. 173–205, 1993.

[4] P. Fryźlewicz and G. P. Nason, "A Haar-Fisz algorithm for Poisson intensity estimation," *J. Comp. Graph. Stat.*, vol. 13, pp. 621–638, 2004.

[5] F. J. Anscombe, "The Transformation of Poisson, Binomial and Negative-Binomial Data," *Biometrika*, vol. 35, pp. 246–254, 1948.

[6] M. Fisz, "The limiting distribution of a function of two independent random variables and its statistical application," *Colloquium Mathematicum*, vol. 3, pp. 138–146, 1955.

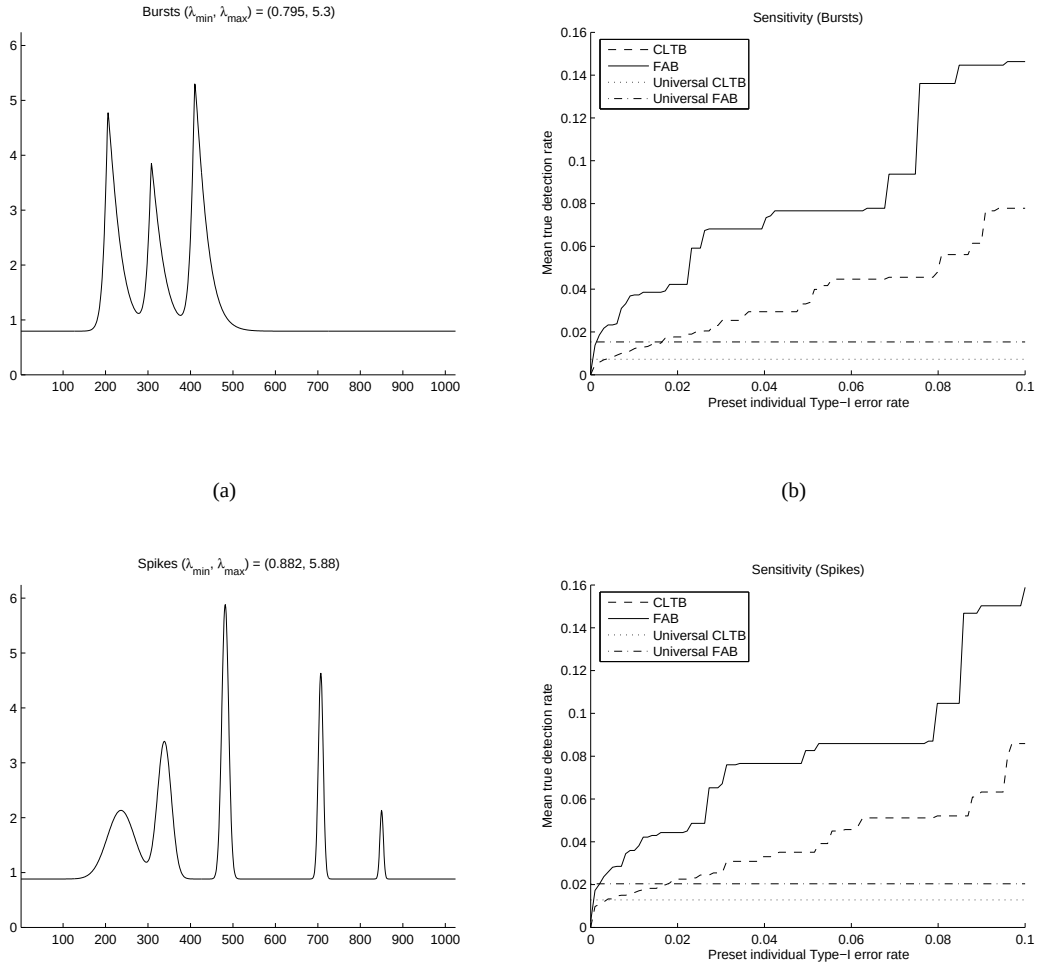


Fig. 1. Sensitivity of the thresholds. Decimated Haar transform is used where $J = 5$. (a) Bursts intensity; (b) mean true detection rate of Bursts; (c) Spikes intensity; (d) mean true detection rate of Spikes.

