Fast Poisson Noise Removal by Biorthogonal Haar Domain Hypothesis Testing

B. Zhang [a,*]

[a] *Quantitative Image Analysis Group URA CNRS 2582 of Institut Pasteur, 75724 Paris, France*

M. J. Fadili [b]

[b] *Image Processing Group GREYC CNRS UMR 6072, 14050 Caen Cedex, France*

J.-L. Starck [c]

[c] *DAPNIA/SEDI-SAP, Service d'Astrophysique, CEA-Saclay, 91191 Gif sur Yvette, France*

S. W. Digel [d]

[d] *Stanford Linear Accelerator Center, 2575 Sand Hill Road, Menlo Park, CA 94025*

Abstract

Methods based on hypothesis tests (HTs) in the Haar domain are widely used to denoise Poisson count data. However, facing large datasets or real-time applications, Haar-based denoisers have to use the decimated transform to meet limited-memory or computation-time constraints. Unfortunately, for regular underlying intensities, decimation yields discontinuous estimates and strong "staircase" artifacts. In this paper, we propose to combine the HT framework with the decimated biorthogonal Haar (Bi-Haar) transform instead of classical Haar. It is shown that the Bi-Haar coefficients converge asymptotically to Haar coefficients in distribution as the transform scale increases. The convergence rate also increases with the data dimension. Thus, we are allowed to directly apply the Haar-based HTs to Bi-Haar coefficients, especially when processing high-dimensional large datasets. By doing so, we benefit from the regular Bi-Haar filter bank to gain a smooth estimate while always maintaining a low computational complexity. The efficiency of this method is also illustrated on an example of hyperspectral-source-flux estimation.

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

* Corresponding author.

Email addresses: bzhang@pasteur.fr (B. Zhang), Jalal.Fadili@greyc.ensicaen.fr (M. J. Fadili), jstarck@cea.fr (J.-L. Starck), digel@slac.stanford.edu (S. W. Digel).

1. Introduction

Astronomical data analysis often requires Poisson noise removal [1]. This problem can be formulated as follows: we observe a discrete dataset of counts $\mathbf{v} = (v_i)_{i\in\mathbb{Z}^q}$ where v_i follows a Poisson distribution of intensity λ_i, i.e. $v_i \sim \mathcal{P}(\lambda_i)$. Here we suppose that v_i's are mutually independent. The denoising aims at estimating the underlying intensity profile $\Lambda = (\lambda_i)_{i\in\mathbb{Z}^q}$ from $\mathbf{v}$.

A host of estimation methods have been proposed in the literature (see the reviews [2][3] and their citations), among which an important family of approaches based on hypothesis tests (HTs) is widely used in astronomy [4,5][6]. These methods rely on Haar transform and the HTs are applied on the Haar coefficients to control a user-specified false positive rate (FPR). When working with large datasets or real-time applications, the decimated Haar transform is generally required to meet limited-memory or computation-time constraints. This is even more true when processing astronomical hyperspectral data, which are usually very large in practice. Unfortunately, for regular underlying intensities, decimation yields discontinuous estimates with strong "staircase" artifacts, thus significantly degrading the denoising performance. Although [7] and [8] attempted to generalize the HTs for wavelets other than Haar, [7] is more computationally complex than Haar-based methods, and [8] adopts an asymptotic approximation which may not allow reasonable solutions in low-count situations.

In this paper, we propose to combine the HT framework with the decimated bi-orthogonal Haar (Bi-Haar) transform. It will be shown that the Bi-Haar coefficients converge asymptotically to Haar coefficients in distribution as the transform scale increases. We will also see that the convergence rate increases with the data dimension q. This implies that we are allowed to directly apply the Haar-based HTs to Bi-Haar coefficients, especially when processing high-dimensional large datasets. By doing so, we benefit from the regular Bi-Haar filter bank to gain a smooth estimate. This approach even exhibits a performance comparable to the more time/space-consuming translation-invariant Haar (TI Haar or undecimated Haar) denoising in some of our experiments. Our method is also applied to source-flux estimation in astronomical hyperspectral data.

