Fast Poisson noise removal by biorthogonal Haar domain hypothesis testing

B. Zhang [a,*], M.J. Fadili [b], J.-L. Starck [c], S.W. Digel [d]

[a] *Quantitative Image Analysis Group URA CNRS 2582 of Institut Pasteur, 75724 Paris, France*
[b] *Image Processing Group GREYC CNRS UMR 6072, 14050 Caen Cedex, France*
[c] *DAPNIA/SEDI-SAP, Service d'Astrophysique, CEA-Saclay, 91191 Gif sur Yvette, France*
[d] *Stanford Linear Accelerator Center, 2575 Sand Hill Road, Menlo Park, CA 94025, United States*

Abstract

Methods based on hypothesis tests (HTs) in the Haar domain are widely used to denoise Poisson count data. 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 the classical Haar. The Bi-Haar filter bank is normalized such that the p-values of Bi-Haar coefficients (p_{BH}) provide good approximation to those of Haar (p_H) for high-intensity settings or large scales; for low-intensity settings and small scales, we show that p_{BH} are essentially upper-bounded by p_H. Thus, we may apply the Haar-based HTs to Bi-Haar coefficients to control a prefixed false positive rate. By doing so, we benefit from the regular Bi-Haar filter bank to gain a smooth estimate while always maintaining a low computational complexity. A Fisher-approximation-based threshold implementing the HTs is also established. The efficiency of this method is illustrated on an example of hyperspectral-source-flux estimation.

© 2008 Elsevier B.V. All rights reserved.

Keywords: Poisson intensity estimation; Biorthogonal Haar wavelets; Wavelet hypothesis testing; Fisher approximation

* Corresponding author. Tel.: +33 (0) 140613891; fax: +33 (0) 140613330.
E-mail 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 q-dimensional (qD) 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 [4–6]. 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 [7,8] 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 an astronomical image decompression context, [9] has also proposed to remove Haar block artifacts by minimizing at each resolution level the ℓ^2-norm of the Laplacian of the solution under some constraints on its wavelet coefficients. It has been shown that this approach was efficient in removing the artifacts, but it requires solving J minimization problems, where J is the number of scales. This can be quite time-consuming and would limit the interest in using Haar for large-dataset analysis.

In this paper, we propose to combine the HT framework with the decimated biorthogonal Haar (Bi-Haar) transform. The Bi-Haar filter bank is normalized such that the p-values of Bi-Haar coefficients (p_{BH}) approximate those of Haar (p_H) for high-intensity settings or large scales; for low-intensity settings and small scales, we show that p_{BH} are essentially upper-bounded by p_H. Thus, we may apply the Haar-based HTs to Bi-Haar coefficients to control a prefixed FPR. By doing so, we benefit from the regular Bi-Haar filter bank to gain a smooth estimate. A Fisher-approximation-based threshold implementing the HTs is also established. We find that 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. The efficiency of this method is also illustrated on an example of hyperspectral-source-flux estimation.

The paper is organized as follows. We begin with the review of the wavelet HTs in Section 2, and then Bi-Haar domain tests are presented in Section 2.2. Section 2.3 details some thresholding operators implementing the tests. The final denoising algorithm is summarized in Section 2.4, 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 $d \in H_0$.

Individual HTs are commonly used to control a user pre-specified FPR in the wavelet domain, say α. The tests are carried out in a coefficient-by-coefficient manner. That is, the p-value of each coefficient p_i is calculated under the null hypothesis H_0. Then, all the coefficients with $p_i > \alpha$ will be zeroed. If we desire to control global statistical error rates, multiple HTs may be adopted such as Bonferroni correction which controls the Family-Wise Error Rate (FWER), and the Benjamini and Hochberg procedure [10,11] controlling the false discovery rate (FDR).

