

# Wavelets, Ridgelets and Curvelets for Poisson Noise Removal

Bo Zhang\*, Jalal M. Fadili and Jean-Luc Starck

## Abstract

In order to denoise Poisson count data, we introduce a variance stabilizing transform (VST) applied on a filtered discrete Poisson process, yielding a near Gaussian process with asymptotic constant variance. This new transform, which can be deemed as an extension of the Anscombe transform to filtered data, is simple, fast and efficient in (very) low-count situations. We combine this VST with the filter banks of wavelets, ridgelets and curvelets, leading to multiscale VSTs (MS-VSTs) and nonlinear decomposition schemes. By doing so, the noise-contaminated coefficients of these MS-VST-modified transforms are asymptotically normally distributed with known variances. A classical hypothesis-testing framework is adopted to detect the significant coefficients, and a sparsity-driven iterative scheme reconstructs properly the final estimate. A range of examples show the power of this MS-VST approach for recovering important structures of various morphologies in (very) low-count images. These results also demonstrate that the MS-VST approach is competitive relative to many existing denoising methods.

## Index Terms

Poisson intensity estimation, filtered Poisson process, multiscale variance stabilizing transform, wavelets, ridgelets, curvelets.

## EDICS Category: RST-DNOI Denoising

B. Zhang is with the Quantitative Image Analysis Group URA CNRS 2582 of Institut Pasteur, 75015 Paris, France. Phone: +33(0)140613974. Fax: +33(0)140613330. E-mail: bzhang@pasteur.fr

J. M. Fadili is with the GREYC CNRS UMR 6072, Image Processing Group, 14050 Caen Cedex, France. Phone: +33(0)231452920. Fax: +33(0)231452698. 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. Phone: +33(0)169085764. Fax: +33(0)169086577. E-mail: jstarck@cea.fr

## I. INTRODUCTION

Denoising images of Poisson counts arise from a variety of applications including astronomy and astrophysics [1], biomedical imaging [2], etc. Typically we observe a discrete dataset of counts  $\mathbf{X} = (X_i)_{i \in \mathbb{Z}^q}$  where  $X_i$  is a Poisson random variable of intensity  $\lambda_i$ , i.e.,  $X_i \sim \mathcal{P}(\lambda_i)$ . Here we suppose that  $X_i$ 's are mutually independent. The denoising aims at estimating the underlying intensity profile  $\Lambda = (\lambda_i)_{i \in \mathbb{Z}^q}$  from  $\mathbf{X}$ .

### *Literature overview*

A host of estimation methods have been proposed in the literature. Major contributions consist of: **1) variance stabilization:** A classical solution is to preprocess the data by applying a variance stabilizing transform (VST) such as the Anscombe transform [3][4]. It can be shown that the transformed data are approximately stationary and Gaussian. Once we are brought to the Gaussian denoising problem, standard approaches such as wavelet thresholding [5][6] are used before the VST is inverted to get the final estimate. Haar-Fisz transform is another widely used VST [7][8], which combines the Fisz transform [9] within the Haar transform. Jansen [10] introduced a conditional variance stabilization (CVS) approach which can be applied in any wavelet domain resulting in stabilized coefficients. **2) wavelet wiener filtering:** Nowak and Baraniuk [11], and Antoniadis and Sapatinas [12] proposed a wavelet domain filter, which can be interpreted as a data-adaptive wiener filter in a wavelet basis; **3) hypothesis testing:** Kolaczyk first introduced a Haar domain threshold [13], which implements a hypothesis testing procedure controlling a user-specified false positive rate (FPR). The hypothesis tests (HTs) have been extended to the biorthogonal Haar domain [14], leading to more regular reconstructions for smooth intensities. [15] derived the probability density function (*pdf*) of any wavelet coefficient, which allows HTs in an arbitrary wavelet basis. However, as the *pdf* has no closed forms, [15] is more computationally complex than Haar-based methods. [16] proposed “corrected” versions of the usual Gaussian-based thresholds for Poisson data. However, the asymptotic approximation adopted by [16] may not allow reasonable solutions in low-count situations. **4) empirical Bayesian and penalized ML estimations:** empirical Bayesian estimators are studied in [17][18][19][10]. The low-intensity case apart, Bayesian approaches generally outperform the direct wavelet filtering [11][12] (see also [20] for a comparative review). Poisson denoising has also been formulated as a penalized maximum likelihood (ML) estimation problem [21][22][23][24] within wavelet, wedgelet and platelet dictionaries. Wedgelet (platelet-) based methods are more efficient than wavelet-based estimators in denoising piecewise constant (smooth) images with smooth contours. To the

best of our knowledge, no Poisson denoising method has been proposed for the ridgelet and curvelet transforms.

### *This paper*

In this paper, we propose a VST to stabilize the variance of a filtered discrete Poisson process, yielding a near Gaussian process. This new transform, which can be deemed as an extension of the Anscombe transform to filtered data, is simple, fast and efficient in (very) low-count situations. The rationale behind the benefits of stabilizing a filtered version of the original process is as follows. It is well known that the performance of the Anscombe VST deteriorates as the intensity becomes low [1] (typically for  $\lambda < 10$ ), i.e., as the SNR decreases. Hence, one can alleviate this limitation and enhance the performance of the VST if the SNR is increased before stabilization. This can be achieved by pre-filtering the original process provided that the filter acts as an “averaging” kernel, or a low-pass filter. A detailed asymptotic analysis will support these claims.

By recognizing that a large family of multiscale transforms are computed from filtering equations (e.g. wavelets), the proposed VST can be seamlessly combined with their filter banks, leading to multiscale VSTs (MS-VSTs). Toward the goal of Poisson denoising, we are allowed to choose or design the most adaptive transform for the sources to be restored based on their morphology. Indeed, owing to recent advances in modern computational harmonic analysis, different multiscale transforms were shown to be very effective in sparsely representing different kinds of information. For example, to represent regular structures with point singularities, a qualified candidate is the wavelet transform [25][1]. The ridgelet transform [26] is very effective in representing global lines in an image. The curvelet system [27][28] is highly suitable for representing smooth ( $\mathcal{C}^2$ ) images away from  $\mathcal{C}^2$  contours. These transforms are also computationally attractive particularly in large-scale applications. We will show that our VST can be easily coupled with these different multiscale geometrical decompositions, yielding normally distributed coefficients with known variances. A classical hypothesis testing framework is then adopted to detect the significant coefficients, and a sparsity-driven iterative scheme is proposed to reconstruct the final estimate. We show that the MS-VST approach provides a very effective denoiser capable of recovering important structures of various (isotropic, line-like and curvilinear) shapes in (very) low-count images.

The paper is organized as follows. In Section II, a detailed analysis is provided to characterize the VST. Section III outlines the general denoising setting for using MS-VST with wavelets. Then, Section III-B and III-C show how the VST can be combined with the isotropic undecimated wavelet transform (IUWT) and the standard separable undecimated wavelet transform (UWT), respectively. Denoising by MS-VST

combined with ridgelets and curvelets are respectively presented in Section IV and V. Section VI provides a discussion on the numerical results obtained, followed by a brief conclusion and the perspectives of our work. Mathematical proofs are deferred to the appendix.

## II. VST OF A FILTERED POISSON PROCESS

Given a Poisson process  $\mathbf{X} := (X_i)_i$  where  $X_i$ 's are independent and  $X_i \sim \mathcal{P}(\lambda_i)$ ,  $Y_j := \sum_i h[i]X_{j-i}$  is the filtered process obtained by convolving  $\mathbf{X}$  with a discrete filter  $h$ . We will use  $Y$  to denote any one of the  $Y_j$ 's. Let us define  $\tau_k := \sum_i (h[i])^k$  for  $k = 1, 2, \dots$ . In addition, we adopt a local homogeneity assumption that  $\lambda_{j-i} = \lambda$  for all  $i$  within the support of  $h$ .

### A. VST-heuristics

It can be seen that the variance of  $Y$  ( $\text{Var}[Y]$ ) is proportional to the intensity  $\lambda$ . To stabilize  $\text{Var}[Y]$ , we seek a transformation  $Z := T(Y)$  such that  $\text{Var}[Z]$  is (asymptotically) constant, say 1, irrespective of the value of  $\lambda$ .

Heuristically, the Taylor expansion gives us  $T(Y) \approx T(\mu_Y) + T'(\mu_Y)(Y - \mu_Y)$ , where  $\mu_Y := \mathbb{E}[Y] = \lambda\tau_1$ . We then have  $\text{Var}[Z] \approx T'(\mu_Y)^2 \cdot \text{Var}[Y] = T'(\mu_Y)^2 \cdot \lambda\tau_2$ . Hence, by setting  $\text{Var}[Z] = 1$ , we obtain a differential equation  $T'(\mu_Y) = \mu_Y^{-1/2} \sqrt{\tau_1/\tau_2}$ , of which the solution is given by  $T(Y) = 2\sqrt{\tau_1/\tau_2}\sqrt{Y}$ . This implies that the square-root transform could serve as a VST. It is possible to use higher order Taylor expansions to find VST of different forms, but solving the associated differential equations is found difficult since they are highly non-linear.

### B. VST-rigor

We define the square-root transform  $T$  as follows:

$$T(Y) := b \cdot \text{sgn}(Y + c)|Y + c|^{1/2} \quad (1)$$

where  $b$  is a normalizing factor. Lemma 1 confirms our heuristics that  $T$  is indeed a VST for a filtered Poisson process (with a nonzero-mean filter) in that  $T(Y)$  is asymptotically normally distributed with a stabilized variance as  $\lambda$  becomes large.

**Lemma 1 (Square root as VST)** *If  $\tau_1 \neq 0$ ,  $\|h\|_2, \|h\|_3 < \infty$ , then we have:*

$$\text{sgn}(Y + c)\sqrt{|Y + c|} - \text{sgn}(\tau_1)\sqrt{|\tau_1|\lambda} \xrightarrow[\lambda \rightarrow +\infty]{\mathcal{D}} \mathcal{N}\left(0, \frac{\tau_2}{4|\tau_1|}\right) \quad (2)$$

where  $\text{sgn}(\cdot)$  is the sign function.

This result holds true for any  $c \in \mathbb{R}$ , of which the value controls the convergence rate in (2). The next section provides an analysis of the asymptotic rate and determines the optimal value of  $c$ .

### C. Optimal parameter of the VST

To simplify the asymptotic analysis, we assume a non-negative filter  $h$  and a positive constant  $c$  (a non-positive  $h$  with a negative  $c$  can also be considered). Thus, our VST is simplified to  $Z := T(Y) = b\sqrt{Y + c}$ . We can now derive the asymptotic expansions of  $\mathbb{E}[Z]$  and  $\text{Var}[Z]$  as stated in Proposition 1. Note that the last point in the proposition results directly from Lemma 1.

#### Proposition 1 (Optimal parameter of the VST)

(i) Define  $Z := b\sqrt{Y + c}$ . Then we have:

$$\mathbb{E}[Z] = b\sqrt{\lambda\tau_1} + b\frac{4c\tau_1 - \tau_2}{8\tau_1^{3/2}}\lambda^{-1/2} + O_{\lambda \rightarrow +\infty}(\lambda^{-1}) \quad (3)$$

$$\begin{aligned} \text{Var}[Z] = & b^2\frac{\tau_2}{4\tau_1} + b^2\left(\frac{7\tau_2^2}{32\tau_1^3} - \frac{2\tau_2c + \tau_3}{8\tau_1^2}\right)\lambda^{-1} + b^2\left(\frac{5\tau_4 + 16c^2\tau_2 + 16c\tau_3}{64\tau_1^3}\right. \\ & \left. - \frac{17\tau_2\tau_3 + 21c\tau_2^2}{32\tau_1^4} + \frac{75\tau_2^3}{128\tau_1^5}\right)\lambda^{-2} + O_{\lambda \rightarrow +\infty}(\lambda^{-5/2}) \end{aligned} \quad (4)$$

(ii) For the VST to be second order accurate and  $Z$  to have asymptotic unit variance,  $b$  and  $c$  must satisfy:

$$c = \frac{7\tau_2}{8\tau_1} - \frac{\tau_3}{2\tau_2}, \quad b = b_1 := 2\sqrt{\frac{\tau_1}{\tau_2}} \quad (5)$$

(iii) For  $b$  and  $c$  as above,  $Z - b_1\sqrt{\tau_1\lambda} \xrightarrow[\lambda \rightarrow +\infty]{\mathcal{D}} \mathcal{N}(0, 1)$ .

Proposition 1 tells us that for the chosen value of  $c$ , the first order term in the expansion (4) disappears, and the variance is almost constant up to a second order residue. As a normalizing factor, the value of  $b$  does not influence the convergence rate. Noted that if there is no filtering ( $h = \delta$ ),  $c$  given by (5) equals  $3/8$ , i.e., the value of the Anscombe VST.

Now fix  $c$  to the value given in (5). Once the asymptotic expectation is normalized to  $\sqrt{\lambda}$ , the coefficient of the higher-order term  $O(\lambda^{-1/2})$  in (3) is given by (6). Similarly, the asymptotic variance being normalized to 1, the coefficient of the term  $O(\lambda^{-2})$  in (4) is shown in (7).

$$C_{\mathbb{E}} = \frac{5\tau_2^2 - 4\tau_1\tau_3}{16\tau_1^2\tau_2} \quad (6)$$

$$C_{\text{Var}} = \frac{5\tau_1^2\tau_2\tau_4 + 13\tau_2^4 - 4\tau_1^2\tau_3^2 - 13\tau_1\tau_2^2\tau_3}{16\tau_1^4\tau_2^2} \quad (7)$$

TABLE I  
 $C_{\mathbb{E}}$  AND  $C_{\text{Var}}$  OF DIFFERENT FILTERS

Filter <sup>1</sup> $h$	$C_{\mathbb{E}}$	$C_{\text{Var}}$
$\delta$ (Anscombe)	$6.25 \times 10^{-2}$	$6.25 \times 10^{-2}$
2D Average = $h_A \otimes h_A$	$6.94 \times 10^{-3}$	$7.72 \times 10^{-4}$
2D $B_3$ -Spline = $h_{B_3} \otimes h_{B_3}$	$-4.94 \times 10^{-4}$	$-3.45 \times 10^{-4}$

<sup>1</sup>  $h_A = [1 \ 1 \ 1]/3$ ;  $h_{B_3} = [1 \ 4 \ 6 \ 4 \ 1]/16$ ;  $\otimes$  denotes the tensor product.