The paper is organized as follows. We begin with the review of the wavelet individual and multiple HTs in Section 2, and then Bi-Haar domain HT is presented in Section 2.3. Section 2.4 details some thresholding operators implementing individual tests. The final denoising algorithm is summarized in Section 2.5, and the numerical results are shown in Section 3. We conclude in Section 4, and the mathematical details are deferred to the appendices.

2. Hypothesis testing in the wavelet domain

Wavelet domain denoising can be achieved by zeroing insignificant coefficients while preserving significant ones. We detect significant coefficients by applying a binary HT on each wavelet coefficient d:

$$H_0 : d = 0 \text{ vs. } H_1 : d \neq 0$$

Note that since any wavelet has a zero mean, if d comes from a signal of constant intensity within the wavelet support, then $d \in H_0$.

2.1. *Individual and multiple hypothesis testing*

Individual HTs are carried out in a coefficient-by-coefficient manner. First, the user pre-specifies a FPR in the wavelet domain, say α. Then the p-value of each coefficient p_i is calculated under the null hypothesis H_0. Finally, all the coefficients with $p_i > \alpha$ will be zeroed.

If we desire to control global statistical error rates, multiple HTs should be used. For example, the Bonferroni overconservative correction controls the probability of erroneously rejecting even one of the true null hypothesis, i.e., Family-Wise Error Rate (FWER). To upper bound FWER by α, the same individual-testing procedure is performed as above but with FPR set to α/N, where N is the number of wavelet coefficients. Alternatively, one can carry out the Benjamini and Hochberg procedure [9] to control the false discovery rate (FDR), i.e., the average fraction of false detections over the total number of detections. The control of FDR has the following advantages over that of FWER: 1) it usually has a greater detection power; 2) it can easily handle correlated data [10]. The latter point allows the FDR control in non-orthogonal wavelet domains. Minimaxity of FDR has also been studied in various cases (see [11][12] for details).

2.2. *p-value of wavelet coefficients under H_0*

Both individual and multiple tests need to compute the p-value of each wavelet coefficient under H_0. Although the probability density function (*pdf*) of a H_0-coefficient is derived [7], this *pdf* has no closed form for a general wavelet. Thus the p-value evaluation in practice is computationally complex.

To obtain distributions of manageable forms, simple wavelets are preferred, such as Haar. To the best of our knowledge, Haar is the only wavelet yielding a closed-form *pdf*, which is given by [13] ($n \geq 0$): $\Pr(d = n; \lambda) = e^{-2\lambda} I_n(2\lambda)$, where $d = X_1 - X_2$, $X_1, X_2 \sim \mathcal{P}(\lambda)$, and I_n is the n-th order modified Bessel function of the first kind. For negative n, the probability can be obtained by symmetry. The tail probability (p-value) is given by [14]:

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

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

2.3. *Bi-Haar domain testing*

Haar wavelet provides us with a manageable distribution under H_0. But due to the lack of continuity of Haar filters, its estimate can be highly irregular with strong "staircase" artifacts when decimation is involved.

To solve this dilemma between distribution manageability and reconstruction regularity, we propose to use the Bi-Haar wavelet. The implementation filter bank (with a normalizing parameter c) is [1]:

$$h = 2^{-c}[1,1], \qquad g = 2^{-c}[\frac{1}{8}, \frac{1}{8}, -1, 1, -\frac{1}{8}, -\frac{1}{8}];$$
$$\tilde{h} = 2^{c-1}[-\frac{1}{8}, \frac{1}{8}, 1, 1, \frac{1}{8}, -\frac{1}{8}], \quad \tilde{g} = 2^{c-1}[1,-1]$$

where (h, g) and $(\tilde{h}, \tilde{g})$ are respectively the analysis and synthesis filter banks. For comparison, the Haar filter bank is $(h = 2^{-c}[1,1], g = 2^{-c}[-1,1], \tilde{h} = 2^{c-1}[1,1], \tilde{g} = 2^{c-1}[1,-1])$. It follows that the synthesis Haar scaling function is discontinuous while that of Bi-Haar is almost Lipschitz [15][16]. Hence, the Bi-Haar reconstruction will be smoother. Furthermore, the following theorem shows that Bi-Haar coefficients converge asymptotically to Haar coefficients.