2.1. p-values of wavelet coefficients under H_0

To carry out HTs, we need to compute the p-value of each wavelet coefficient under H_0. Although the probability density function (*pdf*) of a H_0-coefficient has been derived in [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 [12] ($n \geq 0$): $\Pr(d = n; \lambda) = e^{-2\lambda} I_n(2\lambda)$, where $d = X_1 - X_2$, $X_1, X_2 \sim \mathcal{P}(\lambda)$, and I_n is the nth 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 [13]:

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

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

2.2. 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. Its implementation filter bank is given by [1]:

$$h = 2^{-c}[1, 1], \qquad g = 2^{-c} r \left[\frac{1}{8}, \frac{1}{8}, -1, 1, -\frac{1}{8}, -\frac{1}{8} \right];$$

$$\tilde{h} = 2^{c-1} r \left[-\frac{1}{8}, \frac{1}{8}, 1, 1, \frac{1}{8}, -\frac{1}{8} \right], \qquad \tilde{g} = 2^{c-1}[1, -1]$$

where c and $r = (1 + 2^{-5})^{-1/2}$ are normalizing factors, (h, g) and $(\tilde{h}, \tilde{g})$ are respectively the analysis and synthesis filter banks. Note that our Bi-Haar filter bank has an unusual normalization. The motivation behind this is to ensure that the Bi-Haar coefficients will have the same variance as the Haar ones at each scale. Let us also point out that to correct for the introduction of the factor r, the Bi-Haar coefficients must be multiplied by r^{-1} at each stage of the recursive reconstruction. 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 [14,15]. Hence, the Bi-Haar reconstruction will be smoother.

Table 1
p_H and p_{BH}

$\lambda_j = 10^{-1}$		$\lambda_j = 10^0$		$\lambda_j = 10^1$		$\lambda_j = 10^2$	
$k_0 = 2$	$(1.15 \times 10^{-3},$ $1.17 \times 10^{-4})$	$k_0 = 4$	$(1.12 \times 10^{-3},$ $4.57 \times 10^{-4})$	$k_0 = 9$	$(3.97 \times 10^{-3},$ $2.48 \times 10^{-3})$	$k_0 = 20$	$(2.56 \times 10^{-2},$ $2.28 \times 10^{-2})$
$k_0 = 3$	$(1.91 \times 10^{-5},$ $1.87 \times 10^{-6})$	$k_0 = 5$	$(1.09 \times 10^{-4},$ $4.34 \times 10^{-5})$	$k_0 = 12$	$(2.12 \times 10^{-4},$ $1.26 \times 10^{-4})$	$k_0 = 30$	$(1.62 \times 10^{-3},$ $1.39 \times 10^{-3})$
$k_0 = 4$	$(2.38 \times 10^{-7},$ $2.28 \times 10^{-8})$	$k_0 = 6$	$(8.90 \times 10^{-6},$ $3.49 \times 10^{-6})$	$k_0 = 15$	$(6.60 \times 10^{-6},$ $3.78 \times 10^{-6})$	$k_0 = 40$	$(4.22 \times 10^{-5},$ $3.52 \times 10^{-5})$

Every parenthesis shows (p_H, p_{BH}) for 1D signals, where we always observe that $p_{BH} < p_H$.

At scale $j \geq 1$, let us define $\lambda_j = 2^j \lambda$ where λ is the underlying constant intensity. Then, a Haar coefficient can be written as $d_j^h = 2^{-cj}(X_1 - X_2)$ where $X_1, X_2 \sim \mathcal{P}(\lambda_j/2)$ are independent. We note $p_H := \Pr(d_j^h \geq 2^{-cj} k_0 | H_0)$ to be the p-value of a Haar coefficient where $k_0 = 1, 2, \ldots$. Accordingly, a Bi-Haar coefficient can be written as $d_j^{bh} = 2^{-cj} r(X_3 - X_4 + \frac{1}{8}(X_1 - X_2))$, where $X_1, X_2 \sim \mathcal{P}(\lambda_j)$ and $X_3, X_4 \sim \mathcal{P}(\lambda_j/2)$ are all independent. We note $p_{BH} := \Pr(d_j^{bh} \geq 2^{-cj} k_0 | H_0)$ to be the p-value of a Bi-Haar coefficient at the same critical threshold as for p_H. These definitions can be extended to higher dimensions ($q > 1$) straightforwardly.

For high-intensity settings or for large scales, d_j^h and d_j^{bh} will be asymptotically normal with the same asymptotic variances $\sigma_h^2 = \sigma_{bh}^2 = 2^{qj(1-2c)}\lambda$ due to the normalized filter banks. Thereby, they will have asymptotically equivalent tail probabilities, i.e., $p_{BH} \approx p_H$.