These higher-order coefficients (6) and (7) can be used to evaluate the stabilization efficiency for a given filter. The ideal filters will be those minimizing (6) and (7). Tab. I shows the values of  $C_{\mathbb{E}}$  and  $C_{\text{Var}}$  for different filters, where  $h = \delta$  corresponds to the Anscombe VST (no filtering). Note that the values for the Anscombe VST are 10 or even 100 times larger than for the other cases, indicating the benefits of filtering prior to the stabilization. This is also confirmed by the simulations depicted in Fig. 1, where the estimates of  $\mathbb{E}[Z]$  (resp.  $\text{Var}[Z]$ ) obtained from  $2 \cdot 10^5$  replications are plotted as a function of the intensity  $\lambda$  for Anscombe [3] (dashed-dotted), Haar-Fisz [7] (dashed) and our VST (solid). The asymptotic bounds (dotted), i.e.,  $\sqrt{\lambda}$  for the expectation and 1 for the variance, are also shown. It can be seen that for increasing intensity,  $\mathbb{E}[Z]$  and  $\text{Var}[Z]$  stick to the theoretical bounds at different rates depending on the VST used. Quantitatively, Poisson variables transformed using the Anscombe VST can be reasonably considered to be unbiased and stabilized for  $\lambda \gtrsim 10$ , using Haar-Fisz for  $\lambda \gtrsim 1$ , and using our VST (after low-pass filtering with the chosen  $h$ ) for  $\lambda \gtrsim 0.1$ .

### III. DENOISING BY MS-VST+WAVELETS

#### A. General setting

In this section, the proposed VST will be incorporated within the multiscale framework offered by the (non-necessarily separable) UWT [25][29][30], giving rise to the MS-VST. The undecimated transform is used since it provides translation-invariant denoising. Below, we first discuss the one-dimensional (1D) denoising case, and the multidimensional extension will be straightforward (Section III-B2 and III-C2).

The UWT uses an analysis filter bank  $(h, g)$  to decompose a signal  $a_0$  into a coefficient set  $W = \{d_1, \dots, d_J, a_J\}$ , where  $d_j$  is the wavelet (detail) coefficients at scale  $j$  and  $a_J$  is the approximation coefficients at the coarsest resolution  $J$ . The passage from one resolution to the next one is obtained

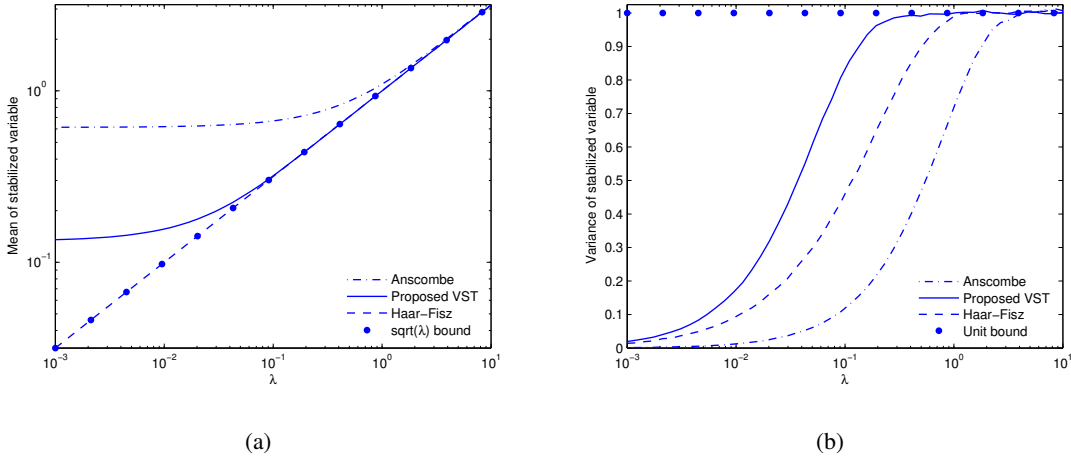


Fig. 1. Behavior of (a)  $\mathbb{E}[Z]$  and (b)  $\text{Var}[Z]$  as a function of the underlying intensity, for the Anscombe VST, 2D Haar-Fisz VST, and the proposed VST with a low-pass filter  $h = 2\text{D } B_3\text{-Spline}$  filter.

using the “à trous” algorithm [31][32]:

$$a_{j+1}[l] = (\bar{h}^{\uparrow j} \star a_j)[l] = \sum_k h[k] a_j[l + 2^j k], \quad d_{j+1}[l] = (\bar{g}^{\uparrow j} \star a_j)[l] = \sum_k g[k] a_j[l + 2^j k] \quad (8)$$

where  $h^{\uparrow j}[l] = h[l]$  if  $l/2^j \in \mathbb{Z}$  and 0 otherwise,  $\bar{h}[n] = h[-n]$ , and “ $\star$ ” denotes convolution. The reconstruction is given by:  $a_j[l] = \frac{1}{2} \left[ (\tilde{h}^{\uparrow j} \star a_{j+1})[l] + (\tilde{g}^{\uparrow j} \star d_{j+1})[l] \right]$ . The filter bank  $(h, g, \tilde{h}, \tilde{g})$  needs to satisfy the exact reconstruction condition.

Now the VST can be combined with the UWT in the following way: since  $(\bar{h}^{\uparrow j})_j$  are low-pass filters (so have nonzero means), we can first stabilize the approximation coefficients  $(a_j)_j$  using the VST, and then compute in the standard way the detail coefficients from the stabilized  $a_j$ ’s. Note that the VST is now scale-dependent (hence MS-VST). By doing so, the asymptotic stabilized Gaussianity of the  $a_j$ ’s will be transferred to the  $d_j$ ’s, as will be shown later. Thus, the distribution of the  $d_j$ ’s being known (Gaussian), we can detect the significant coefficients by classical HTs. With the knowledge of the detected coefficients, the final estimate can be reconstructed. In summary, UWT denoising with the MS-VST involves the following three main steps:

- 1) **Transformation (Sections III-B and III-C):** Compute the UWT in conjunction with the MS-VST;
- 2) **Detection (Section III-D):** Detect significant detail coefficients by HTs;
- 3) **Estimation (Section III-E):** Reconstruct the final estimate iteratively using the knowledge of the detected coefficients.

The last step needs some explanation. The signal reconstruction requires inverting the MS-VST-combined

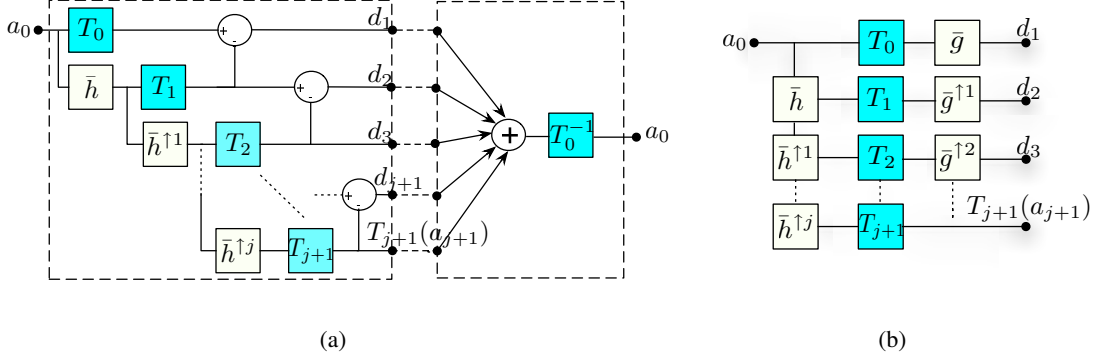


Fig. 2. Diagrams of the MS-VST in 1D. (a) MS-VST combined with the IUWT. The left dashed frame shows the decomposition part and the right one illustrates the direct inversion; (b) MS-VST combined with the standard UWT. The decomposition part is shown and no direct inversion exists.

UWT after the detection step. However, the nonlinearity of the MS-VST makes a direct inversion impossible in the general case. Even for the IUWT, which uses special filter banks yielding an invertible MS-VST, the direct inverse will be seen to be suboptimal. Hence, we propose to reformulate the reconstruction as a convex sparsity-promoting optimization problem and solve it by an iterative steepest descent algorithm (Section III-E).

### B. MS-VST+IUWT

The IUWT [33] uses the filter bank  $(h, g = \delta - h, \tilde{h} = \delta, \tilde{g} = \delta)$  where  $h$  is typically a symmetric low-pass filter such as the  $B_3$ -Spline filter. The particular structure of the analysis filters  $(h, g)$  leads to the iterative decomposition scheme shown in the left part of (9). The reconstruction is trivial, i.e.,  $a_0 = a_J + \sum_{j=1}^J d_j$ . This algorithm is widely used in astronomical applications [1] and biomedical imaging [34] to detect isotropic objects.

As stated in Section III-A, we apply the VST on the  $a_j$ 's resulting in the stabilization procedure shown in the right part of (9):

$$\text{IUWT} \begin{cases} a_j = \bar{h}^{\uparrow j-1} \star a_{j-1} \\ d_j = a_{j-1} - a_j \end{cases} \Rightarrow \begin{matrix} \text{MS-VST} \\ + \\ \text{IUWT} \end{matrix} \begin{cases} a_j = \bar{h}^{\uparrow j-1} \star a_{j-1} \\ d_j = T_{j-1}(a_{j-1}) - T_j(a_j) \end{cases} \quad (9)$$

Note that the filtering step on  $a_{j-1}$  can be rewritten as a filtering on  $a_0 := \mathbf{X}$ , i.e.,  $a_j = h^{(j)} \star a_0$ , where  $h^{(j)} = \bar{h}^{\uparrow j-1} \star \dots \star \bar{h}^{\uparrow 1} \star \bar{h}$  for  $j \geq 1$  and  $h^{(0)} = \delta$ .  $T_j$  is the VST operator at scale  $j$  (see Lemma 1):

$$T_j(a_j) = b^{(j)} \text{sgn}(a_j + c^{(j)}) \sqrt{|a_j + c^{(j)}|} \quad (10)$$



Let us define  $\tau_k^{(j)} := \sum_i \left( h^{(j)}[i] \right)^k$ . Then according to (5), the constant  $c^{(j)}$  associated to  $h^{(j)}$  should be set to

$$c^{(j)} := \frac{7\tau_2^{(j)}}{8\tau_1^{(j)}} - \frac{\tau_3^{(j)}}{2\tau_2^{(j)}} \quad (11)$$

This stabilization procedure is directly invertible as we have:

$$a_0 = T_0^{-1} \left[ T_J(a_J) + \sum_{j=1}^J d_j \right] \quad (12)$$

The decomposition scheme and the inversion of MSVST+IUWT are also illustrated in Fig. 2(a).

1) *Asymptotic distribution of the detail coefficients:*

**Theorem 1 (Asymptotic distribution of  $d_j$ )** *Setting  $b^{(j)} := \text{sgn}(\tau_1^{(j)})/\sqrt{|\tau_1^{(j)}|}$ , if  $\lambda$  is constant within the support of the filter  $h^{(j)}[k - \cdot]$ , then we have:*

$$d_j[k] \xrightarrow{\lambda \rightarrow +\infty} \mathcal{N} \left( 0, \frac{\tau_2^{(j-1)}}{4\tau_1^{(j-1)^2}} + \frac{\tau_2^{(j)}}{4\tau_1^{(j)^2}} - \frac{\langle h^{(j-1)}, h^{(j)} \rangle}{2\tau_1^{(j-1)}\tau_1^{(j)}} \right) \quad (13)$$

Here  $\langle \cdot, \cdot \rangle$  represents the scalar product. This is a very useful result showing that the detail coefficients issued from locally homogeneous parts of the signal (null hypothesis  $H_0$ , see Section III-D) follow asymptotically a centered normal distribution with an intensity-independent variance. The variance only depends on the filter  $h$  and the current scale. Hence, the stabilized variance (also the constants  $b^{(j)}$ ,  $c^{(j)}$ ,  $\tau_k^{(j)}$ ) can all be pre-computed once for any given  $h$ .

2) *Extension to the multi-dimensional case:* The filter bank in  $q$ D ( $q > 1$ ) becomes ( $h_{qD}, g_{qD} = \delta - h_{qD}, \tilde{h}_{qD} = \delta, \tilde{g}_{qD} = \delta$ ) where  $h_{qD} = \otimes_{i=1}^q h$ . Note that  $g_{qD}$  is in general nonseparable. The MS-VST decomposition scheme remains the same as (9), and the asymptotic result above holds true. The complexity for pre-computing  $b^{(j)}$ ,  $c^{(j)}$ ,  $\tau_k^{(j)}$  and the stabilized variance in (13) remains the same as in the 1D case.