Theorem 1 *Let d_j^{bh} and d_j^h denote respectively the Bi-Haar and Haar coefficients at scale j. For any dimensional data, if $c > 1$ then:*

$$D_j := d_j^{bh} - d_j^h \xrightarrow{\mathcal{D}} 0, \quad j \to +\infty \tag{2}$$

In the one-dimensional (1D) case, Proposition 2 provides an upper-bound for the convergence rate of (2) in terms of the deviation on the cumulative distribution functions (*cdf*). We note the *cdf* of D_j (by (A.1) and (1)):

$$F_{D_j}(x) = \Pr(D_j \leq x) = \begin{cases} 1 - F_{\chi^2_{2\lfloor 2^{cj+3}x \rfloor + 2}}(2^{j+1}\lambda), & x \geq 0 \\ F_{\chi^2_{2\lceil -2^{cj+3}x \rceil}}(2^{j+1}\lambda), & x < 0 \end{cases} \tag{3}$$

Proposition 2 *Define the Heaviside function $H(x) = 0$ if $x < 0$ and $H(x) = 1$ otherwise. We have for $q = 1$:*

– ***(Uniform bound)*** *For any $x_0 > 0$, we have (4) uniformly for all $|x| \geq x_0$,*

$$|F_{D_j}(x) - H(x)| \leq \begin{cases} \exp(-n\alpha_n), & x \geq x_0 \\ \exp(-m\alpha_m), & x \leq -x_0 \end{cases} \tag{4}$$

where $n = 2\lfloor 2^{cj+3}x_0 \rfloor + 2$, $m = 2\lceil 2^{cj+3}x_0 \rceil$, $\Delta = 2^{j+1}\lambda$, $\eta_k = \Delta/k$, and

$$\alpha_k = \eta_k - \sqrt{\eta_k^2 + \frac{1}{4}} + \frac{1}{2}\ln\left(\frac{1 + \sqrt{1 + 4\eta_k^2}}{2\eta_k}\right) > 0 \tag{5}$$

– ***(Asymptotic uniform bound)*** *If $c > 1$, then (4) holds with $\alpha_k = O_{j \to +\infty}(j)$ in (5).*

The proof of Theorem 1 in Appendix A also implies that the convergence rate increases as q becomes higher. Theorem 1 and Proposition 2 justify that for $c > 1$, the p-value of Bi-Haar coefficients can be approximated by that of Haar, i.e., by (1).

2.4. *Thresholds for individual tests*

For individual tests controlling the FPR, the HTs can be implemented by thresholding operators. In other words, one can find t_j such that $\Pr(|d_j^{bh}| \geq t_j | H_0) = \alpha$ where α represents the controlled FPR. Owing to the results in Section 2.3, we can now consider the Haar case and the derived t_j will be directly applicable for Bi-Haar coefficients. We point out that to simplify the presentation, t_j is supposed to be scale-dependent only,

but scale *and* location-dependent thresholds can be derived using the same procedure presented below.

2.4.1. *CLTB threshold [4,8,5,6]*

The Haar coefficient can be written as $d_j^h = 2^{-cjq}(X_1 - X_2)$ where $X_1, X_2 \sim \mathcal{P}(\lambda_j/2)$ are independent. It follows from (1) that:

$$\Pr(d_j^h \geq t_j | H_0) = \Pr\left(\chi^2_{(2m_j)}(\lambda_j) < \lambda_j\right) \approx \Pr(\gamma\chi^2_{(f)} < \lambda_j) \tag{6}$$

$$\approx \Pr\left(Z > \frac{f - \lambda_j/\gamma}{\sqrt{2f}}\right) \tag{7}$$

where $m_j = 2^{cjq}t_j$, $\gamma = (2m_j + 2\lambda_j)/(2m_j + \lambda_j)$, $f = (2m_j + \lambda_j)^2/(2m_j + 2\lambda_j)$, $\chi^2_{(v)}$ is a central chi-square variable and $Z \sim \mathcal{N}(0,1)$. Here, two stages of approximation are used: 1) the non-central chi-square distribution is first approximated by a central one (6) [17]; 2) the central chi-square variable is then approximated by a normal one (7) using the central limit theorem (CLT). t_j is thus called the CLT-based (CLTB) threshold. Consequently, it remains to solve the equation (7) $= \alpha/2$, and the solution is given by:

$$t_j = 2^{-cjq-1}\left(z^2_{\alpha/2} + \sqrt{z^4_{\alpha/2} + 4 \cdot \lambda_j z^2_{\alpha/2}}\right) \tag{8}$$

where $z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$, and Φ is the standard normal *cdf*. Universal threshold can also be obtained by setting $z_{\alpha/2} = \sqrt{2\ln N_j}$ in (8) where N_j is the total number of coefficients in one band at scale j.

2.4.2. *FAB threshold*

An improvement of CLTB threshold can be achieved by replacing (7) with an approximation of faster convergence, e.g., the following one proposed by Fisher [18]:

$$\sqrt{2\chi^2_{(f)}} \to \mathcal{N}(\sqrt{2f-1}, 1), \quad f \to \infty \tag{9}$$

Therefore, (7) is changed to:

$$\Pr\left(\gamma\chi^2_{(f)} < \lambda_j\right) \approx \Pr\left(Z > \sqrt{2f-1} - \sqrt{\frac{2\lambda_j}{\gamma}}\right) \tag{10}$$

Let us denote:

$$G(m_j) := \sqrt{2f-1} - \sqrt{\frac{2\lambda_j}{\gamma}} = \sqrt{\frac{(2m_j + \lambda_j)^2}{m_j + \lambda_j} - 1} - \sqrt{\frac{\lambda_j(2m_j + \lambda_j)}{m_j + \lambda_j}} \tag{11}$$

It remains to solve $G(m_j) = z_{\alpha/2}$, which leads to a quartic equation in m_j:

$$16m_j^4 + \left[16\lambda_j - 8(z^2_{\alpha/2} + 1)\right] m_j^3 + \left[(z^2_{\alpha/2} + 1)^2 - (20z^2_{\alpha/2} + 12)\lambda_j + 4\lambda_j^2\right] m_j^2$$
$$+ \left[2(z^2_{\alpha/2} + 1)^2\lambda_j - 16z^2_{\alpha/2}\lambda_j^2 - 4\lambda_j^2\right] m_j + (z^2_{\alpha/2} + 1)^2\lambda_j^2 - 4z^2_{\alpha/2}\lambda_j^3 = 0 \tag{12}$$

The final Fisher-approximation-based (FAB) threshold t_j is obtained from m_j^*, the solution of (12). Owing to the following results, we do not need to write out the explicit expression of m_j^*, which could be rather complex:

Proposition 3 *The feasible condition for m_j is given by (13), and the feasible solution m_j^* exists and is unique.*

$$m_j \geq \frac{1}{8}\left[z_{\alpha/2}^2 - 2\lambda_j + 1 + \left(z_{\alpha/2}^4 + (12\lambda_j + 2)z_{\alpha/2}^2 + 4\lambda_j^2 + 12\lambda_j + 1\right)^{1/2}\right] \tag{13}$$

Proposition 3 implies that we can use any numerical quartic-equation solver, e.g. Hacke's method [19], to find the four solutions of (12). One and only one of the solutions will satisfy (13), which is m_j^*. The universal threshold can also be derived in the same way as in the CLTB case.

2.5. *Summary of the denoising algorithm*

Note that the thresholds t_j (more generally the p-value of d_j^h) depend on the background rate at scale j (i.e. λ_j). Without any prior knowledge, it can be estimated by the values of the approximation coefficients at scale $j+1$ (i.e. a_{j+1}). Here, the wavelet denoising should be carried out in a coarse-to-fine manner, outlined as follows:

Algorithm 1 Poisson noise removal by HTs in the Bi-Haar domain

1: Bi-Haar transform of $\mathbf{v}$ up to $j = J$ to obtain a_J and d_j^{bh} $(1 \leq j \leq J)$
2: **for** $j = J$ down to 1 **do**
3: $\quad\hat{\lambda}_j = 2^{jq}\lambda$ if λ is known; otherwise $\hat{\lambda}_j = \max(2^{cjq}a_j, 0)$
4: $\quad$Hypothesis testing on d_j^{bh} by either computing p-values or applying thresholds
5: $\quad$Reconstruct a_{j-1} by inverse Bi-Haar transform
6: **end for**
7: Positivity projection: $\hat{\Lambda} = \max(a_0, 0)$

3. Results

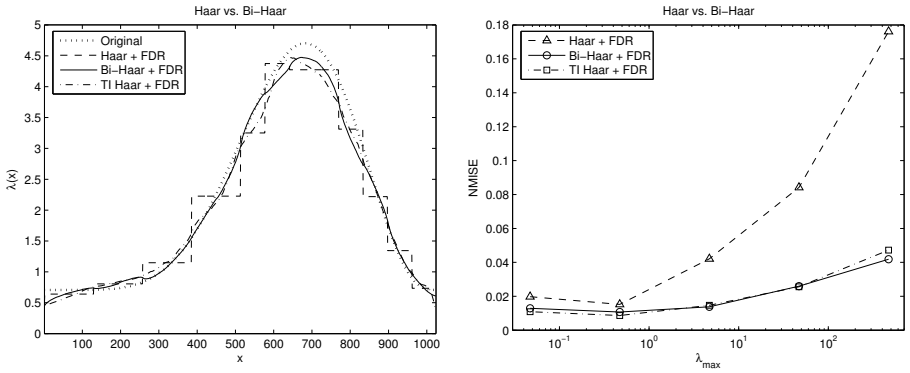
3.1. *Haar vs. Bi-Haar denoising for regular intensities*

To compare Haar and Bi-Haar denoising for regular intensities, we generate noisy signals from the "Smooth" function [2] (see Fig.1(a)) and measure the Normalized Mean Integrated Square Error (NMISE) per bin from the denoised signals. The NMISE is defined as: NMISE $:= \mathbb{E}[(\sum_{i=1}^N(\hat{\lambda}_i - \lambda_i)^2/\lambda_i)/N]$, where $(\hat{\lambda}_i)_i$ is the intensity estimate. Note that the denominator λ_i plays the role of variance stabilization in the error measure.

Fig.1(a) shows the denoising examples given by Haar, Bi-Haar and TI Haar estimations, where FDR tests are applied. The original intensity function is scaled to cover a wide range of intensities, and Fig.1(b) compares the NMISEs (measured from 100 replications) of the three estimators as functions of the underlying peak intensity.

It can be seen that the Bi-Haar estimate is much more regular than the Haar one, and is even almost as good as TI Haar at every intensity level under the NMISE criterion.

This surprising performance is gained with the same complexity as in the Haar denoising, i.e., $O(N)$ only, as opposed to $O(N \log N)$ in the TI Haar case.