For low-intensity settings ($\lambda \ll 1$) and small scales, the following proposition (proof in Appendix A) shows for 1D signals that p_{BH} is essentially upper-bounded by p_H under H_0. The bounds for multi-dimensional data ($q > 1$) are also studied in Appendix A.

Proposition 1. *We have the following upper-bound for 1D signals*

$$p_{BH} \leq p_H + A(\lambda_j)(1 - 2p_H) \tag{2}$$

where

$$A(\lambda_j) = \frac{1}{2}\left[1 - e^{-2\lambda_j}\left(I_0(2\lambda_j) + 2\sum_{m=1}^{8} I_m(2\lambda_j)\right)\right].$$

As $\lambda \to 0+$, $A(\lambda_j) = \frac{2^{9j-7}}{2835}\lambda^9 + o(\lambda^9)$.

This theoretical bound is clearly confirmed by the numerical simulations shown in Table 1. Here we show the results for $\lambda_j \in [10^{-1}, 10^2]$ and different critical thresholds k_0 at the tails of the distributions. We indeed observe that p_{BH} is always strictly smaller than p_H.

2.3. Thresholds controlling FPR

For individual tests controlling FPR, the HTs can be implemented by thresholding operators. In other words, one can find $\tilde{t}_j$ such that $\Pr(|d_j^{bh}| \geq \tilde{t}_j | H_0) \leq \alpha$ where α represents the controlled FPR. Now consider the Haar case and suppose that we have derived the Haar threshold t_j under the controlled FPR. Then, by setting $\tilde{t}_j := 2^{-cjq}\lceil 2^{cjq} t_j \rceil$ the results in Section 2.2 allow us to conclude that the FPR for a Bi-Haar test will always be upper-bounded by α. We point out that

to simplify the presentation, t_j and $\tilde{t}_j$ are supposed to be scale-dependent only, but scale *and* location-dependent thresholds can be derived using the same procedure presented below.

2.3.1. CLTB threshold [4,8,5,6]

The Haar coefficient for qD data 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:

$$\Pr(d_j^h \geq t_j | H_0) = \Pr\left(\chi_{(2m_j)}^2(\lambda_j) < \lambda_j\right) \approx \Pr(\gamma \chi_{(f)}^2 < \lambda_j) \tag{3}$$

$$\approx \Pr\left(Z > \frac{f - \lambda_j/\gamma}{\sqrt{2f}}\right) \tag{4}$$

where $m_j = 2^{cjq}t_j$, $\gamma = (2m_j + 2\lambda_j)/(2m_j + \lambda_j)$, $f = (2m_j + \lambda_j)^2/(2m_j + 2\lambda_j)$, $\chi_{(v)}^2$ 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 (3) [16]; (2) the central chi-square variable is then approximated by a normal one (4) using the central limit theorem (CLT). t_j is thus called the CLT-based (CLTB) threshold. Consequently, it remains to solve the equation (4)$= \alpha/2$, and the solution is given by:

$$t_j = 2^{-cjq-1}(z_{\alpha/2}^2 + \sqrt{z_{\alpha/2}^4 + 4 \cdot \lambda_j z_{\alpha/2}^2}) \tag{5}$$

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 (5) where N_j is the total number of coefficients in one band at scale j.

2.3.2. FAB threshold

An improvement of CLTB threshold can be achieved by replacing (4) with an approximation of faster convergence, e.g., the following one proposed by Fisher [17]:

$$\sqrt{2\chi_{(f)}^2} \to \mathcal{N}(\sqrt{2f - 1}, 1), \quad f \to \infty. \tag{6}$$

Therefore, (4) is changed to:

$$\Pr\left(\gamma \chi_{(f)}^2 < \lambda_j\right) \approx \Pr\left(Z > \sqrt{2f - 1} - \sqrt{\frac{2\lambda_j}{\gamma}}\right). \tag{7}$$

Let us denote:

$$G(m_j) := \sqrt{2f - 1} - \sqrt{\frac{2\lambda_j}{\gamma}} = \sqrt{\frac{(2m_j + \lambda_j)^2}{m_j + \lambda_j} - 1} - \sqrt{\frac{\lambda_j(2m_j + \lambda_j)}{m_j + \lambda_j}}. \tag{8}$$