### C. MS-VST+Standard UWT

In this section, we show how the MS-VST can be used to stabilize the wavelet coefficients of a standard separable UWT. In the same vein as (9), we apply the VST on the approximation coefficients  $(a_j)_j$ , leading to the following scheme (see also the block-diagram of Fig. 2(b)):

$$\text{UWT} \begin{cases} a_j &= \bar{h}^{\uparrow j-1} \star a_{j-1} \\ d_j &= \bar{g}^{\uparrow j-1} \star a_{j-1} \end{cases} \Rightarrow \begin{matrix} \text{MS-VST} \\ + \\ \text{UWT} \end{matrix} \begin{cases} a_j &= \bar{h}^{\uparrow j-1} \star a_{j-1} \\ d_j &= \bar{g}^{\uparrow j-1} \star T_{j-1}(a_{j-1}) \end{cases} \quad (14)$$

where  $T_j(a_j) = b^{(j)} \text{sgn}(a_j + c^{(j)}) \sqrt{|a_j + c^{(j)}|}$ , and  $c^{(j)}$  is defined as in (11).

1) *Asymptotic distribution of the detail coefficients:*

**Theorem 2 (Asymptotic distribution of  $d_j$ )** Setting  $b^{(j)} := 2\sqrt{|\tau_1^{(j)}|/\tau_2^{(j)}}$ , if  $\lambda$  is constant within the support of the filter  $(\bar{g}^{\uparrow j-1} \star h^{(j-1)})[k - \cdot]$ , then  $d_j[k] \xrightarrow[\lambda \rightarrow +\infty]{\mathcal{D}} \mathcal{N}(0, \sigma_j^2)$ , where

$$\sigma_j^2 = \frac{1}{\tau_2^{(j-1)}} \sum_{m,n} \bar{g}^{\uparrow j-1}[m] \bar{g}^{\uparrow j-1}[n] \sum_k h^{(j-1)}[k] h^{(j-1)}[k + m - n] \quad (15)$$

Parallel to Theorem 1, Theorem 2 shows the asymptotic normality of the wavelet detail coefficients obtained from locally homogeneous parts of the signal (null hypothesis  $H_0$ , see Section III-D). Here, the values of  $b^{(j)}$ ,  $c^{(j)}$ ,  $\tau_k^{(j)}$  and  $\sigma_j$  can all be pre-computed once the wavelet has been chosen.

2) *Extension to the multi-dimensional case:* The scheme (14) can be extended straightforwardly to higher dimensional cases, and the asymptotic result above holds true. For example, in the 2D case, the UWT is given by the left part of (16) and the version combined with the MS-VST is given on the right:

$$\text{UWT} \begin{cases} a_j &= \bar{h}^{\uparrow j-1} \bar{h}^{\uparrow j-1} \star a_{j-1} \\ d_j^1 &= \bar{g}^{\uparrow j-1} \bar{h}^{\uparrow j-1} \star a_{j-1} \\ d_j^2 &= \bar{h}^{\uparrow j-1} \bar{g}^{\uparrow j-1} \star a_{j-1} \\ d_j^3 &= \bar{g}^{\uparrow j-1} \bar{g}^{\uparrow j-1} \star a_{j-1} \end{cases} \Rightarrow \begin{matrix} \text{MS-VST} \\ + \\ \text{UWT} \end{matrix} \begin{cases} a_j &= \bar{h}^{\uparrow j-1} \bar{h}^{\uparrow j-1} \star a_{j-1} \\ d_j^1 &= \bar{g}^{\uparrow j-1} \bar{h}^{\uparrow j-1} \star T_{j-1}(a_{j-1}) \\ d_j^2 &= \bar{h}^{\uparrow j-1} \bar{g}^{\uparrow j-1} \star T_{j-1}(a_{j-1}) \\ d_j^3 &= \bar{g}^{\uparrow j-1} \bar{g}^{\uparrow j-1} \star T_{j-1}(a_{j-1}) \end{cases} \quad (16)$$

where  $hg \star a$  is the convolution of  $a$  by the separable filter  $hg$ , i.e., convolution first along the rows by  $h$  and then along the columns by  $g$ . The complexity of pre-computing the constants  $b^{(j)}$ ,  $c^{(j)}$ ,  $\tau_k^{(j)}$  and  $\sigma_j$  remains the same as in the 1D case.

#### D. Detection by wavelet-domain hypothesis testing

Our wavelet-domain detection is formulated by HTs [35], i.e.,  $H_0 : d_j[k] = 0$  vs.  $H_1 : d_j[k] \neq 0$ . The coefficients of the null hypothesis are insignificant ones. Note that wavelet coefficients computed from locally homogeneous parts of the signal are insignificant. Indeed, if there were no noise, these coefficients obtained by applying the classical UWT scheme would be zero-valued, since any wavelet has a zero mean. Thanks to Theorems 1 and 2, the distribution of  $d_j[k]$  under the null hypothesis  $H_0$  is now known (Gaussian).

HTs can be carried out individually in a coefficient-by-coefficient manner. First, the user pre-specifies a FPR in the wavelet domain, say  $\alpha$ . Then the  $p$ -value of each coefficient  $p := 2[1 - \Phi(|d|/\sigma)]$  is calculated under  $H_0$ . Here  $\Phi(x)$  is the standard normal cumulative distribution function, and  $\sigma$  is the asymptotic standard deviation of  $d$  after being stabilized by the MS-VST. Finally, all the coefficients with  $p > \alpha$  will be zeroed.

If we desire a control over global statistical error rates, multiple HTs should be used. For example, the Bonferroni over-conservative correction controls the probability of erroneously rejecting even one of the true null hypothesis, i.e., the Family-Wise Error Rate (FWER). Alternatively, one can carry out the Benjamini and Hochberg procedure [36] 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 [37]. The latter point allows the FDR control in non-orthogonal wavelet domains. Minimality of FDR has also been studied in various settings (see [38][39] for details).

### E. Iterative reconstruction

Following the detection step, we have to invert the MS-VST scheme to reconstruct the estimate. For the standard UWT case, direct reconstruction procedure is unavailable since the convolution (by  $\bar{g}^{\uparrow j-1}$ ) operator and the nonlinear VST operator  $T_{j-1}$  do not commute in (14). For the IUWT case, although the direct MS-VST inversion is possible (by (12)), it can not guarantee a positive reconstruction (the Poisson intensity is always nonnegative). A positivity projection may be applied on the direct inverse, but could entail a loss of important structures in the estimate (see results in Section III-F). Thus, we propose to reformulate the reconstruction as a convex optimization problem described below, and solve it iteratively. In the following, we will concentrate on the 1D case for clarity.

We suppose that the underlying intensity function  $\Lambda$  is sparsely represented in the wavelet domain. We define the multiresolution support [40]  $\mathcal{M}$ , which is determined by the set of detected significant coefficients at each scale  $j$  and location  $k$ , i.e.,

$$\mathcal{M} := \{(j, k) \mid \text{if } d_j[k] \text{ is significant}\} \quad (17)$$

The estimation is then formulated as a constrained sparsity-promoting minimization problem in terms of the wavelet coefficients  $\mathbf{d}$ . A component of  $\mathbf{d}$  can be indexed by the usual scale-location index  $(j, k)$  (i.e.  $d_j[k]$ ). The indices can also be renumbered so that  $\mathbf{d}$  is mapped to a vector in  $\mathbb{R}^L$ . In this case, a component of  $\mathbf{d}$  is indexed in a 1D way, i.e.,  $d[i]$ . Hereafter, both notations will be used. Our optimization problem is given by

$$\begin{aligned} \min_{\mathbf{d} \in \mathcal{C}} J(\mathbf{d}), \quad J(\mathbf{d}) &:= \|\mathbf{d}\|_1 \\ \text{where } \mathcal{C} &:= \mathcal{S}_1 \cap \mathcal{S}_2, \quad \mathcal{S}_1 := \{\mathbf{d} \mid d_j[k] = (\mathcal{W}\mathbf{X})_j[k], (j, k) \in \mathcal{M}\}, \quad \mathcal{S}_2 := \{\mathbf{d} \mid \mathcal{R}\mathbf{d} \geq 0\} \end{aligned} \quad (18)$$

where  $\mathcal{W}$  represents the wavelet transform operator, and  $\mathcal{R}$  its (weak-generalized) left inverse (synthesis operator). Recall that  $\mathbf{X}$  is the observed count data vector. Clearly, we seek the sparsest solution by

minimizing the  $\ell^1$ -objective [41][42] within the feasible set  $\mathcal{C} := \mathcal{S}_1 \cap \mathcal{S}_2$ . The set  $\mathcal{S}_1$  requires that the significant elements of  $\mathbf{d}$  preserve those of the data  $\mathbf{X}$ ; the set  $\mathcal{S}_2$  ensures a positive intensity estimate.

(18) is a convex optimization problem which can be cast as a Linear Program (LP) and solved using interior-point methods. However, the computational complexity of the LP solver increases dramatically with the size of the problem. Classical projected (sub-)gradient method is also difficult to apply here since the projector on the feasible set is unknown. Below we propose an alternative based on the hybrid steepest descent (HSD) [43]. The HSD approach allows minimizing convex functionals over the intersection of fixed point sets of nonexpansive mappings. It is much faster than LP, and in our problem, the nonexpansive mappings do have closed forms.

**Theorem 3** *Let  $\mathbf{d} \in \mathbb{R}^L$ . Define the following regularized optimization problem ( $\epsilon \geq 0$ ):*

$$\min_{\mathbf{d} \in \mathcal{C}_B} J_\epsilon(\mathbf{d}), \quad J_\epsilon(\mathbf{d}) := \sum_{i=1}^L \sqrt{d[i]^2 + \epsilon} \quad (19)$$

where  $\mathcal{C}_B := \mathcal{S}_1 \cap \mathcal{S}_2 \cap \mathcal{S}_3$ ,  $\mathcal{S}_3 := \{\mathbf{d} \mid \|\mathbf{d}\|_2 \leq B, B \geq \|\mathcal{W}\mathbf{x}\|_1\}$

*Define the HSD iteration scheme [43] ( $k \geq 0$ ):*

$$\mathbf{d}_\epsilon^{(k+1)} := T_{\mathcal{C}_B} \mathbf{d}_\epsilon^{(k)} - \beta_{k+1} \nabla J_\epsilon(T_{\mathcal{C}_B} \mathbf{d}_\epsilon^{(k)}) \quad (20)$$

where  $\nabla J_\epsilon$  is the gradient of  $J_\epsilon$ , and  $T_{\mathcal{C}_B} := P_{\mathcal{S}_3} \circ P_{\mathcal{S}_1} \circ Q_{\mathcal{S}_2}$ ,

$$P_{\mathcal{S}_3} \mathbf{d} := \frac{\mathbf{d}}{\|\mathbf{d}\|_2} \cdot \min(\|\mathbf{d}\|_2, B); \quad (P_{\mathcal{S}_1} \mathbf{d})_j[k] := \begin{cases} (\mathcal{W}\mathbf{X})_j[k] & (j, k) \text{ in } \mathcal{M} \\ d_j[k] & \text{otherwise} \end{cases}; \quad Q_{\mathcal{S}_2} \mathbf{d} := \mathcal{W}P_+ \mathcal{R} \mathbf{d} \quad (21)$$

where  $P_+$  represents the projection onto the nonnegative orthant, and  $P_{\mathcal{S}_1}$  and  $P_{\mathcal{S}_3}$  are the projectors onto their respective constraint sets. The step sequence satisfies:

$$\lim_{k \rightarrow \infty} \beta_k = 0, \quad \sum_{k \geq 1} \beta_k = +\infty \quad \text{and} \quad \sum_{k \geq 1} |\beta_k - \beta_{k+1}| < +\infty \quad (22)$$

Suppose that in (ii)-(v) below  $\mathcal{W}$  represents a tight frame decomposition and  $\mathcal{R}$  its pseudo-inverse operator. Then we have:

- (i) The solution set of (18) is the same as that of (19) with  $\epsilon = 0$ ;
- (ii)  $T_{\mathcal{C}_B}$  is nonexpansive, and its fix point set is  $\text{Fix}(T_{\mathcal{C}_B}) = \mathcal{C}_B \neq \emptyset$ ;
- (iii)  $\forall \epsilon > 0$ , with any  $\mathbf{d}_\epsilon^{(0)} \in \mathbb{R}^N$ ,  $\mathbf{d}_\epsilon^{(k)} \xrightarrow[k \rightarrow +\infty]{} \mathbf{d}_\epsilon^*$ , where  $\mathbf{d}_\epsilon^*$  is the unique solution to (19);
- (iv) As  $\epsilon \rightarrow 0^+$ , the sequence  $(\mathbf{d}_\epsilon^*)_{\epsilon > 0}$  is bounded. Therefore, it has at least one limit point;
- (v) As  $\epsilon \rightarrow 0^+$ , every limit point of the sequence  $(\mathbf{d}_\epsilon^*)_{\epsilon > 0}$  is a solution to (18).

Theorem 3 implies that in practice instead of directly solving (18), one can solve its smoothed version (19) by applying (20) with a small  $\epsilon$ . In real problems,  $T_{\mathcal{C}_B}$  is simplified to  $T_{\mathcal{C}_B} = T_{\mathcal{C}} := P_{\mathcal{S}_1} \circ Q_{\mathcal{S}_2}$ .