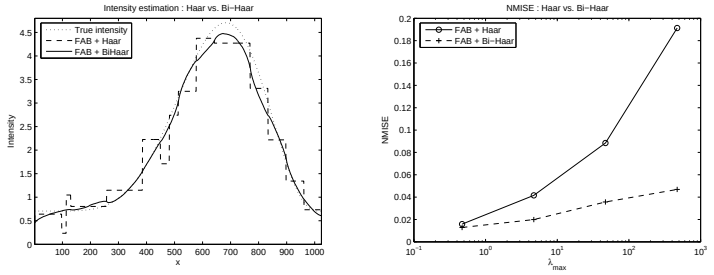


Fig. 2. Haar vs. Bi-Haar. Universal FAB is applied, $J = 7$ and the signal length is 1024. (a) denoising results; (b) NMISE.

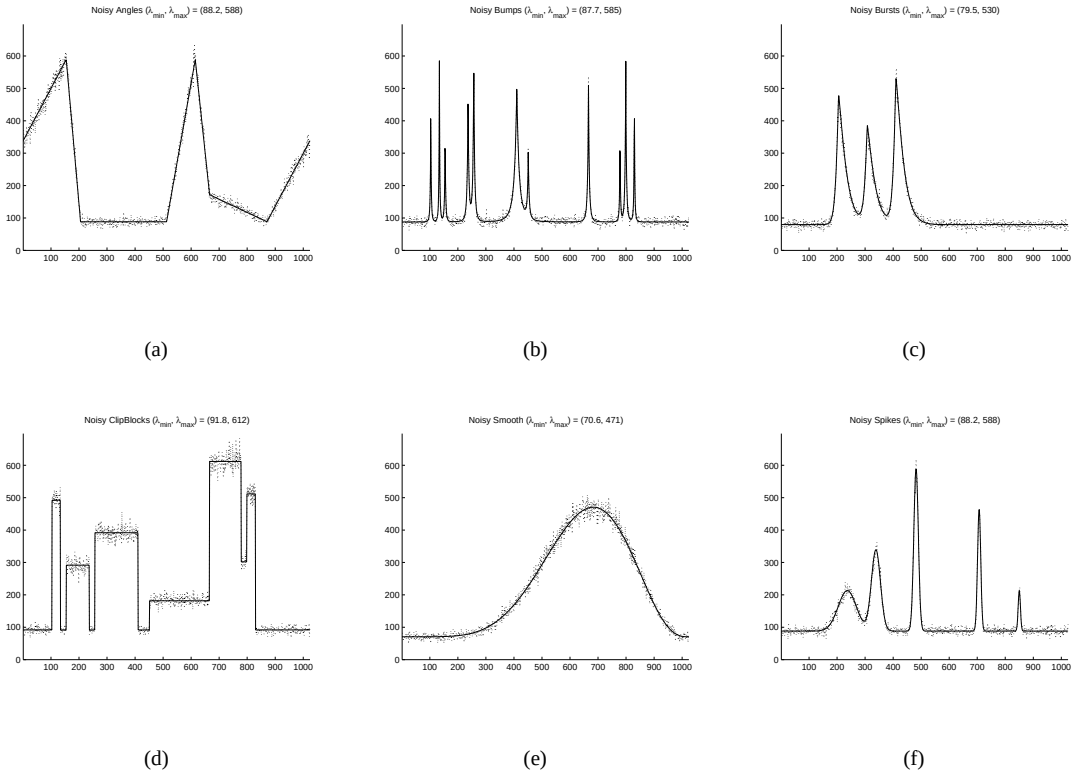


Fig. 3. Test intensities with examples of generated Poisson noise. Scaling factor $s = 100$. Length of each signal is 1024. (a) Angles; (b) Bumps; (c) Bursts; (d) Clipped Blocks; (e) Smooth; (f) Spikes.

[7] R. D. Nowak and R. G. Baraniuk, "Wavelet-Domain Filtering for Photon Imaging Systems," *IEEE Transactions on Image Processing*, vol. 8, no. 5, pp. 666–678, May 1999.

[8] A. Antoniadis and T. Sapatinas, "Wavelet shrinkage for natural exponential families with quadratic variance functions," *Biometrika*, vol. 88, pp. 805–820, 2001.

[9] E. D. Kolaczyk, "Bayesian multiscale models for Poisson processes," *J. Amer. Statist. Ass.*, vol. 94, no. 447, pp. 920–933, Sept. 1999.

[10] K. E. Timmermann and R. D. Nowak, "Multiscale Modeling and Estimation of Poisson Processes with Application to Photon-Limited Imaging," *IEEE Transactions on Information Theory*, vol. 45, no. 3, pp. 846–862, Apr. 1999.