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

$$16m_j^4 + [16\lambda_j - 8(z_{\alpha/2}^2 + 1)]m_j^3 + [(z_{\alpha/2}^2 + 1)^2 - (20z_{\alpha/2}^2 + 12)\lambda_j + 4\lambda_j^2]m_j^2$$
$$+ [2(z_{\alpha/2}^2 + 1)^2\lambda_j - 16z_{\alpha/2}^2\lambda_j^2 - 4\lambda_j^3]m_j + (z_{\alpha/2}^2 + 1)^2\lambda_j^2 - 4z_{\alpha/2}^2\lambda_j^3 = 0. \tag{9}$$

The final Fisher-approximation-based (FAB) threshold t_j is obtained from m_j^*, the solution of (9). Owing to the following results, we do not need to write out the explicit expression of m_j^*, which could be rather complex:

Proposition 2. *The feasible condition for m_j is given by* (10)*, and the feasible solution m_j^* exists and is unique.*

$$m_j \geq \frac{1}{8}[z_{\alpha/2}^2 - 2\lambda_j + 1 + (z_{\alpha/2}^4 + (12\lambda_j + 2)z_{\alpha/2}^2 + 4\lambda_j^2 + 12\lambda_j + 1)^{1/2}]. \tag{10}$$

Proposition 2 implies that we can use any numerical quartic-equation solver, e.g. Hacke's method [18], to find the four solutions of (9). One and only one of the solutions will satisfy (10), which is m_j^*. The universal threshold can also be derived in the same way as in the CLTB case.

2.4. Summary of the denoising controlling FPR

Note that the thresholds $\tilde{t}_j$ 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 **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:    $\hat{\lambda}_j = 2^{jq}\lambda$ if λ is known; otherwise $\hat{\lambda}_j = \max(2^{cjq}a_j, 0)$
4:    Testing d_j^{bh} by applying thresholds $\tilde{t}_j$ for a prefixed FPR $= \alpha$
5:    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 FAB thresholds are applied to control a FPR $\alpha = 10^{-3}$. 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.
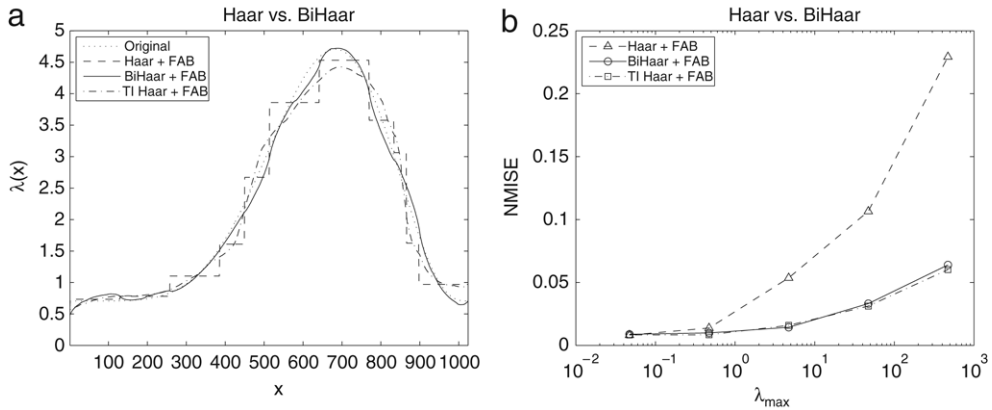
Fig. 1. Denoising the "Smooth" function (length = 1024). Estimates from Haar, Bi-Haar and TI Haar (undecimated) are compared. $\alpha = 10^{-3}$ and $J = 7$. (a) denoising results; (b) NMISEs.

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 jth spatial scale and the jth spectral scale, respectively. 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 2.8 GHz PC is 13 s for our Bi-Haar denoising, i.e., more than 50 times faster than the TI Haar denoising (665 s)), not to mention the memory space required by the redundant 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 at that band are depicted in Fig. 2(d). The denoising results using Haar and Bi-Haar transforms are respectively shown in Fig. 2(b) and (e), where FAB thresholds are 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 normalized ℓ^2-loss for Haar and Bi-Haar flux estimates, i.e. $\frac{1}{\sqrt{N}}\|\hat{S} - S\|_{\ell^2}$, are 14.4 and 7.7 respectively.