Indeed, as we always deal with finite dimensional data with bounded values, the exact value of  $B$  is not important, and it can be considered to be sufficiently large so that the constraint  $\mathcal{S}_3$  is always satisfied. We also point out that although Theorem 3 assumes a tight frame decomposition and pseudo-inverse reconstruction, in our experiments, we observed that the iterations (20) applied equally to general frame decompositions and inverses, and performed very well even with  $\epsilon = 0$  (see results in Section III-F). For  $\epsilon = 0$ , (20) rewrites:

$$\mathbf{d}^{(k+1)} := T_C \mathbf{d}^{(k)} - \beta_{k+1} \nabla J \left( T_C \mathbf{d}^{(k)} \right) \quad (23)$$

where the sub-gradient  $\nabla J(\mathbf{d})[i] = \text{sgn}(\mathbf{d}[i])$ . (23) is implemented in practice as a soft thresholding with a threshold  $\beta_{k+1}$  (noted as  $\text{ST}_{\beta_{k+1}}$ ). Now the MS-VST denoising using the IUWT and the standard UWT is presented in Algorithm 1 and 2 respectively. In Algorithm 1, step 1 – 6 obtain a first estimate of

---

**Algorithm 1** MS-VST + Isotropic UWT

---

**Require:**  $a_0 := \mathbf{X}$ ; a low-pass filter  $h$ ,

**Detection**

- 1: **for**  $j = 1$  to  $J$  **do**
- 2:   Compute  $a_j$  and  $d_j$  using (9).
- 3:   HTs on  $d_j$  assuming the normal statistics (Theorem 1), get the estimate  $\hat{d}_j$ , and update  $\mathcal{M}$ .
- 4: **end for**

**Estimation**

- 5: Estimate  $\mathbb{E}[T_0(a_0)]$  by:  $\widehat{T_0 a_0} = \sum_{j=1}^J \hat{d}_j + T_J(a_J)$
  - 6: Estimate  $\mathbb{E}[a_0]$  by:  $\hat{a}_0 = \text{Var}[T_0(a_0)] + \widehat{T_0 a_0}^2 - c^{(0)}$
  - 7: Initialize  $\mathbf{d}^{(0)} = \mathcal{W}P_+ \hat{a}_0$
  - 8: **for**  $k = 1$  to  $N_{\max}$  **do**
  - 9:    $\tilde{\mathbf{d}} := P_{S_1} \circ Q_{S_2} \mathbf{d}^{(k-1)}$
  - 10:    $\hat{\mathbf{d}} := \mathbf{d}^{(k)} := \text{ST}_{\beta_k}[\tilde{\mathbf{d}}]$ .
  - 11: **end for**
  - 12: Get the estimate  $\hat{\Lambda} = P_+ \mathcal{R} \hat{\mathbf{d}}$ .
- 

$\Lambda$  by directly inverting MS-VST+IUWT after zeroing the insignificant wavelet coefficients. The direct inverse serves as the initialization of the iterations. In step 6, the term  $\text{Var}[T_0(a_0)]$  corrects the bias due to squaring an estimate. Indeed, if  $Z = \sqrt{a_0 + c^{(0)}}$ , then  $\lambda = \mathbb{E}[a_0] = \mathbb{E}[Z^2] - c^{(0)} = \text{Var}[Z] + \mathbb{E}[Z]^2 - c^{(0)}$ .

In Algorithm 2 the initialization is provided by the detected significant wavelet coefficients (step 5). For both algorithms,  $N_{\max}$  is the maximum number of iterations. A possible choice of the step sequence  $(\beta_k)_k$  is a linearly decreasing one:  $\beta_k = \frac{N_{\max} - k}{N_{\max} - 1}$ ,  $k = 1, 2, \dots, N_{\max}$ . It can be noted that for  $(\beta_k)_k$  chosen as above, the conditions in (22) are all satisfied as  $N_{\max} \rightarrow \infty$ . The computational cost of the whole denoising is dominated by the iterative estimation step. This step involves an analysis and a

---

**Algorithm 2** MS-VST + Standard UWT

---

**Require:**  $a_0 := \mathbf{X}$ ; a wavelet filter bank  $(h, g, \tilde{h}, \tilde{g})$ ,

**Detection**

- 1: **for**  $j = 1$  to  $J$  **do**
- 2:   Compute  $a_j$  and  $d_j$  using (14).
- 3:   HTs on  $d_j$  assuming the normal statistics (Theorem 2) and update  $\mathcal{M}$ .
- 4: **end for**

**Estimation**

- 5: Initialize  $\mathbf{d}_j^{(0)}[k] = (\mathcal{W}\mathbf{X})_j[k]$ , if  $(j, k) \in \mathcal{M}$ ; 0 otherwise.
  - 6: **for**  $k = 1$  to  $N_{\max}$  **do**
  - 7:    $\tilde{\mathbf{d}} := P_{S_1} \circ Q_{S_2} \mathbf{d}^{(k-1)}$
  - 8:    $\hat{\mathbf{d}} := \mathbf{d}^{(k)} := \text{ST}_{\beta_k}[\tilde{\mathbf{d}}]$ .
  - 9: **end for**
  - 10: Get the estimate  $\hat{\Lambda} = P_+ \mathcal{R} \hat{\mathbf{d}}$ .
- 

synthesis at each iteration and thus has a complexity of  $O(2N_{\max}V)$ , where  $V = O(N \log N)$  is the complexity of UWT and  $N$  is the number of data samples.

#### F. Applications

1) *Simulated biological image restoration:* We have simulated an image containing disk-like isotropic sources on a constant background (see Fig. 3(a)) where the pixel size is  $100nm \times 100nm$ . From the leftmost column to the rightmost one, source radii increase from  $50nm$  to  $350nm$ . This image has been convolved with a Gaussian function with a standard deviation of  $103nm$  which approximates a confocal microscope PSF [44]. The source amplitudes range from 0.08 to 4.99, and the background level is 0.03. This spot grid can be deemed as a model for cellular vesicles of different sizes and intensities. A realization of the photon-count image is shown in Fig. 3(b). We present the restoration results given by Anscombe [4] (Fig. 3(c)), Haar-Fisz [7] (Fig. 3(d)), CVS [10] (Fig. 3(e)), HT-based Haar thresholding [13] (Fig. 3(f)), platelet estimation [45][23][24] (Fig. 3(g)), and the MS-VST denoiser using iterative (Fig. 3(h)) and direct (Fig. 3(i)) reconstructions. IUWT has been used to produce the results in Fig. 3(c)(d)(e)(h)(i); standard Haar UWT is used in Fig. 3(f); cycle spinning with a total of 25 shifts is employed in Fig. 3(d)(g) to attenuate the block artifacts. The controlled FPR in all the wavelet-based methods is set to  $5 \times 10^{-3}$ ; for the platelet approach, the trade-off factor between the likelihood and the penalization  $\gamma$  is set to  $1/3$  (see [24]).

As revealed by Fig. 3, all the estimators perform comparatively well at high intensity levels (right part of the images). For low-intensity sources, Haar-Fisz, CVS, Platelets and the MS-VST are the most

sensitive approaches. We can see that the IUWT-based methods preserve better the isotropic source shapes than the other methods. Some residual noise can be seen in the estimate of CVS.

We also quantify the performances in terms of the normalized mean integrated square error (NMISE) per bin from the denoised signals. The NMISE is defined as:  $\text{NMISE} := \mathbb{E}[(\sum_{i=1}^N (\hat{\lambda}_i - \lambda_i)^2 / \lambda_i) / N]$ , where  $(\hat{\lambda}_i)_i$  is the intensity estimate. Note that the denominator  $\lambda_i$  plays the role of variance stabilization in the error measure. In our experiments, NMISEs are evaluated based on 5 replications. The MS-VST denoiser provides the second lowest error, which is slightly larger than that of the platelet estimate. The platelet estimator offers an efficient piecewise linear approximation to the image. However, on the isolated smooth spots, it tends to alter the isotropic shapes and produces some artifacts. The regularity in the result could be improved by averaging a larger number of cyclic shifts, but leading to a very time-consuming procedure (a computation-time benchmark is shown for a real example in Section V-C2).

Finally, we can also observe that the iterative reconstruction Fig. 3(i) improves restoration of low-flux sources (see the upper part of the image) compared to the direct inverse Fig. 3(j). This phenomenon is clearly expected.

2) *Astronomical image restoration:* Fig. 4 compares the restoration methods on a galaxy image. The FDR control is employed in Anscombe, Haar-Fisz, CVS, Haar HTs, and the MS-VST methods. Among all the results, Haar-Fisz, CVS, Platelets and the MS-VST estimates detect more faint sources. It is found that Haar-Fisz, Haar HTs, Platelets and the MS-VST with iterative construction generate comparable low NMISE values, among which the iterative MS-VST leads to the smallest one.

#### IV. DENOISING BY MS-VST+RIDGELETS

##### A. The Ridgelet Transform

The ridgelet transform [26] has been shown to be very effective for representing global lines in an image. Ridgelet analysis may be constructed as a wavelet analysis in the Radon domain. Recall that the 2D Radon transform of an object  $f$  is the collection of line integrals indexed by  $(\theta, t) \in [0, 2\pi) \times \mathbb{R}$  given by

$$Rf(\theta, t) = \int_{\mathbb{R}^2} f(x_1, x_2) \delta(x_1 \cos \theta + x_2 \sin \theta - t) dx_1 dx_2 \quad (24)$$

where  $\delta$  is the Dirac distribution. Then the ridgelet transform is precisely the application of a 1D wavelet transform to the slices of the Radon transform where the angular variable  $\theta$  is constant and  $t$  is varying. For each scale  $s > 0$ , position  $t \in \mathbb{R}$  and angle  $\theta \in [0, 2\pi)$ , the 2D ridgelet function  $\psi_{s,t,\theta}$  is defined from a 1D wavelet function  $\psi$  as:

$$\psi_{s,t,\theta}(x_1, x_2) = s^{-1/2} \cdot \psi((x_1 \cos \theta + x_2 \sin \theta - t)/s) \quad (25)$$

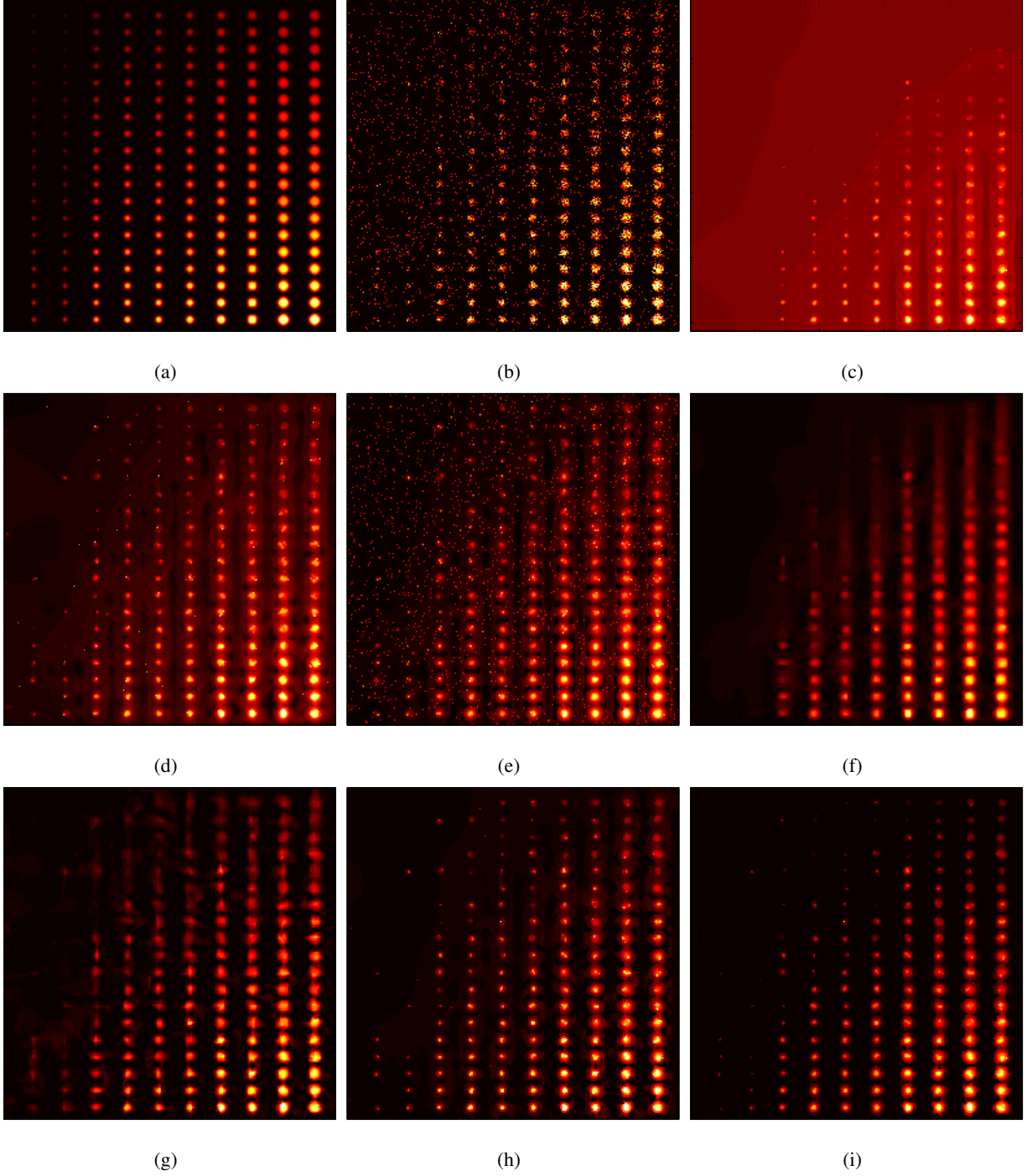


Fig. 3. Denoising an image of simulated spots of different radii (image size:  $256 \times 256$ ). (a) simulated sources (amplitudes  $\in [0.08, 4.99]$ ; background = 0.03); (b) observed counts; (c) Anscombe-denoised image (IUWT,  $J = 5$ ,  $\text{FPR} = 5 \times 10^{-3}$ ,  $\text{NMISE} = 2.34$ ); (d) Haar-Fisz-denoised image (IUWT,  $J = 5$ ,  $\text{FPR} = 5 \times 10^{-3}$ , 25 cyclic shifts where 5 for each of the axes,  $\text{NMISE} = 0.33$ ); (e) CVS-denoised image (IUWT,  $J = 5$ ,  $\text{FPR} = 5 \times 10^{-3}$ ,  $\text{NMISE} = 0.81$ ); (f) image denoised by HT-based Haar thresholding (Haar UWT,  $J = 5$ ,  $\text{FPR} = 5 \times 10^{-3}$ ,  $\text{NMISE} = 0.10$ ); (g) platelet-denoised image ( $\gamma = 1/3$ , 25 random cyclic shifts,  $\text{NMISE} = 0.059$ ); (h) MS-VST-denoised image (IUWT,  $J = 5$ ,  $\text{FPR} = 5 \times 10^{-3}$ ,  $N_{\max} = 20$  iterations,  $\text{NMISE} = 0.069$ ); (i) MS-VST-denoised image (IUWT,  $J = 5$ ,  $\text{FPR} = 5 \times 10^{-3}$ , direct inverse,  $\text{NMISE} = 0.073$ ).