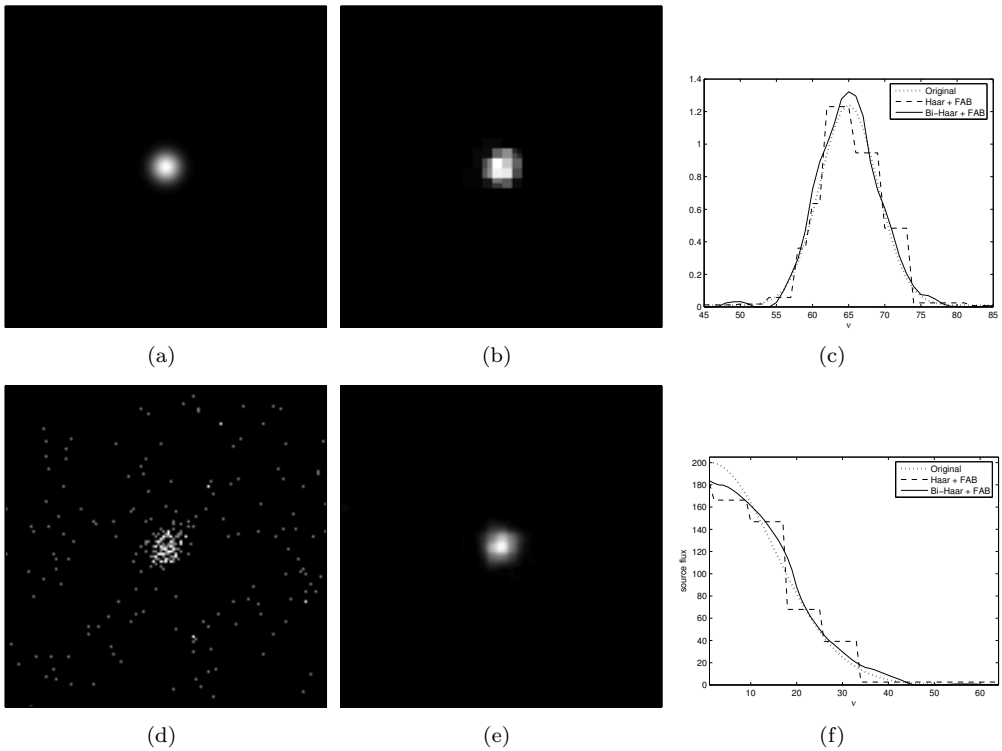
(a) (b)

Fig. 1. Denoising the "Smooth" function (length = 1024). Estimates from Haar, Bi-Haar and TI Haar (undecimated) are compared. FDR is set to 0.05 and $J = 7$. (a) denoising results; (b) NMISEs.

3.2. *Source-flux estimation in astronomical hyperspectral data*

We apply our method to source-flux estimation in astronomical hyperspectral images. A hyperspectral image $\mathbf{v}(x, y, \nu)$ is a "2D+1D" volume, where x and y define the spatial coordinates and ν indexes the spectral band. Each bin records the detected number of photons. As the three axes of our data have different physical meanings, we are motivated to apply a "2D+1D" wavelet transform instead of using the classical 3D transform. That is, we first carry out a complete 2D wavelet transform for spatial planes, and then a 1D transform along the spectral direction. We use j_{xy} and j_ν to denote the j-th spatial scale and the j-th spectral scale, respectively. It can be shown that the convergence result (2) still holds as j_{xy} and j_ν increase. Hyperspectral data in practice can be very large, implying that fast denoising is only possible with decimated transforms (the execution time of the example below on a P4 2.8GHz PC is 13s for our Bi-Haar denoising, i.e., more than 50 times faster than the TI Haar denoising (688s)). Let alone the memory space required by the TI transform.

Our simulated data contain a source having a Gaussian profile. The source amplitude A_ν decreases from 2 to 10^{-4} as ν increases. One example band is shown in Fig.2(a). The observed counts is presented in Fig.2(d). The denoising results using Haar and Bi-Haar transforms are respectively shown in Fig.2(b) and (e). Here, the FAB thresholding is applied. Fig.2(c) illustrates the estimation smoothness gained by Bi-Haar by comparing a line profile of the estimated source from different methods. In hyperspectral imaging, the source flux $S(\nu)$ is an important quantity, which equals to the integral of the source intensity over its spatial support at band ν. Fig.2(f) compares the flux given by different denoisers. Clearly, the Haar-based approach leads to a piecewise constant estimate, whereas Bi-Haar provides a regular flux which is more accurate: the L^2-errors for Haar and Bi-Haar flux estimates, i.e. $\|\hat{S} - S\|_{L^2}$, are 14.4 and 7.4 respectively.



(a) (b) (c)

(d) (e) (f)

Fig. 2. Source-flux estimation in a hyperspectral image (size: $129 \times 129 \times 64$). $A_\nu \in [10^{-4}, 2]$; $J_{xy} = 3$, $J_\nu = 5$, FAB thresholding with $\alpha = 10^{-5}$. (a) intensities at $\nu = 15$; (b) Haar-denoised data ($\nu = 15$), $\|\hat{\Lambda} - \Lambda\|_{L^2} = 0.017$; (c) estimated source profile at $\nu = 15$ (intensity along a line passing through the source center); (d) Poisson count image; (e) Bi-Haar-denoised data ($\nu = 15$), $\|\hat{\Lambda} - \Lambda\|_{L^2} = 0.011$; (f) estimated flux (Respectively for Haar and Bi-Haar estimates: $\|\hat{S} - S\|_{L^2} = 14.4$ and 7.4.)