[11] R. D. Nowak and E. D. Kolaczyk, "A statistical multiscale framework for Poisson inverse problems," *IEEE Transactions on Information Theory*, vol. 46, no. 5, pp. 1811–1825, Aug. 2000.

[12] H. Lu, Y. Kim, and J. M. M. Anderson, "Improved Poisson Intensity Estimation: Denoising Application Using Poisson Data," *IEEE Transactions on Image Processing*, vol. 13, no. 8, pp. 1128–1135, Aug. 2004.

[13] P. Besbeas, I. D. Feis, and T. Sapatinas, "A Comparative Simulation Study of Wavelet Shrinkage Estimators for Poisson Counts," *Internat. Statist. Rev.*, vol. 72, no. 2, pp. 209–237, 2004.

[14] M. Jansen, "Multiscale Poisson data smoothing," TU Eindhoven SPOR, Tech. Rep. 03-29, 2003.

[15] R. M. Willett and R. D. Nowak, "Multiscale likelihood analysis and image reconstruction," in *SPIE Wavelets X*, vol. 5207, San Diego, CA, Aug. 2003, pp. 97–111.

[16] S. Sardy, A. Antoniadis, and P. Tseng, "Automatic smoothing with wavelets for a wide class of distributions," *J. Comput. Graph. Stat.*, vol. 13, no. 2, pp. 399–421, June 2004.

[17] E. D. Kolaczyk, "Estimation of intensities of burst-like Poisson processes using Haar wavelets," *Biometrika*, vol. 46, pp. 352–363, 1996.

[18] ——, "Nonparametric estimation of intensity maps using Haar wavelets and Poisson noise characteristics," *The Astrophysical Journal*, vol.