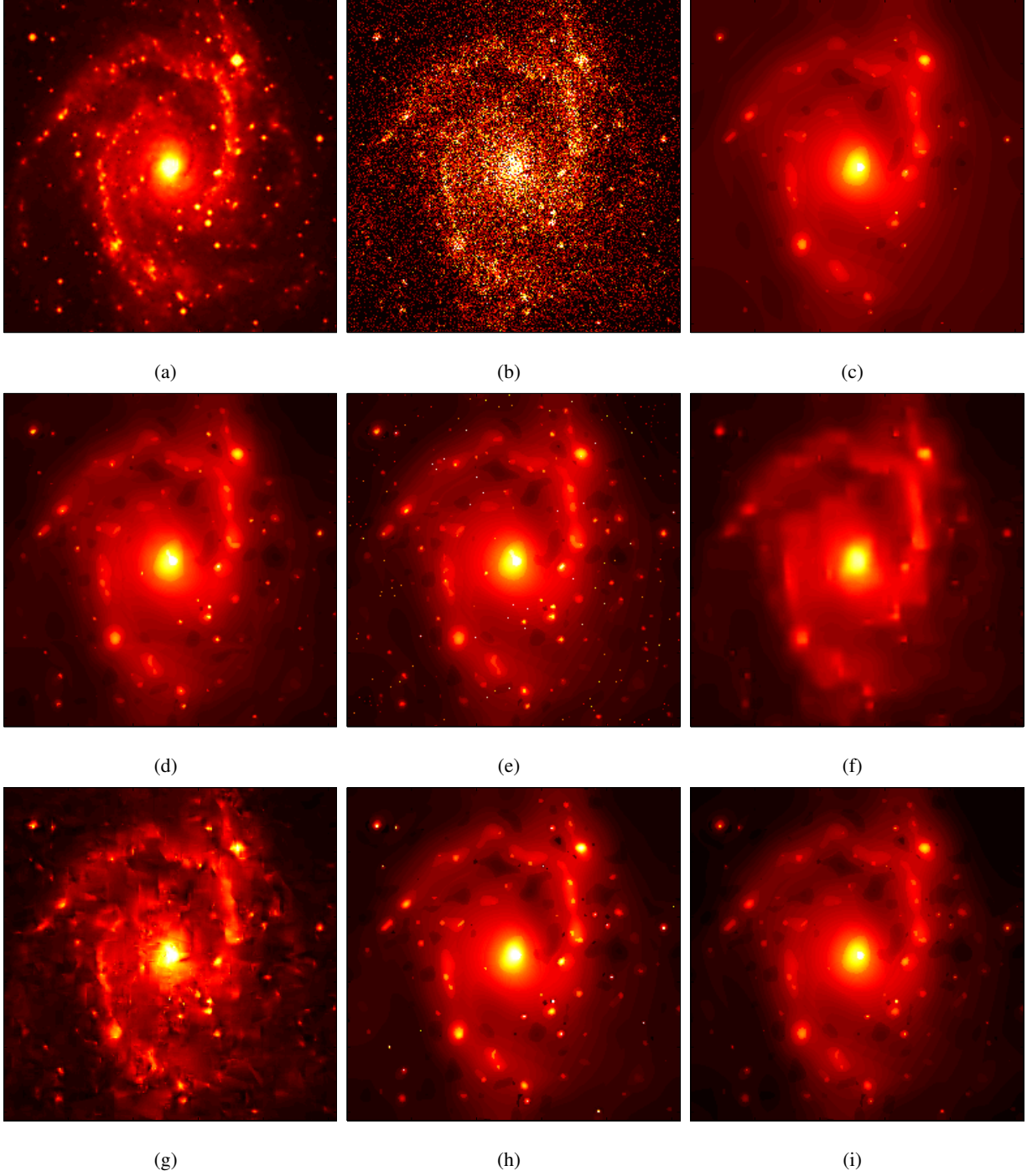


Fig. 4. Denoising a galaxy image (image size:  $256 \times 256$ ). (a) galaxy image (intensity  $\in [0, 5]$ ); (b) observed counts; (c) Anscombe-denoised image (IUWT,  $B_3$ -spline filter bank,  $J = 5$ , FDR = 0.1, NMISE = 0.15); (d) Haar-Fisz-denoised image (IUWT,  $B_3$ -spline filter bank,  $J = 5$ , FDR = 0.1, 25 cyclic shifts where 5 for each of the axes, NMISE = 0.04); (e) CVS-denoised image (IUWT,  $B_3$ -spline filter bank,  $J = 5$ , FDR = 0.1, NMISE = 0.074); (f) denoised image by Haar HTs (Haar UWT,  $J = 5$ , FDR = 0.1, NMISE = 0.036); (g) Platelet-denoised image ( $\gamma = 1/3$ , 25 random cyclic shifts, NMISE = 0.038); (h) MS-VST-denoised image (IUWT,  $B_3$ -spline filter bank,  $J = 5$ , FDR = 0.1,  $N_{max} = 20$  iterations, NMISE = 0.035); (i) MS-VST-denoised image (IUWT,  $B_3$ -spline filter bank,  $J = 5$ , FDR = 0.1, direct inverse, NMISE = 0.051).

A ridgelet is constant along the lines  $x_1 \cos \theta + x_2 \sin \theta = \text{const.}$  Transverse to a ridge is a wavelet.

Thus, the basic strategy for calculating the continuous ridgelet transform is first to compute the Radon transform  $Rf(t, \theta)$  and second, to apply a 1D wavelet transform to the slices  $Rf(\cdot, \theta)$ . Different digital ridgelet transforms can be derived depending on the choice of both the Radon algorithm and the wavelet decomposition [46]. For example, the Slant Stack Radon (SSR) transform [47][48] is a good candidate, which has the advantage of being geometrically accurate, and is used in our experiments. The inverse SSR has however the drawback to be iterative. If computation time is an issue, the recto-polar Radon transform is a good alternative. More details on the implementation of these Radon transforms can be found in [28][47][48][46].

### B. MS-VST with Ridgelets

As a Radon coefficient is obtained from an integration of the pixel values along a line, the noise in the Radon domain follows also a Poisson distribution. Thus, we can apply the 1D MS-VST wavelet detection described in Section III to the slices of the Radon transform. Let  $\mathcal{M} := \{(\theta, j, k)\}$  denote the ridgelet multi-resolution support, where  $(\theta, j, k)$  indicates that the stabilized ridgelet coefficient at projection angle  $\theta$ , scale  $j$  and location  $k$  is significant.  $\mathcal{M}$  being available, we can formulate a constrained  $\ell^1$ -minimization problem in exactly the same way as in the wavelet case (Section III-E), which is then solved by HSD iterations. Hence, the Ridgelet Poisson denoising algorithm consists of the following three steps:

---

#### Algorithm 3 MS-VST + Ridgelets

---

- 1: Apply the Radon transform.
  - 2: For each Radon slice, apply the 1D MS-VST+UWT detection and update  $\mathcal{M}$ .
  - 3: Apply the HSD iterations to the ridgelet coefficients before getting the final estimate.
- 

### C. Results

We have simulated an image with smooth ridges shown in Fig. 5(a). The peak intensities of the vertical ridges vary progressively from 0.1 to 0.5; the inclined ridge has a maximum intensity of 0.3; the background level is 0.05. A Poisson-count image is shown in Fig. 5(b). The biorthogonal 7/9 filter bank [25] is used in the Anscombe (Fig. 5(c)), Haar-Fisz (Fig. 5(d)), CVS (Fig. 5(e)), and MS-VST+UWT (Fig. 5(g)) approaches. Denoising using Haar HTs is given by Fig. 5(f). The estimates by Platelets and by MS-VST+Ridgelets are shown in Fig. 5(h) and Fig. 5(i), respectively. Due to the very low-count setting, the Anscombe estimate is highly biased. Among all the wavelet-based methods, MS-VST+UWT leads

to the smallest error, but is outperformed by the Platelet and the MS-VST-based ridgelet estimates. The two latter methods result in the lowest NMISE values among all the competitors. Clearly, this is because wavelets are less adapted to line-like sources. It can also be seen that the shape of the ridges is better preserved by the ridgelet-based estimate.

## V. DENOISING BY MS-VST+CURVELETS

### A. *The First Generation Curvelet Transform*

The ridgelet transform is efficient for finding only the lines of the size of the image. To detect line segments, a partitioning need to be introduced. The image is first decomposed into smoothly overlapping blocks of side-length  $B$  pixels, and the ridgelet transform is applied independently on each block. This is called the local ridgelet transform. The curvelet transform [49][50] opens the possibility to analyze an image with different block sizes, but with a single transform. The idea is to first decompose the image into a set of wavelet bands using the IUWT, and to analyze each band with a local ridgelet transform. The block size is changed at every other scale. The coarsest resolution of the image ( $a_J$ ) is not processed. This transform has been shown to provide optimal approximation rate for piecewise  $C^2$  images away from  $C^2$  contours, and is very effective in detecting anisotropic structures of different lengths. More details can be found in [49][28].

### B. *MS-VST with Curvelets*

As the first step of the algorithm is an IUWT, we can stabilize each resolution level in the same way as described in Section III-B. We then apply the local ridgelet transform on each stabilized wavelet band. Significant Gaussianized curvelet coefficients will be detected by HTs from which the curvelet multiresolution support  $\mathcal{M}$  is derived. Finally, the same to the wavelet and ridgelet case, we solve a constrained  $\ell^1$ -minimization problem on the curvelet coefficients by HSD iterations before reconstructing the estimate. We now present a sketch of the Poisson curvelet denoising algorithm:

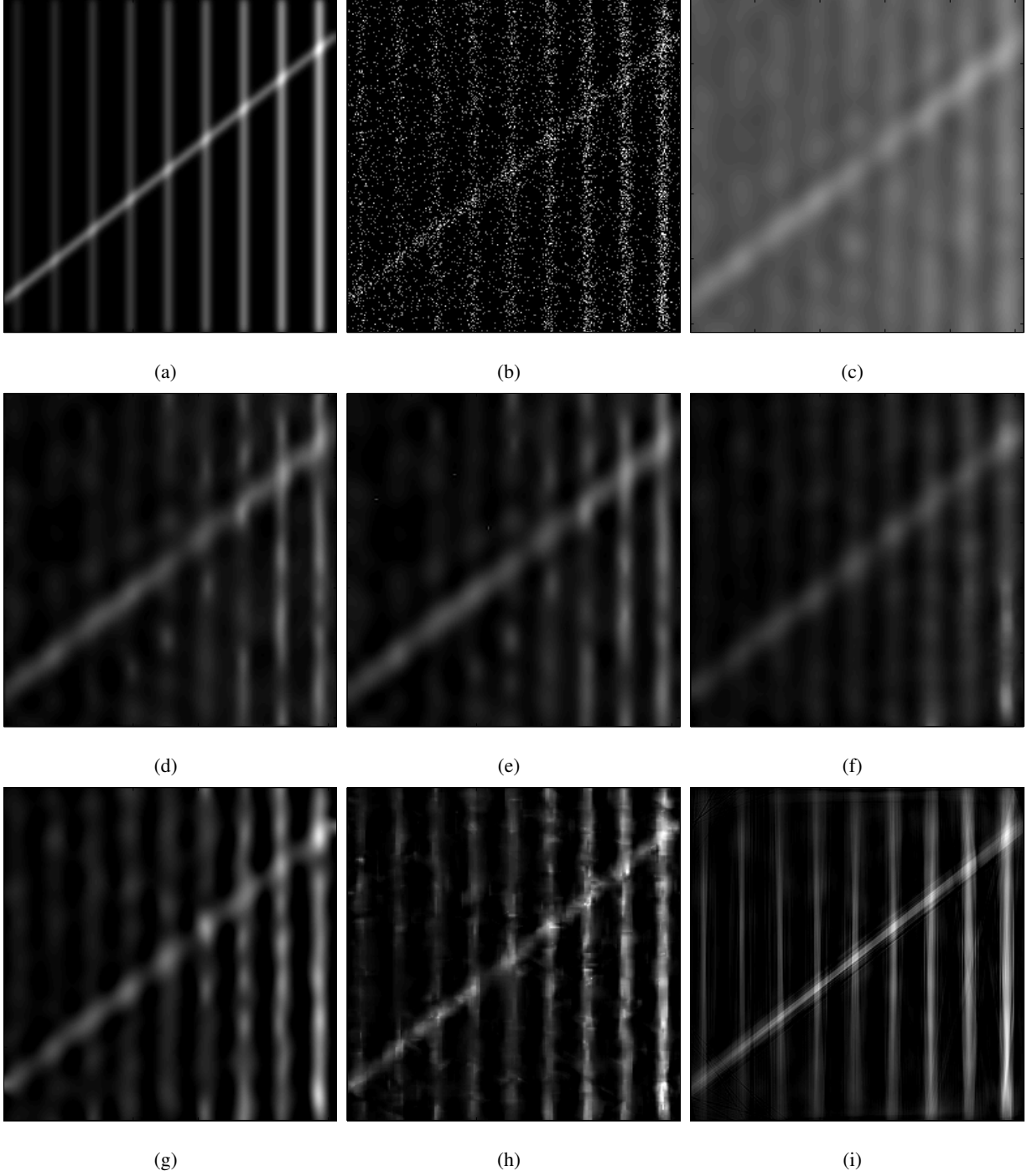


Fig. 5. Poisson denoising of smooth ridges (image size:  $256 \times 256$ ). (a) intensity image (the peak intensities of the 9 vertical ridges vary progressively from 0.1 to 0.5; the inclined ridge has a maximum intensity of 0.3; background = 0.05); (b) Poisson noisy image; (c) Anscombe-denoised image (UWT, 7/9 filter bank,  $J = 4$ ,  $\text{FDR} = 10^{-7}$ ,  $\text{NMISE} = 0.83$ ); (d) Haar-Fisz-denoised image (UWT, 7/9 filter bank,  $J = 4$ ,  $\text{FDR} = 10^{-7}$ , 25 cyclic shifts where 5 for each of the axes,  $\text{NMISE} = 0.035$ ); (e) CVS-denoised image (UWT, 7/9 filter bank,  $J = 4$ ,  $\text{FDR} = 10^{-7}$ ,  $\text{NMISE} = 0.034$ ); (f) image denoised by Haar+FDR ( $J = 4$ ,  $\text{FDR} = 10^{-7}$ ,  $\text{NMISE} = 0.044$ ); (g) image denoised by MS-VST+UWT (7/9 filter bank,  $J = 4$ ,  $\text{FDR} = 10^{-7}$ ,  $N_{max} = 10$  iterations,  $\text{NMISE} = 0.023$ ); (h) Platelet-denoised image ( $\gamma = 1/3$ , 25 random cyclic shifts,  $\text{NMISE} = 0.017$ ); (i) MS-VST+Ridgelets ( $J = 4$ ,  $\text{FDR} = 10^{-7}$ ,  $N_{max} = 10$  iterations,  $\text{NMISE} = 0.017$ ).

November 5, 2007

DRAFT

---

**Algorithm 4** MS-VST + Curvelets
 

---

- 1: Apply the MS-VST+IUWT with  $J$  scales to get the stabilized wavelet subbands  $(d_j)_j$ .
  - 2: set  $B_1 = B_{\min}$
  - 3: **for**  $j = 1$  to  $J$  **do**
  - 4:   Partition the subband  $d_j$  with blocks of side-length  $B_j$  and apply the digital ridgelet transform to each block to obtain the stabilized curvelet coefficients.
  - 5:   HTs on the stabilized curvelet coefficients to obtain  $\mathcal{M}$ .
  - 6:   **if**  $j$  modulo 2 = 1 **then**
  - 7:      $B_{j+1} = 2B_j$
  - 8:   **else**
  - 9:      $B_{j+1} = B_j$
  - 10:   **end if**
  - 11: **end for**
  - 12: Apply the HSD iterations to the curvelet coefficients before getting the final estimate.
- 

It is not as straightforward as with the wavelet and ridgelet transforms to derive the asymptotic noise variance in the stabilized curvelet domain. In our experiments, we derived them using simulated data with Poisson noise only. After having checked that the standard deviation in the curvelet bands becomes stabilized as the intensity level  $\lambda$  increases (which means that the stabilization is working properly), we stored this standard deviation  $\sigma_{j_1, j_2, l}$  for each wavelet scale  $j_1$ , each ridgelet scale  $j_2$ , and each direction angle  $l$ . Then, once the stabilized curvelet transform is applied to our data, these values of  $(\sigma_{j_1, j_2, l})_{j_1, j_2, l}$  serve in the hypothesis testing framework described in Section III-D to test the significance of each stabilized curvelet coefficient at each scale  $(j_1, j_2)$  and direction angle  $l$ .

### C. Applications

1) *Natural image restoration*: Fig. 6 compares different restoration methods on the Barbara image. The original image is heavily scaled down to simulate a low-intensity setting (Fig. 6(a), intensity  $\in [0.93, 15.73]$ ). The FDR control is employed in Anscombe (Fig. 6(c)), Haar-Fisz (Fig. 6(d)), CVS (Fig. 6(e)), Haar HTs (Fig. 6(f)), MS-VST+UWT (Fig. 6(g)), and MS-VST+Curvelet (Fig. 6(i)). As the image is piecewise regular with smooth contours, platelets and curvelets take their full power and provide the best results. In terms of NMISE, MS-VST+Curvelet results in the most accurate estimate. Visually, MS-VST+Curvelet best preserves the fine textures.

2) *Biological image restoration*: Fig. 7 compares the methods on a fluorescent tubulin filaments stained with Bodipy FL goat anti-mouse IgG<sup>1</sup>. The same denoising settings are used as for Fig. 6. MS-VST+UWT

<sup>1</sup>The image is available on the ImageJ site <http://rsb.info.nih.gov/ij>



Fig. 6. Poisson denoising of the Barbara image (image size:  $256 \times 256$ ). (a) intensity image (intensity  $\in [0.93, 15.73]$ ); (b) Poisson noisy image; (c) Anscombe-denoised image (UWT, 7/9 filter bank,  $J = 4$ , FDR = 0.1, NMISE = 0.26); (d) Haar-Fisz-denoised image (UWT, 7/9 filter bank,  $J = 4$ , FDR = 0.1, 25 cyclic shifts where 5 for each of the axes, NMISE = 0.28); (e) CVS-denoised image (UWT, 7/9 filter bank,  $J = 4$ , FDR = 0.1, NMISE = 0.28); (f) denoised image by Haar+FDR (Haar UWT,  $J = 4$ , FDR = 0.1, NMISE = 0.29); (g) denoised image by MS-VST+UWT (UWT, 7/9 filter bank,  $J = 4$ ,  $N_{\max} = 5$  iterations, FDR = 0.1, NMISE = 0.26); (h) platelet-denoised image ( $\gamma = 1/3$ , 25 random cyclic shifts, NMISE = 0.18); (i) denoised image by MS-VST+Curvelets ( $J = 4$ ,  $N_{\max} = 5$  iterations, FDR = 0.1, NMISE = 0.17).

outperforms all the wavelet-based methods; among all the compared approaches, MS-VST+Curvelet leads to the best result both quantitatively and visually. For this example, we also evaluated the computation time of the tested methods on a 1.1GHz PC, giving: Anscombe (C++ codes, 4 sec), Haar-Fisz (C++ codes, 90 sec), CVS (Matlab codes, 3 sec), Haar HTs (C++ codes, 8 sec), MS-VST+UWT (C++ codes, 18 sec), Platelets (Matlab MEX codes, 2404 sec), MS-VST+Curvelet (Matlab codes, 1287 sec). This benchmark shows that our MS-VST+UWT provides a fast solution among the wavelet-based estimators; MS-VST+Curvelet is more computationally intensive but is about one time faster than platelet denoising in our example.

## VI. DISCUSSION AND CONCLUSION

In this paper, we have introduced a novel variance stabilization method and shown that it can be easily combined with various multiscale transforms such as the undecimated wavelet (isotropic and standard), the ridgelet and the curvelet transforms. Based on our multiscale stabilization, we were able to propose a new strategy for removing Poisson noise and our approach enjoys the following advantages:

- It is efficient and sensitive in detecting faint features at a very low-count rate;
- We have the choice to integrate the VST with the multiscale transform we believe to be the most suitable for restoring a given kind of morphological feature (isotropic, line-like, curvilinear, etc);
- The computation time is similar to that of a Gaussian denoising, which makes our denoising method capable of processing large data sets.

Comparison to competing methods in the literature show that the MS-VST is very competitive offering performance as good as state-of-the-art approaches, with low computational burden. This work can be extended along several lines in the future. First, the curvelet denoising could be improved if the VST is applied after the Radon transform in the local ridgelet transform step, rather than on the wavelet coefficients as proposed here. This is however not trivial and requires further investigations. Second, new multiscale transforms have been recently proposed such as the fast curvelet transform [52] and the wave atom transform [53], and it would also be very interesting to investigate how our MS-VST could be linked to them. Finally, here we have considered the denoising with a single multiscale transform only. If the data contains features with different morphologies, it could be better to introduce several multiscale transforms in the denoising algorithm. This could be done in a very similar way as in the Gaussian noise case [54].

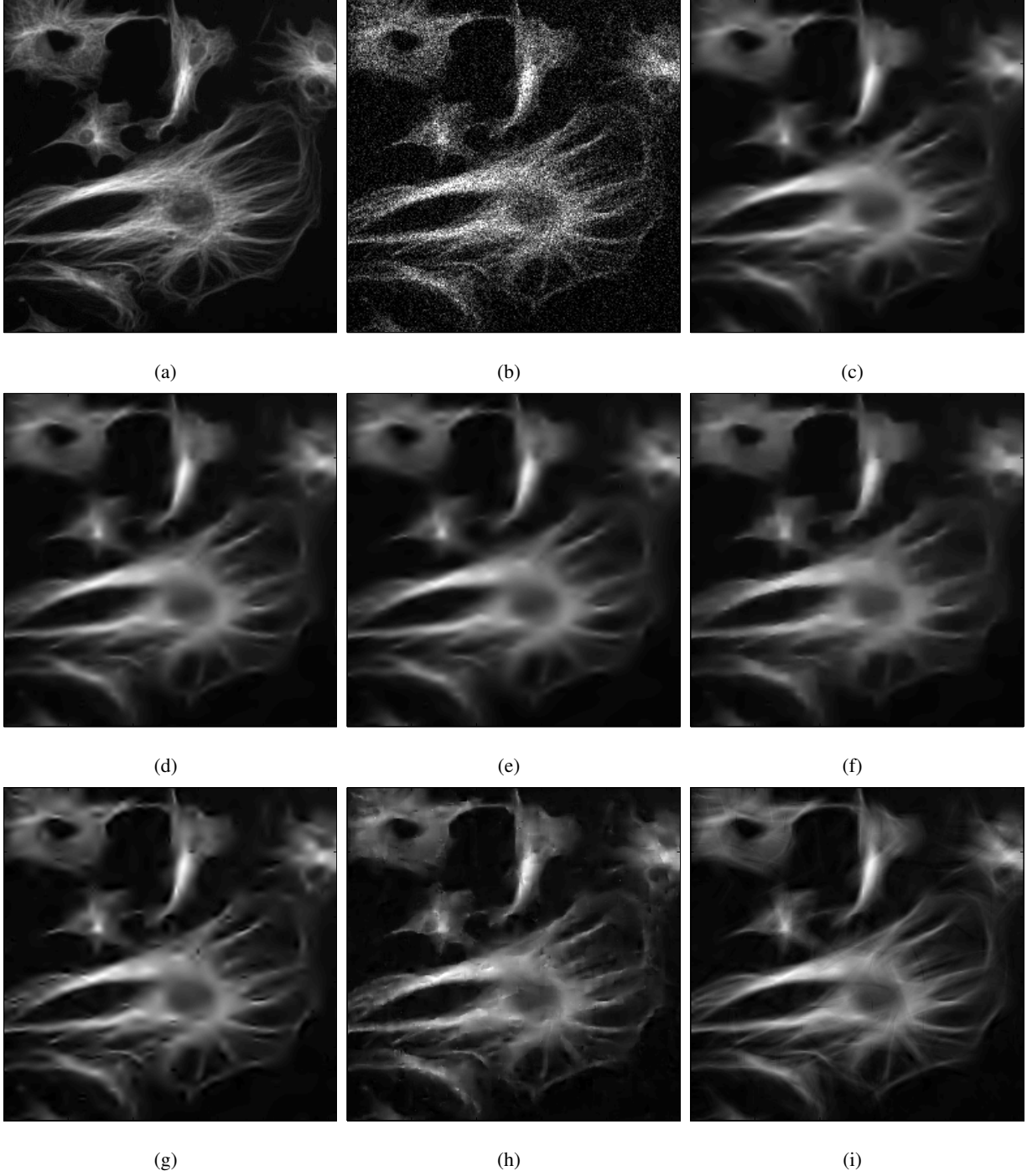


Fig. 7. Poisson denoising of fluorescent tubulins (image size:  $256 \times 256$ ). (a) intensity image (intensity  $\in [0.53, 16.93]$ ); (b) Poisson noisy image; (c) Anscombe-denoised image (UWT, 7/9 filter bank,  $J = 4$ , FDR = 0.1, NMISE = 0.095); (d) Haar-Fisz-denoised image (UWT, 7/9 filter bank,  $J = 4$ , FDR = 0.1, 25 cyclic shifts where 5 for each of the axes, NMISE = 0.096); (e) CVS-denoised image (UWT, 7/9 filter bank,  $J = 4$ , FDR = 0.1, NMISE = 0.10); (f) denoised image by Haar+FDR (Haar UWT,  $J = 4$ , FDR = 0.1, NMISE = 0.10); (g) denoised image by MS-VST+UWT (UWT, 7/9 filter bank,  $J = 4$ ,  $N_{\max} = 5$  iterations, FDR = 0.1, NMISE = 0.090); (h) platelet-denoised image ( $\gamma = 1/3$ , 25 random cyclic shifts, NMISE = 0.079); (i) denoised image by MS-VST+Curvelets ( $J = 4$ ,  $N_{\max} = 5$  iterations, FDR = 0.1, NMISE = 0.078).



## ACKNOWLEDGMENT

This work is partly supported by the Institut Pasteur, CNRS and CEA.