4. Conclusion

In this paper, we proposed to combine the HT framework with the decimated Bi-Haar transform instead of the classical Haar for denoising large datasets of Poisson counts. We showed that the Haar-based HTs can be directly applied to Bi-Haar coefficients. By doing so, we benefit from the regular Bi-Haar filter bank to gain a smooth estimate with no "staircase" artifacts, while always maintaining a low computational complexity. This approach could be extended in the future to fast deconvolution of Poisson data.

Appendix A. Proof of Theorem 1

The following lemma is needed, which can be proven using Lévy's continuity theorem.

Lemma 4 *$\kappa_{n,p}$ and $\kappa_{0,p}$ denote the p-th cumulant of $(X_n)_{n\geq 1}$ and of X_0 respectively. If $\sup_{p\geq 1} |\kappa_{n,p} - \kappa_{0,p}| \to 0$ as $n \to \infty$, then $X_n \xrightarrow{\mathcal{D}} X_0$.*

We will now prove Theorem 1.

PROOF. We first show the 1D case. Our signal is a sequence of independent Poisson counts $(v_n)_n$. The approximation coefficients of Haar and Bi-Haar transforms are identical. D_j is given by:

$$D_j = 2^{-cj}(Z_1 - Z_2 + 2^{-3}(Z_3 - Z_4)) - 2^{-cj}(Z_1 - Z_2) = 2^{-cj-3}(Z_3 - Z_4) \qquad \text{(A.1)}$$

where Z_1 and Z_2 are the sums of 2^{j-1} independent variables of $(v_n)_n$, and Z_3 and Z_4 are the sums of 2^j independent variables. Moreover, Z_1, Z_2, Z_3 and Z_4 are mutually independent.

We have for $\kappa_p^{D_j}$ i.e. the p-th cumulant of D_j ($p \geq 1$):

$$|\kappa_p^{D_j}| = 2^{-cpj-3p}|\kappa_p^{Z_3} + (-1)^p \kappa_p^{Z_4}| \leq 2^{j(1-cp)-3p+1}\lambda_{max} =: r \qquad \text{(A.2)}$$

where λ_{max} is the maximum intensity of $(v_n)_n$. If $c > 1$ then $r \to 0$ with an exponential decreasing rate for any p as j increases. We conclude by applying Lemma 4.

For $q = 2$, we proceed in the same manner as in the 1D case. We have for the p-th cumulants of D_j at band hg, gh and gg:

$$|\kappa_p^{D_{j,hg}}| \leq 2^{2j(1-cp)-3p+1}\lambda_{max}, \ |\kappa_p^{D_{j,gh}}| \leq 2^{2j(1-cp)-3p+1}\lambda_{max}$$
$$|\kappa_p^{D_{j,gg}}| \leq 2^{2j(1-cp)-3p+2}(1 + 8^{-p})\lambda_{max}$$

Notice that we have a higher convergence rate in 2D than in 1D. It is clear that we can continue this calculation for $q > 2$, and we conclude always by the same arguments as in the 1D case.

Appendix B. Proof of Proposition 2

PROOF. We will prove the case of $x \geq x_0$ and the case $x \leq -x_0$ can be established similarly. A non-central chi-square variable $\chi_n^2(\Delta)$ can be written as a sum of i.i.d. normal variables:

$$\chi_n^2(\Delta) = \sum_{i=1}^{n} Y_i, \quad Y_i := X_i^2, \quad X_i \sim \mathcal{N}(\mu, 1), \quad \mu = \sqrt{\frac{\Delta}{n}}$$

Here, we have that $n = 2\lfloor 2^{cj+3}x \rfloor + 2$ and $\Delta = 2^{j+1}\lambda$. Define $y := 2^{j+1}\lambda/n = \Delta/n = \mu^2$. The moment generating function of Y_i is given by:

$$M(t) := \mathbb{E}(e^{tY_i}) = \exp\left(\frac{\mu^2 t}{1-2t}\right)/\sqrt{1-2t}$$