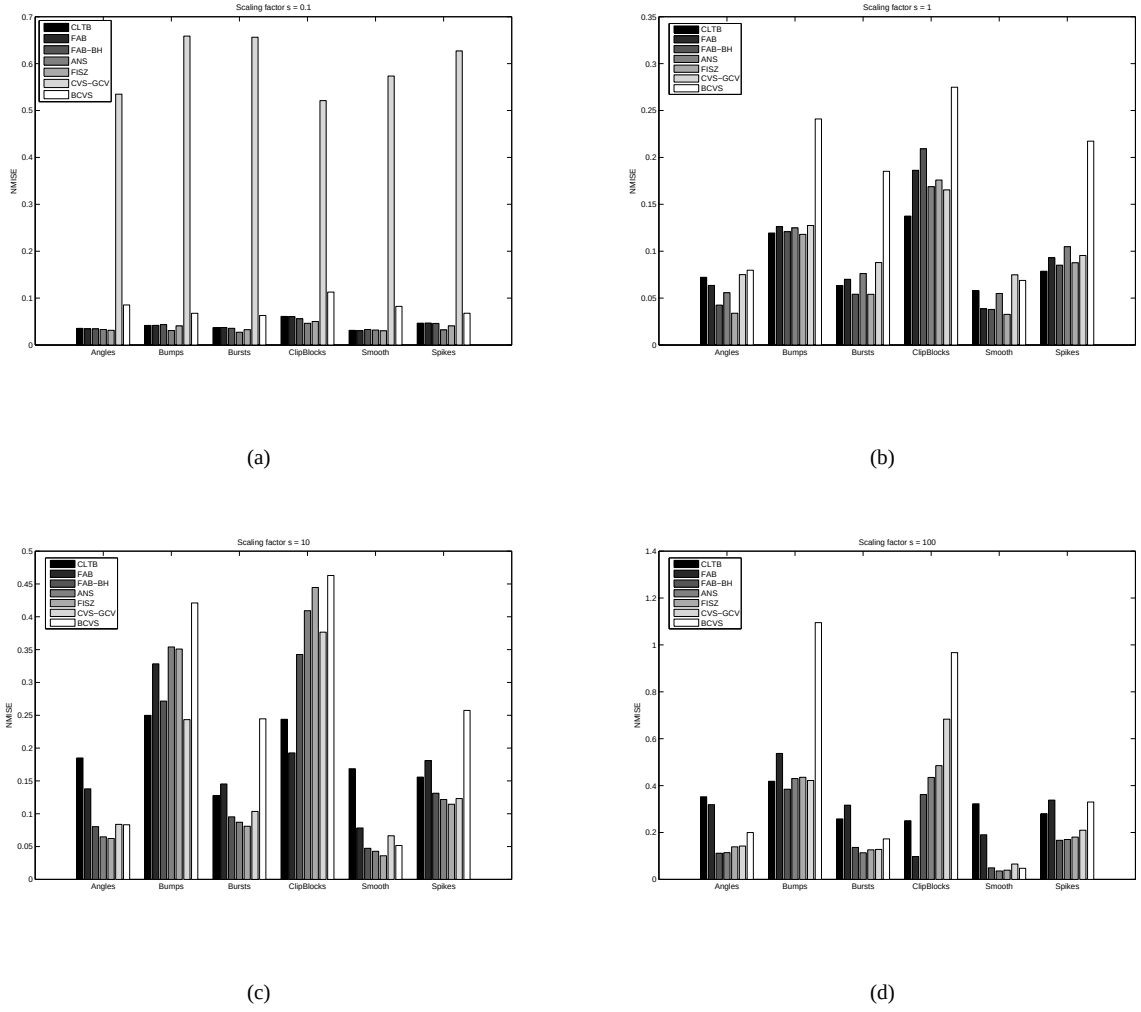


Fig. 4. NMISE of the estimators. In each group of bars, the bar from left to right (or from dark to bright) corresponds to the NMISE of CLTB, FAB, FAB-BH, ANS, FISZ, CVS-GCV and BCVS. Scaling factor (a) $s = 0.1$; (b) $s = 1$; (c) $s = 10$; (d) $s = 100$.

534, pp. 490–505, 2000.

[19] C. Charles and J. P. Rasson, "Wavelet denoising of Poisson-distributed data and applications," *Computational Statistics and Data Analysis*, vol. 43, no. 2, pp. 139–148, 2003.

[20] E. D. Kolaczyk, "Wavelet shrinkage estimation of certain Poisson intensity signals using corrected thresholds," *Statist. Sinica*, vol. 9, pp. 119–135, 1999.

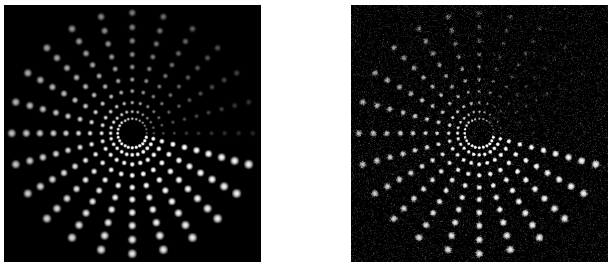
[21] R. G. Miller, *Simultaneous Statistical Inference*, 2nd ed. Springer-Verlag, 1981.

[22] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *J. Roy. Statist. Soc. ser. B*, vol. 57, no. 1, pp. 289–300, 1995.

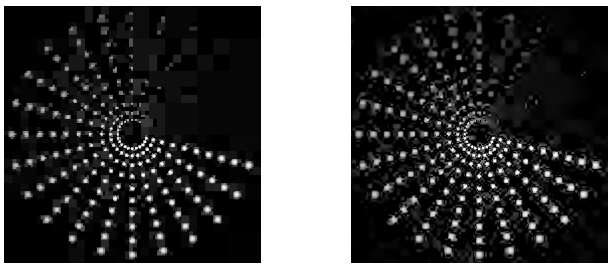
[23] Y. Benjamini and Y. Yekutieli, "The control of the false discovery rate in multiple testing under dependency," *Ann. Statist.*, vol. 29, no. 4, pp. 1165–1188, 2001.

[24] R. A. Fisher, *Contributions to mathematical statistics*. New York: Wiley, 1950.

[25] A. Bijaoui and G. Jammal, "On the distribution of the wavelet coefficient for a Poisson noise," *Signal Processing*, vol. 81, pp. 1789–1800, 2001.



(a) (b)

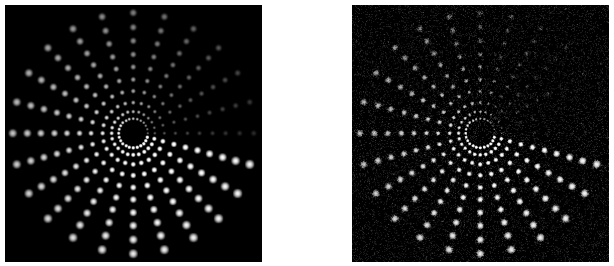


(c) (d)

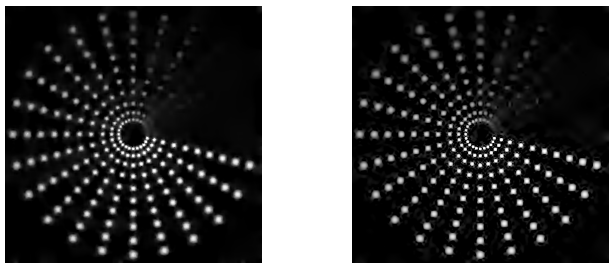
(e) (f)