## APPENDIX

## A. Proof of Lemma 1

*Proof:* Suppose a filtered Poisson process  $Y := \sum_i h[i]X_i$ , where  $X_i \sim \mathcal{P}(\lambda)$  and all  $(X_i)_i$  are independent. Assuming  $c \in \mathbb{R}$ ,  $\tau_1 < \infty$ ,  $\tau_2 < +\infty$  and  $\|h\|_3 < +\infty$ , Lévy's continuity theorem shows that

$$\sqrt{\tau_2 \lambda} \left( \frac{Y + c}{\tau_2 \lambda} - \frac{\tau_1}{\tau_2} \right) \xrightarrow[\lambda \rightarrow +\infty]{\mathcal{D}} \mathcal{N}(0, 1) \quad (26)$$

Then, by applying the Delta-method [55] with the function  $f(x) := \text{sgn}(x)\sqrt{|x|}$  and (26), Lemma 1 follows. ■

## B. Proof of Proposition 1

*Proof:* Expand  $T(Y)$  in the neighborhood of  $Y = \mu_Y$ , we obtain

$$T(Y) = b\sqrt{Y + c} = b\sqrt{\mu_Y + c} + b\frac{1}{2}\frac{Y - \mu_Y}{\sqrt{\mu_Y + c}} - b\frac{(Y - \mu_Y)^2}{8(\mu_Y + c)^{3/2}} + \dots + R_s \quad (27)$$

where the Lagrangian form of the remainder  $R_s$  is given by

$$R_s := b\frac{(-1)^{s-1}(2s-3)!!}{2^s s!} \frac{(Y - \mu_Y)^s}{(\xi + c)^{s-1/2}} \quad (s > 1) \quad (28)$$

with  $\xi$  strictly between  $\mu_Y$  and  $Y$ . The following lemma gives an asymptotic bound on the expectation of the remainder  $R_s$ .

**Lemma 2** Consider  $Y := \sum_i h[i]X_i$  a filtered Poisson process where  $h$  is a nonnegative FIR filter with  $\tau_1 > 0$ . If  $s > 1$  and  $c > 0$ , then  $\mathbb{E}[|R_s|] = O_{\lambda \rightarrow +\infty}(\lambda^{-\frac{s-1}{2}})$ .

Proposition 1 results immediately from Lemma 2. Using (27) and (28), we can derive the Taylor expansion of  $\mathbb{E}[Z]$  about  $\lambda = +\infty$  up to order  $s = 3$ . Then, (3) follows from Lemma 2. (4) can be proved similarly. (ii) can be easily verified, and the last statement (iii) follows from Lemma 1. ■

It remains to prove Lemma 2. We will make use of the Cramér-Chernoff inequality [56].

**Lemma 3 (Cramér-Chernoff)** Let  $(X_i)_{1 \leq i \leq n}$  be  $n$  i.i.d. real random variables. Consider the sum  $S_n := \sum_{i=1}^n X_i$ . Let  $M(t) := \mathbb{E}[e^{tX_1}]$  be the moment generating function (mgf) of  $X_1$  and define  $I_X(x) := \sup_{t \in \mathbb{R}} (tx - \log M(t))$  for  $x \in \mathbb{R}$  ( $I_X$  is thus  $[0, +\infty]$  valued). Then, we have for all  $n \geq 1$ ,

$$\Pr(S_n \leq nx) \leq e^{-nI_X(x)}, \quad x \leq \mathbb{E}[X_1]$$

$I_X(x)$  is strictly positive if  $x \neq \mathbb{E}[X_1]$ . It can also be shown that  $F(t)$  is concave and is strictly concave if  $X_i$  is not almost surely a constant. Now, we have the following lemma,

**Lemma 4** *Consider a filtered Poisson process  $Y := \sum_{i=1}^n h[i]U_i$  where  $U_i \sim \mathcal{P}(\lambda)$  are independent, and  $h$  is a filter of length  $n$  with  $\tau_1 > 0$ . Then, for all  $c^* \in (0, \tau_1/\sqrt{\tau_2})$ , there exists  $\beta > 0$  depending only on  $h$  and  $c^*$  such that,*

$$\Pr(Y \leq \lambda(\tau_1 - c^* \sqrt{\tau_2})) \leq e^{-\lambda\beta}$$

*Proof:* Rewrite  $Y$  as follows:

$$Y := \sum_{i=1}^n h[i]U_i = \sum_{i=1}^n h[i] \sum_{j=1}^{\lambda/a} W_{i,j} = \sum_{j=1}^{\lambda/a} \sum_{i=1}^n h[i]W_{i,j} = \sum_{j=1}^{\lambda/a} T_j, \quad T_j := \sum_{i=1}^n h[i]W_{i,j}$$

where  $\exists a > 0$  such that  $\lambda/a \in \mathbb{N}$  and  $W_{i,j} \sim \mathcal{P}(a)$  are i.i.d. Poisson variables. It can be noted that  $(T_j)_j$  are also i.i.d. variables. We will apply Lemma 3 on  $Y$ . First let us calculate  $I_T(x)$  as follows:

$$I_T(x) := \sup_{t \in \mathbb{R}} (tx - \log M_T(t)) = \sup_{t \in \mathbb{R}} \left( tx - \sum_{i=1}^n a (e^{h[i]t} - 1) \right) \quad (29)$$

where  $M_T$  is the mgf of  $T_1$ . We will evaluate  $I_T(x)$  at  $x_0 := a(\tau_1 - c^* \sqrt{\tau_2}) > 0$ . Since  $T_1$  is not almost surely a constant,  $I_T(x_0)$  must be attained at a unique  $t_0$ . Thus, setting  $x = x_0$ , we take the derivative of the sup argument in (29) and set it to zero, resulting in the equation necessarily satisfied by  $t_0$ :

$$\sum_{i=1}^n h[i] (1 - e^{h[i]t_0}) = c^* \sqrt{\tau_2} \quad (30)$$

$I_T(x_0)$  is given by:

$$I_T(x_0) = a\beta, \quad \beta = t_0(\tau_1 - c^* \sqrt{\tau_2}) - \sum_{i=1}^n (e^{h[i]t_0} - 1) \quad (31)$$

Both (30) and (31) show that  $t_0$  and  $\beta$  depend only on  $h$  and  $c^*$ . We have in addition  $I_T(x_0) > 0$ , since  $x_0 < \tau_1 a$ . We can now apply Lemma 3, giving:

$$\Pr(Y \leq x_0 \lambda/a) = \Pr(Y \leq \lambda(\tau_1 - c^* \sqrt{\tau_2})) \leq e^{-I_T(x_0)\lambda/a} = e^{-\lambda\beta}$$

■

Now we are at the point to prove Lemma 2.

*Proof:* It can be seen from (28) that  $R_s$  satisfies:

$$|R_s| \leq B_s := \frac{|b| |Y - \mu_Y|^s}{2s |\xi + c|^{s-\frac{1}{2}}} \quad (32)$$

Denote  $\mu_Y := \lambda\tau_1$  and  $\sigma_Y := \sqrt{\lambda\tau_2}$ . We have,

$$\begin{aligned}\mathbb{E}[B_s] &= \int \mathbf{1}(y \geq \mu_Y - c^* \lambda^{\frac{1}{2}} \sigma_Y) B_s dP_Y + \int \mathbf{1}(0 \leq y < \mu_Y - c^* \lambda^{\frac{1}{2}} \sigma_Y) B_s dP_Y \\ &\leq \frac{|b|}{2s} \frac{\mathbb{E}[|Y - \mu_Y|^s]}{(\mu_Y - c^* \lambda^{\frac{1}{2}} \sigma_Y + c)^{s-\frac{1}{2}}} + \frac{|b|}{2s} \frac{\mu_Y^s}{c^{s-\frac{1}{2}}} \Pr(0 \leq Y < \lambda(\tau_1 - c^* \sqrt{\tau_2})) \\ &\leq \frac{|b|}{2s} \frac{\mathbb{E}[|Y - \mu_Y|^s]}{(\lambda(\tau_1 - c^* \sqrt{\tau_2}) + c)^{s-\frac{1}{2}}} + \frac{|b|}{2s} \frac{\lambda^s \tau_1^s}{c^{s-\frac{1}{2}}} \cdot e^{-\lambda\beta}\end{aligned}\quad (33)$$

where there exists  $c^* \in (0, \tau_1/\sqrt{\tau_2})$  and the second term in (33) results from Lemma 4. Then,

$$(33) = \mathbb{E}[|Y - \mu_Y|^s] \cdot O_{\lambda \rightarrow +\infty}(\lambda^{-s+\frac{1}{2}})$$

We will conclude by showing that  $\tilde{M}_s := \mathbb{E}[|Y - \mu_Y|^s] = O_{\lambda \rightarrow +\infty}(\lambda^{s/2})$ . The moment  $M_n$  and the cumulant  $\kappa_n$  of the centered random variable  $(Y - \mu_Y)$  are related by:

$$M_n = \kappa_n + \sum_{p=2}^{n-2} C_{n-1}^p M_p \kappa_{n-p} \quad (n \geq 2) \quad (34)$$

It can be shown by induction that  $M_n$  is a polynomial of  $\kappa_2, \dots, \kappa_n$ , which has a minimal order 1 and a maximal order  $\lfloor n/2 \rfloor$ . The  $p$ -th cumulant of  $(Y - \mu_Y)$  is  $\kappa_p = \lambda\tau_p$  for  $p \geq 2$ . Therefore  $M_n = O_{\lambda \rightarrow +\infty}(\lambda^{\lfloor n/2 \rfloor})$ . Consequently,  $\tilde{M}_k$  satisfies:

$$\begin{aligned}\tilde{M}_{2k} &:= \mathbb{E}[|Y - \mu_Y|^{2k}] = M_{2k} = O_{\lambda \rightarrow +\infty}(\lambda^k) \\ \tilde{M}_{2k+1} &:= \mathbb{E}[|Y - \mu_Y|^{2k+1}] = \mathbb{E}[|Y - \mu_Y|^k |X - \mu_Y|^{k+1}] \leq M_{2k}^{1/2} M_{2k+2}^{1/2} = O_{\lambda \rightarrow +\infty}(\lambda^{\frac{2k+1}{2}})\end{aligned}$$

This shows that  $\tilde{M}_s = O_{\lambda \rightarrow +\infty}(\lambda^{s/2})$ . ■

### C. Proofs of Theorem 1 and 2

We will prove Theorem 1 below, and Theorem 2 can be proved in the same way.

*Proof:* Let  $\mathbf{F}_j := [a_{j-1} + c^{(j-1)}, a_j + c^{(j)}]^T$  and  $\boldsymbol{\mu}_j := [\tau_1^{(j-1)}, \tau_1^{(j)}]^T$ . Suppose  $\tau_1^{(j-1)}, \tau_1^{(j)} < \infty$ ,  $0 < \tau_2^{(j-1)}, \tau_2^{(j)} < +\infty$ , and  $\|h^{(j-1)}\|_3, \|h^{(j)}\|_3 < +\infty$ . Then Lévy's continuity theorem results in

$$\sqrt{\lambda} \left( \frac{\mathbf{F}_j}{\lambda} - \boldsymbol{\mu}_j \right) \xrightarrow[\lambda \rightarrow +\infty]{\mathcal{D}} \mathcal{N}(0, \Sigma_j), \quad \Sigma_j = \begin{bmatrix} \tau_2^{(j-1)} & \langle h^{(j-1)}, h^{(j)} \rangle \\ \langle h^{(j-1)}, h^{(j)} \rangle & \tau_2^{(j)} \end{bmatrix} \quad (35)$$

Define  $g(x_1, x_2) := b^{(j-1)} \text{sgn}(x_1) \sqrt{|x_1|} - b^{(j)} \text{sgn}(x_2) \sqrt{|x_2|}$ . We obtain the desired result by applying the multivariate Delta-method with the function  $g$  and (35). ■

#### D. proof of Theorem 3

We will first need to prove Lemma 5. Given a Hilbert space  $\mathcal{H}$  with inner product  $\langle \cdot, \cdot \rangle_{\mathcal{H}}$  and induced norm  $\| \cdot \|_{\mathcal{H}}$ , we call a mapping  $V : \mathcal{H} \rightarrow \mathcal{H}$  *nonexpansive* if for all  $x, y \in \mathcal{H}$ ,  $\|Vx - Vy\|_{\mathcal{H}} \leq \|x - y\|_{\mathcal{H}}$ . Suppose that a mapping  $V : \mathcal{H} \rightarrow \mathcal{H}$  is nonexpansive and  $\text{Fix}(V) \neq \emptyset$ . Then  $V$  is *attracting* (w.r.t.  $\text{Fix}(V)$ ) if for every  $x \notin \text{Fix}(V), y \in \text{Fix}(V)$ , we have  $\|Vx - y\|_{\mathcal{H}} < \|x - y\|_{\mathcal{H}}$ . Properties of nonexpansive and attracting mappings can be found in [57]. For a given set  $S \subset \mathcal{H}$ , a mapping  $V : \mathcal{H} \rightarrow \mathcal{H}$  is  $\eta$ -strongly monotone over  $S$  if there exists  $\eta > 0$  such that  $\langle Vx - Vy, x - y \rangle_{\mathcal{H}} \geq \eta \|x - y\|_{\mathcal{H}}^2$  for all  $x, y \in S$ . Let us point out that in our case,  $\mathcal{H}$  is  $\mathbb{R}^L$ .