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 individual HTs can be applied to Bi-Haar coefficients to control a prefixed FPR. By doing so, we benefit from the regular Bi-Haar filter bank to gain a smooth estimate with
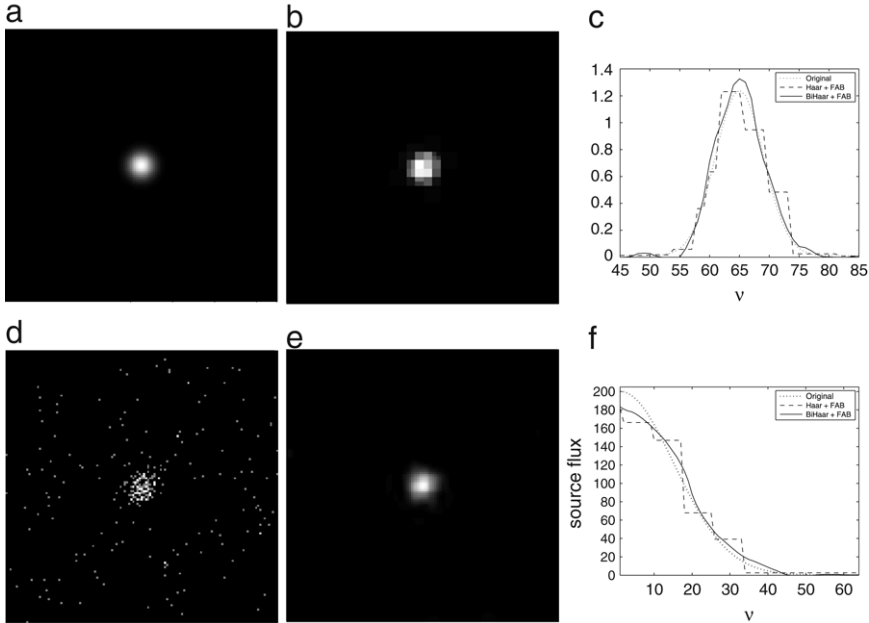
Fig. 2. Source-flux estimation in a hyperspectral image (size: $129 \times 129 \times 64$). $A_\nu \in [10^{-4}, 2]$; $J_{xy} = 3$, $J_\nu = 5$, FAB thresholding with $\alpha = 10^{-5}$. (a) Intensities at $\nu = 15$; (b) Haar-denoised data ($\nu = 15$); (c) estimated source profile at $\nu = 15$ (intensity along a line passing through the source center); (d) Poisson count image; (e) Bi-Haar-denoised data ($\nu = 15$); (f) estimated flux (respectively for Haar and Bi-Haar estimates: $\frac{1}{\sqrt{N}} \|\hat{S} - S\|_{\ell^2} = 14.4$ and 7.7.).

no "staircase" artifacts, while always maintaining a low computational complexity. A Fisher-approximation-based threshold implementing HTs is also designed. This approach could be extended in the future to fast deconvolution of Poisson data.

Appendix A. Proof of Proposition 1

Proof. We note that

$$p_H = \Pr(d_j^h \geq 2^{-cj} k_0 | H_0) = \sum_{k \geq k_0} e^{-\lambda_j} I_k(\lambda_j) = \sum_{k \leq -k_0} e^{-\lambda_j} I_{|k|}(\lambda_j)$$

where $k_0 \geq 1$. The p-value of d_{BH} is given by

$$p_{BH} = \Pr(X_1 - X_2 + 8(X_3 - X_4) \geq \lceil 8k_0/r \rceil | H_0)$$

$$= \sum_{k \in \mathbb{Z}} \Pr(X_3 - X_4 = k | H_0) \sum_{n = \lceil 8k_0/r \rceil}^{\infty} \Pr(X_1 - X_2 = n - 8k | H_0)$$

where $X_{1,2} \sim \mathcal{P}(\lambda_j)$, $X_{3,4} \sim \mathcal{P}(\lambda_j/2)$, and $(X_i)_i$ are independent. Now we have,

$$p_{BH} = \sum_{k \geq k_0} \Pr(X_3 - X_4 = k | H_0)$$

$$\times \sum_{n = \lceil 8k_0/r \rceil}^{\infty} \Pr(X_1 - X_2 = n - 8k | H_0) + \Pr(X_1 - X_2 = n + 8k | H_0)$$

$$+ \sum_{|k|<k_0} \Pr(X_3 - X_4 = k|H_0) \sum_{n=\lceil 8k_0/r \rceil}^{\infty} \Pr(X_1 - X_2 = n - 8k|H_0) \tag{A.1}$$

$$\leq p_H + \sum_{|k|<k_0} \mathrm{e}^{-\lambda_j} I_{|k|}(\lambda_j) \underbrace{\sum_{n=\lceil 8k_0/r \rceil}^{\infty} \mathrm{e}^{-2\lambda_j} I_{|n-8k|}(2\lambda_j)}_{T}.$$