We define:

$$I_Y(y) := \sup_{t \in \mathbb{R}}(ty - \ln M(t)) = \frac{2y - \sqrt{4y^2+1}}{2} + \frac{1}{2}\ln\left(\frac{1+\sqrt{1+4y^2}}{2y}\right)$$

Since $y < \mathbb{E}(Y_i) = y + 1$, it can be shown that $I_Y(y) > 0$. Note that $y = \eta_n$, so $I_Y(y) = I_Y(\eta_n) = \alpha_n$. Then, the uniform bound is obtained by applying Cramér-Chernoff theorem [20] and by noticing that α_n is an increasing function of $n(> 0)$. The asymptotic behavior of α_n is obtained by setting $j \to +\infty$ with the assumption $c > 1$.

Appendix C. Proof of Proposition 3

PROOF. The facts that $G(m_j) = z_{\alpha/2}$, $z_{\alpha/2} > 0$, $2f - 1 \geq 0$, $m_j > 0$ and $\lambda_j \geq 0$ show (13).

Next, when the equality in (13) holds, we have:

$$G(m_j) = \sqrt{z_{\alpha/2}^2 + \frac{\lambda_j}{2m_j}(z_{\alpha/2}^2 + 1)} - \sqrt{\frac{\lambda_j}{2m_j}(z_{\alpha/2}^2 + 1)} \leq z_{\alpha/2}$$

The existence and uniqueness of the feasible solution follow from the fact that G is a strictly increasing function under (13), and that $G(m_j) \to +\infty$ as $m_j \to +\infty$.

References

[1] J.-L. Starck, F. Murtagh, and A. Bijaoui. *Image Processing and Data Analysis: The Multiscale Approach*. Cambridge University Press, 1998.

[2] P. Besbeas, I. De Feis, and T. Sapatinas. A Comparative Simulation Study of Wavelet Shrinkage Estimators for Poisson Counts. *Internat. Statist. Rev.*, 72(2):209–237, 2004.

[3] R. Willett. Multiscale Analysis of Photon-Limited Astronomical Images. *SCMA IV*, 2006. in press.

[4] E. D. Kolaczyk. Nonparametric Estimation of Gamma-Ray Burst Intensities Using Haar Wavelets. *The Astrophysical Journal*, 483:340–349, 1997.

[5] E. D. Kolaczyk. Nonparametric estimation of intensity maps using Haar wavelets and Poisson noise characteristics. *The Astrophysical Journal*, 534:490–505, 2000.

[6] C. Charles and J. P. Rasson. Wavelet denoising of Poisson-distributed data and applications. *Computational Statistics and Data Analysis*, 43(2):139–148, 2003.

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

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

[9] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. ser. B*, 57(1):289–300, 1995.

[10] Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.*, 29(4):1165–1188, 2001.

[11] F. Abramovich, Y. Benjamini, D. Donoho, and I. Johnstone. Adapting to Unknown Sparsity by controlling the False Discovery Rate. *Annals of Statistics*, 34(2):584–653, 2006.

[12] D. L. Donoho and J. Jin. Asymptotic Minimaxity of FDR for Sparse Exponential Model. *Annals of Statistics*, 2006. To appear.

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

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

[15] J. D. Villasenor, B. Belzer, and J. Liao. Wavelet filter evaluation for image compression. *IEEE Transactions on Image Processing*, 4(8):1053–1060, 1995.

[16] O. Rioul. Simple regularity criteria for subdivision schemes. *SIAM Journal on Mathematical Analysis*, 23(6):1544–1576, 1992.

[17] P. B. Patnaik. The non-central χ^2- and F-distributions and their applications. *Biometrika*, 36:202–232, 1949.

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

[19] J. E. Hacke. Solving the quartic. *Amer. Math. Monthly*, 48:327–328, 1941.

[20] S. R. S. Varadhan. *Large Deviations and Applications*. SIAM, Philadelphia, 1984.