**Lemma 5** *With the same notations as in Theorem 3, we have:*

- (a)  $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$  and  $\mathcal{C}_B$  are all closed convex nonempty sets;
- (b)  $P_{\mathcal{S}_1}$  and  $P_{\mathcal{S}_3}$  are attracting, and  $\text{Fix}(P_{\mathcal{S}_1}) = \mathcal{S}_1$  and  $\text{Fix}(P_{\mathcal{S}_3}) = \mathcal{S}_3$ ;
- (c)  $\text{Fix}(Q_{\mathcal{S}_2}) = \mathcal{S}_2$ , and if  $\mathcal{W}$  represents a tight frame and  $\mathcal{R}$  is the pseudo-inverse operator, then  $Q_{\mathcal{S}_2}$  is nonexpansive;
- (d) If  $V_1$  is attracting,  $V_2$  is nonexpansive, and  $\text{Fix}(V_1) \cap \text{Fix}(V_2) \neq \emptyset$ , then  $V := V_1 \circ V_2$  is nonexpansive with  $\text{Fix}(V) = \text{Fix}(V_1) \cap \text{Fix}(V_2)$ .

*Proof:* (a) and (b) can be easily verified. (c) results from the fact that  $\|\mathcal{W}\|\|\mathcal{R}\| = 1$  ([25]) and that  $P_+$  is a projector (so nonexpansive). To prove (d),  $V$  can be easily verified to be nonexpansive. It is obvious that  $\text{Fix}(V_1) \cap \text{Fix}(V_2) \subseteq \text{Fix}(V_1 \circ V_2)$ . To prove the other inclusion, pick  $x \in \text{Fix}(V_1 \circ V_2)$ . It is sufficient to show that  $x \in \text{Fix}(V_2)$ . Suppose that  $x \notin \text{Fix}(V_2)$ , then necessarily  $V_2x \notin \text{Fix}(V_1)$ . Now pick any  $y \in \text{Fix}(V_1) \cap \text{Fix}(V_2)$ . Since  $V_1$  is attracting, we have:

$$\|x - y\|_{\mathcal{H}} = \|V_1 \circ V_2x - y\|_{\mathcal{H}} < \|V_2x - y\|_{\mathcal{H}} = \|V_2x - V_2y\|_{\mathcal{H}} \leq \|x - y\|_{\mathcal{H}}$$

which is absurd. Thus  $\text{Fix}(V_1) \cap \text{Fix}(V_2) = \text{Fix}(V_1 \circ V_2) = \text{Fix}(V)$ . ■

Let us now prove Theorem 3.

*Proof:* (i) can be easily verified. (ii) is a direct result of Lemma 5(d). To prove (iii), we note that  $J_{\epsilon}$  is convex and  $\nabla J_{\epsilon}(\mathbf{d})[i] = d[i](d[i]^2 + \epsilon)^{-1/2}$ . It can be verified that  $\nabla J_{\epsilon}(\mathbf{d})$  is  $\epsilon^{-1/2}$ -Lipschitzian and  $\epsilon(B^2 + \epsilon)^{-3/2}$ -strongly monotone over  $T_{\mathcal{C}_B}(\mathbb{R}^N)$ . Then (iii) results from the convergence theorem of HSD [43]. (iv) is obvious. To prove (v), we have for any convergent subsequence of  $d_{\epsilon}^*$ , say  $d_{\epsilon_j}^* \xrightarrow{\epsilon \rightarrow 0^+} d_0^*$ , that

$$\forall \mathbf{d} \in \mathcal{C}_B, \quad J(\mathbf{d}_{\epsilon_j}^*) = \sum_{i=1}^N |\mathbf{d}_{\epsilon_j}^*[i]| \leq \sum_{i=1}^N \sqrt{\mathbf{d}_{\epsilon_j}^*[i]^2 + \epsilon_j} \leq \sum_{i=1}^N \sqrt{\mathbf{d}[i]^2 + \epsilon_j} \quad (36)$$

Then by taking the limit  $\epsilon \rightarrow 0^+$  on both sides of (36), we have  $\|\mathbf{d}_0^*\|_1 \leq \|\mathbf{d}\|_1$ .  $\mathbf{d}_0^* \in \mathcal{C}_B$  since  $\mathcal{C}_B$  is closed.  $\mathbf{d}_0^*$  is thus a solution to (19) with  $\epsilon = 0$ , and hence also a solution to (18) by (i). ■

## REFERENCES

- [1] J.-L. Starck, F. Murtagh, and A. Bijaoui, Image Processing and Data Analysis: The Multiscale Approach. Cambridge University Press, 1998.
- [2] J. Pawley, Handbook of Biological Confocal Microscopy, 3rd ed. Springer, 2006.
- [3] F. J. Anscombe, “The Transformation of Poisson, Binomial and Negative-Binomial Data,” Biometrika, vol. 35, pp. 246–254, 1948.
- [4] D. L. Donoho, “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.
- [5] —, “Ideal spatial adaptation by wavelet shrinkage,” Biometrika, vol. 81, pp. 425–455, 1994.
- [6] —, “De-noising by soft-thresholding,” IEEE Transactions on Information Theory, vol. 41, no. 3, pp. 613–627, 1995.
- [7] P. Fryzlewicz and G. P. Nason, “A Haar-Fisz algorithm for Poisson intensity estimation,” J. Comp. Graph. Stat., vol. 13, pp. 621–638, 2004.
- [8] P. Fryzlewicz, V. Delouille, and G. P. Nason, “GOES-8 X-ray sensor variance stabilization using the multiscale data-driven Haar-Fisz transform,” J. Roy. Statist. Soc. ser. C, vol. 56, no. 1, pp. 99–116, 2007.
- [9] 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.
- [10] M. Jansen, “Multiscale Poisson data smoothing,” J. Roy. Statist. Soc. ser. B, vol. 68, no. 1, pp. 27–48, 2006.
- [11] 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.
- [12] A. Antoniadis and T. Sapatinas, “Wavelet shrinkage for natural exponential families with quadratic variance functions,” Biometrika, vol. 88, pp. 805–820, 2001.
- [13] E. D. Kolaczyk, “Nonparametric estimation of intensity maps using Haar wavelets and Poisson noise characteristics,” The Astrophysical Journal, vol. 534, pp. 490–505, 2000.
- [14] B. Zhang, M. J. Fadili, J.-L. Starck, and S. W. Digel, “Fast Poisson Noise Removal by Biorthogonal Haar Domain Hypothesis Testing,” Statistical Methodology, 2007, revised.
- [15] A. Bijaoui and G. Jammal, “On the distribution of the wavelet coefficient for a Poisson noise,” Signal Processing, vol. 81, pp. 1789–1800, 2001.
- [16] E. D. Kolaczyk, “Wavelet shrinkage estimation of certain Poisson intensity signals using corrected thresholds,” Statist. Sinica, vol. 9, pp. 119–135, 1999.
- [17] —, “Bayesian multiscale models for Poisson processes,” J. Amer. Statist. Ass., vol. 94, no. 447, pp. 920–933, Sep. 1999.
- [18] K. E. Timmermann and R. D. Nowak, “Multiscale Modeling and Estimation of Poisson Processes with Application to Photon-Limited Imaging,” IEEE Trans. Inf. Theo., vol. 45, no. 3, pp. 846–862, Apr. 1999.
- [19] 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.
- [20] 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.

- [21] 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, Jun. 2004.
- [22] R. M. Willett and R. D. Nowak, “Fast, Near-Optimal, Multiresolution Estimation of Poisson Signals and Images,” in EUSIPCO, San Diego, CA, 2004.
- [23] —, “Multiscale Poisson Intensity and Density Estimation,” Duke University, Tech. Rep., 2005.
- [24] R. Willett, “Multiscale Analysis of Photon-Limited Astronomical Images,” SCMA IV, 2006, in press.
- [25] S. G. Mallat, A Wavelet Tour of Signal Processing, 2nd ed. Academic Press, 1998.
- [26] E. Candès and D. Donoho, “Ridgelets: the key to high dimensional intermittency?” Philosophical Transactions of the Royal Society of London A, vol. 357, pp. 2495–2509, 1999.
- [27] E. J. Candès and D. L. Donoho, “Curvelets – A Surprisingly Effective Nonadaptive Representation For Objects with Edges,” in Curve and Surface Fitting: Saint-Malo 1999, A. Cohen, C. Rabut, and L. Schumaker, Eds. Nashville, TN: Vanderbilt University Press, 1999.
- [28] J.-L. Starck, E. Candès, and D. Donoho, “The Curvelet transform for image denoising,” IEEE Transactions on Image Processing, vol. 11, no. 6, pp. 131–141, 2002.
- [29] G. P. Nason and B. W. Silverman, Lecture Notes in Statistics. Springer-Verlag, 1995, vol. 103, ch. The stationary wavelet transform and some statistical applications, pp. 281–299.
- [30] R. R. Coifman and D. L. Donoho, Lecture Notes in Statistics. Springer-Verlag, 1995, vol. 103, ch. Translation-invariant de-noising, pp. 125–150.
- [31] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian, “A Real-Time Algorithm for Signal Analysis with the Help of the Wavelet Transform,” in Wavelets: Time-Frequency Methods and Phase-Space. Springer-Verlag, 1989, pp. 286–297.
- [32] M. J. Shensa, “Discrete wavelet transforms: Wedding the à trous and Mallat algorithms,” IEEE Transactions on Signal Processing, vol. 40, pp. 2464–2482, 1992.
- [33] J.-L. Starck, M. Fadili, and F. Murtagh, “The Undecimated Wavelet Decomposition and its Reconstruction,” IEEE Transactions on Image Processing, vol. 16, no. 2, pp. 297–309, 2007.
- [34] J. C. Olivo-Marin, “Extraction of spots in biological images using multiscale products,” Pattern Recognition, vol. 35, no. 9, pp. 1989–1996, 2002.
- [35] A. Antoniadis, J. Bigot, and T. Sapatinas, “Wavelet Estimators in Nonparametric Regression: A Comparative Simulation Study,” Journal of Statistical Software, vol. 6, no. 6, 2001.
- [36] 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.
- [37] 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.
- [38] F. Abramovich, Y. Benjamini, D. Donoho, and I. Johnstone, “Adapting to Unknown Sparsity by controlling the False Discovery Rate,” Annals of Statistics, vol. 34, no. 2, pp. 584–653, 2006.
- [39] D. L. Donoho and J. Jin, “Asymptotic minimaxity of false discovery rate thresholding for sparse exponential data,” Annals of Statistics, vol. 34, no. 6, pp. 2980–3018, 2006.
- [40] J. L. Starck, A. Bijaoui, and F. Murtagh, “Multiresolution Support Applied to Image Filtering and Deconvolution,” CVIP: Graphical Models and Image Processing, vol. 57, no. 5, pp. 420–431, Sep. 1995.

- [41] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via  $\ell^1$  minimization," PNAS, vol. 100, no. 5, pp. 2197–2202, 2003.
- [42] D. L. Donoho, "For Most Large Underdetermined Systems of Linear Equations, the minimal  $\ell^1$ -norm solution is also the sparsest solution," Department of Statistics of Stanford Univ., Tech. Rep., 2004.
- [43] I. Yamada, "The Hybrid Steepest Descent Method for the Variational Inequality Problem over the Intersection of Fixed Point Sets of Nonexpansive Mappings," in Inherently Parallel Algorithm in Feasibility and Optimization and their Applications. New York: Elsevier, 2001.
- [44] B. Zhang, J. Zerubia, and J.-C. Olivo-Marin, "Gaussian approximations of fluorescence microscope point-spread function models," Applied Optics, vol. 46, no. 10, pp. 1819–1829, 2007.
- [45] R. M. Willett and R. D. Nowak, "Platelets: a multiscale approach for recovering edges and surfaces in photon-limited medical imaging," IEEE Trans. Med. Imag., vol. 22, no. 3, pp. 332–350, 2003.
- [46] J.-L. Starck, M. Elad, and D. Donoho, "Redundant multiscale transforms and their application for morphological component analysis," Advances in Imaging and Electron Physics, vol. 132, 2004.
- [47] A. Averbuch, R. R. Coifman, D. L. Donoho, M. Israeli, and J. Waldén, "Fast Slant Stack: A notion of Radon transform for data in a cartesian grid which is rapidly computible, algebraically exact, geometrically faithful and invertible," SIAM J. Sci. Comput., 2001, to appear.
- [48] D. Donoho and A. Flesia, "Digital Ridgelet Transform Based on True Ridge Functions," in Beyond Wavelets, J. Schmeidler and G. Welland, Eds. Academic Press, 2002.
- [49] D. L. Donoho and M. R. Duncan, "Digital curvelet transform: strategy, implementation and experiments," in Proc. Aerosense 2000, Wavelet Applications VII, H. H. Szu, M. Vetterli, W. Campbell, and J. R. Buss, Eds., vol. 4056. SPIE, 2000, pp. 12–29.
- [50] J.-L. Starck, E. Candes, and D. Donoho, "Astronomical image representation by the curvelet transform," Astronomy and Astrophysics, vol. 398, pp. 785–800, 2003.
- [51] [Http://http://rsb.info.nih.gov/ij](http://rsb.info.nih.gov/ij).
- [52] E. Candes, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," SIAM Multiscale Model. Simul., vol. 5, no. 3, pp. 861–899, 2006.
- [53] L. Demanet and L. Ying, "Wave Atoms and sparsity of oscillatory patterns," Appl. Comput. Harmon. Anal., 2006, submitted.
- [54] J. L. Starck, "Very high quality image restoration by combining wavelets and curvelets," Proc. SPIE, vol. 4478, pp. 9–19, Dec. 2001.
- [55] P. J. Bickel and K. A. Doksum, Mathematical statistics: basic ideas and selected topics, 2nd ed. Prentice-Hall, London, 2001, vol. I.
- [56] S. Varadhan, Large Deviations and Applications. Philadelphia: SIAM, 1984.
- [57] H. H. Bauschke and J. M. Borwein, "On Projection Algorithms for Solving Convex Feasibility Problems," SIAM Review, vol. 38, no. 3, pp. 367–426, 1996.