To bound T, we use the identity [19] $\mathrm{e}^x = I_0(x) + 2\sum_{n=1}^{\infty} I_n(x)$. As $r < 1$, we have

$$T \leq \mathrm{e}^{-2\lambda_j} \sum_{n\geq 9} I_n(2\lambda_j) = \frac{1}{2}\left[1 - \mathrm{e}^{-2\lambda_j}\left(I_0(2\lambda_j) + 2\sum_{m=1}^{8} I_m(2\lambda_j)\right)\right] =: A(\lambda_j).$$

Thus, $p_{BH} \leq p_H + A(\lambda_j)(1 - 2p_H)$. As $\lambda \to 0+$, we have that $A(\lambda_j) = \frac{2^{9j-7}}{2835}\lambda^9 + o(\lambda^9)$. $\quad\square$

The same arguments can be carried out to bound p_{BH} in multi-dimensional cases. As an example, let us consider 2D data. A 2D wavelet transform will produce bands of horizontal, vertical and diagonal detail coefficients, i.e., $d_{j;H}$, $d_{j;V}$, and $d_{j;D}$. For horizontal and vertical coefficients, we have that $p_{BH} \leq p_H + A(\lambda_j)(1 - 2p_H)$, where $\lambda_j := 4^j\lambda$. For diagonal coefficients, it can be shown that $p_{BH} \leq p_H + B(\lambda_j)(1 - 2p_H)$, where

$$B(\lambda_j) := \frac{1}{2}\left[1 - \sum_{n=-64}^{64} \sum_{k=-\infty}^{\infty} \mathrm{e}^{-8\lambda_j} I_{|k|}(4\lambda_j) I_{|n-8k|}(4\lambda_j)\right]. \tag{A.2}$$

To see the behavior of $B(\lambda_j)$ as the intensity becomes small, we note $B(\lambda_j) = B_K(\lambda_j) + \epsilon_K$. Here, B_K is given by (A.2) with k ranging from $-K$ to K, and ϵ_K is the residual which can be made arbitrary small as K increases. Then, we have for all $K \geq 8$ that $B_K(\lambda_j) = \frac{8}{567}\lambda_j^9 + o(\lambda_j^9)$. Clearly, this procedure can be continued for higher-dimensional cases ($q > 2$).

Appendix B. Proof of Proposition 2

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 (10).
Next, when the equality in (10) 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 (10), and that $G(m_j) \to +\infty$ as $m_j \to +\infty$. $\quad\square$

References

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

[2] P. Besbeas, I. De Feis, T. Sapatinas, A comparative simulation study of wavelet shrinkage estimators for poisson counts, International Statistical Review 72 (2) (2004) 209–237.

[3] R. Willett, Multiscale analysis of photon-limited astronomical images, in: Statistical Challenges in Modern Astronomy (SCMA) IV, 2006.

[4] E.D. Kolaczyk, Nonparametric estimation of gamma-ray burst intensities using haar wavelets, The Astrophysical Journal 483 (1997) 340–349.

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

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

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

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

[9] Y. Bobichon, A. Bijaoui, A regularized image restoration algorithm for lossy compression in astronomy, Experimental Astronomy 7 (1997) 239–255.

[10] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, Journal of the Royal Statistical Society Series B 57 (1) (1995) 289–300.

[11] Y. Benjamini, D. Yekutieli, The control of the false discovery rate in multiple testing under dependency, The Annals of Statistics 29 (4) (2001) 1165–1188.

[12] J.G. Skellman, The frequency distribution of the difference between two Poisson variates belonging to different populations, Journal of the Royal Statistical Society Series A 109 (1946) 296.

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

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

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

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

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

[18] J.E. Hacke, Solving the quartic, The American Mathematical Monthly 48 (1941) 327–328.

[19] M. Abramowitz, I.A. Stegun, Handbook of Mathematical Functions, Dover, 1970.