(g) (h)

Fig. 5. XMM image denoising (shown with logarithmic intensity). Image size $= 512 \times 512$. $J = 7$. (a) intensity image $(\lambda_{min} = 0.05, \lambda_{max} = 200.05)$; (b) Poisson noisy image; (c) intensity estimated by CLTB where $\lambda_B = \lambda_{min}$; (d) intensity estimated by FAB-BH; (e)(g) local zoom of (c); (f)(h) local zoom of (d).



(a) (b)



(c) (d)

(e) (f)

(g) (h)

Fig. 6. XMM image TI denoising (shown with logarithmic intensity). Image size $= 512 \times 512$. $J = 7$. (a) intensity image ($\lambda_{min} = 0.05, \lambda_{max} = 200.05$); (b) Poisson noisy image; (c) intensity estimated by CLTB (TI) where $\lambda_B = \lambda_{min}$; (d) intensity estimated by FAB-BH (TI); (e)(g) local zoom of (c); (f)(h) local zoom of (d).

[26] G. de Castro, “Note on differences of Bernoulli and Poisson variables,” *Portugaliae Mathematica*, vol. 11, pp. 173–175, 1952.

[27] J. G. Skellman, “The frequency distribution of the difference between two Poisson variates belonging to different populations,” *J. Roy. Statist. Soc. ser. A*, vol. 109, p. 296, 1946.

[28] N. L. Johnson, “On an extension of the connexion between Poisson and χ^2-distributions,” *Biometrika*, vol. 46, pp. 352–363, 1959.

[29] P. B. Patnaik, “The non-central χ^2- and F-distributions and their applications,” *Biometrika*, vol. 36, pp. 202–232, 1949.

[30] M. R. Leadbetter, G. Lindgren, and H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes*. New York: Springer-Verlag, 1983.

[31] N. L. Johnson, S. Kots, and N. Balakrishnan, *Continuous Univariate Distributions*, 2nd ed. John Wiley & Sons, 1995, vol. 2.

[32] J. E. Hacke, “Solving the quartic,” *Amer. Math. Monthly*, vol. 48, pp. 327–328, 1941.

[33] M. Jansen, “Generalized Cross Validation for wavelet thresholding,” *Signal Processing*, vol. 56, no. 1, pp. 33–44, 1997.

[34] C. J. Miller, C. Genovese, R. C. Nichol, L. Wasserman, A. Connolly, D. Reichart, A. Hopkins, J. Schneider, and A. Moore, “Controlling the False Discovery Rate in Astrophysical Data Analysis,” *Astro-ph/0107034*, July 2001.

[35] F. Abramovich, Y. Benjamini, D. Donoho, and I. Johnstone, “Adapting to Unknown Sparsity by controlling the False Discovery Rate,” Dept. of Statistics, Stanford Universty, Tech. Rep. 2000-19, 2000.

[36] D. L. Donoho and R. R. Coifman, “Translation-Invariant De-Noising,” in *Wavelets and Statistics*, ser. Lecture Notes in Statistics, A. Antoniadis and G. Oppenheim, Eds. New York: Springer-Verlag, 1995, vol. 103.

[37] S. Durand and J. Forment, “Reconstruction of wavelet coefficients using total variation minimization,” *SIAM Journal on Scientific Computing*, vol. 24, no. 5, pp. 1754–1767, 2003.

[38] F. Malgouyres, “Minimizing the total variation under a general convex constraint for image restoration,” *ITIP*, vol. 11, no. 2, pp. 1450–1456, 2002.

[39] E. J. Candès and F. Guo, “New Multiscale Transforms, Minimum Total Variation Synthesis: Applications to Edge-Preserving Image Reconstruction,” *Signal Processing*, vol. 82, no. 5, pp. 1516–1543, 2002.

[40] J. L. Starck, “Very high quality image restoration by combining wavelets and curvelets,” *Proc. SPIE*, vol. 4478, pp. 9–19, Dec. 2001.

[41] J. L. Starck, J. Fadili, and F. Murtagh, “The Undecimated Wavelet Decomposition and its Reconstruction,” Submitted to Elsevier Science, Mar. 2005.

[42] J. Y. Ouvrard, *Probabilité*. Cassini, 2000, vol. 2, pp. 313–314.