# MULTI-SCALE VARIANCE STABILIZING TRANSFORM FOR MULTI-DIMENSIONAL POISSON COUNT IMAGE DENOISING

*B. Zhang[a], M.J. Fadili[b] and J-L. Starck[c]*

[a] Quantitative Image Analysis Unit
Institut Pasteur 75724 Paris France

[b] GREYC UMR CNRS 6072
14050 Caen France

[c] DAPNIA/SEDI-SAP CEA-Saclay
91191 Gif-sur-Yvette France

## ABSTRACT

We propose in this paper a Multi-Scale Variance Stabilizing Transform (MSVST) for approximately Gaussianizing and stabilizing the variance of a sequence of independent Poisson random variables (RVs) filtered by a low-pass linear filter. This approach is shown to be fast, very well adapted to extremely low-count situations and easily applicable to any dimensional data. It is shown that the RV transformed using Anscombe VST can be reasonably considered as stabilized for an intensity $\lambda \gtrsim 10$, using Fisz VST for $\lambda \gtrsim 1$ and using our VST (after low-pass filtering) for $\lambda \gtrsim 0.1$. We then use the MSVST technique to stabilize the detail coefficients of the Isotropic Undecimated Wavelet Transform (IUWT) of multi-dimensional Poisson count data. We use a hypothesis testing framework in the wavelet domain to denoise the Gaussianized and stabilized coefficients, and then apply the inverse MSVST-IUWT to get the estimated intensity image underlying the Poisson data. Finally, potential applicability of our approach is illustrated on an astronomical example where isotropic structures must be recovered.

## 1. INTRODUCTION

The ability to restore the underlying intensity from an inhomogeneous Poisson process is crucial for many applications. We observe a discrete dataset of counts $\mathbf{x} = (X_n)_{n \in I}$ where $I$ is the index set. Each count $X_n$ is independently Poisson distributed with a mean $\lambda_n$, i.e. $X_n \sim \mathcal{P}(\lambda_n)$.

A host of estimation methods have been proposed in the literature. A common solution is to use a variance stabilizing transform (VST), which "Gaussianizes" the Poisson noise before applying the standard wavelet thresholding denoising on the transformed signal. For example, Anscombe transform [1] and Fisz transform are respectively proposed in [2] and [3]. Besides nonlinear VST, direct wavelet filtering has been studied in [4, 5]. The state-of-the-art methods are Bayesian approaches (see overview in [6]). A great part of the above methods are based on the Haar wavelet transform until recently, Jansen [7] introduced a conditional variance stabilization, which generalizes the idea in [3] and is applicable to any family of wavelet transforms. A Bayesian scheme was also derived within this framework, which can be deemed as an extension of [8, 9, 10]. Other technique such as penalized maximum likelihood estimation [11, 12] can also be considered Bayesian, since the penalization term implicitly introduces a prior on the underlying intensity. In general, Bayesian methods outperform those of direct wavelet filtering [4, 5]. However Bayesian approaches require having enough "useful signals" in the observations in order to well fit the prior (usually by estimating prior parameters). In the low-intensity case where we generally lack such "useful signals", the final estimation can suffer from a large bias produced by an improperly fitted model. Another previous important contribution to this field is the wavelet-domain hypothesis testing framework first introduced in [13, 14]. Thresholds based on user-specified false detection rate are derived for Haar coefficients both in the constant and the model-based background situations. Universal thresholds are also found available in the above two cases [14, 15]. However, this method is only adapted to piecewise constant and burst like intensities as it uses the Haar wavelet, which will yield stair-case-like artifacts in estimating regular intensities. Although Kolaczyk [16] derived the Poisson-corrected version of the Gaussian-based threshold for any wavelet, the asymptotic approximation used in [16] may not allow reasonable threshold solution in very low intensity settings. A recent work [17] has tackled these two drawbacks by proposing a more sensitive bi-orthogonal Haar domain hypothesis testing procedure.

In this paper, we propose a VST for approximately Gaussianizing and stabilizing the variance of a sequence of independent Poisson random variables (RVs) filtered by a low-pass linear filter. This approach is shown to be fast, very well adapted to extremely low-count situations and easily applicable to any dimensional data. We then adapt the VST to a Multi-Scale (MSVST) context to stabilize the detail coefficients of the IUWT of multi-dimensional Poisson count data. We use a hypothesis testing framework in the wavelet domain to threshold the Gaussianized and stabilized coefficients, and then apply the inverse MSVST to get the estimated intensities underlying the Poisson data. Global statistical error rate is also controlled by using a multi-test scheme, i.e. the False Discovery Rate (FDR) procedure.

## 2. VST OF FILTERED POISSON DATA

Let $(Y_n)$ be a sequence of RVs observed at the output of a finite impulse response (FIR) filter $h \in l^2(\mathbb{Z})$:

$$y_n = \sum_i h[i] x_{n-i} \tag{1}$$

where $(X_n)$ are independent Poisson RVs[1]. We additionally assume that the $h[i]$'s are non-negative samples for $i = 1, \ldots, N_h$ (typically $h$ is a low-pass filter). Our goal is to stabilize the variance of $Y_n$.

It is known that if $h = \delta$, the Anscombe transform [1] of $Y_n$ (hence $X_n$) acts as if the data arose from a Gaussian distribution with unit variance, under the assumption that the intensity $\lambda_n$ is large. This is why the Anscombe VST performs poorly in low-count settings. But, if the filter $h$ acts as an "averaging" kernel (more generally a low-pass filter), one can reasonably expect that stabilizing $Y_n$ would be more beneficial, since the signal-to-noise ratio measured at the output of $h$ is expected to be higher.

A transformation of $Y_n$ is sought, such that its variance is constant irrespective of the value of $Y_n$. The general form of the VST $\mathcal{A}$ is derived by delta-method argument, giving that:

$$\mathcal{A}Y = Z(Y) = b\sqrt{Y + c} \tag{2}$$

In the following, although the general case can be treated easily, for the sake of simplicity we shall assume that $\forall n, \lambda_n = \lambda$. For instance, in the a multi-scale wavelet transform context, this amounts to considering that the intensity is constant within the support of the wavelet. The next lemma summarizes the main properties of this transformation:

**Lemma 1** *(i) From the Taylor series expansion of the RV $Z(Y)/b$ at $Y = Y$, its mean and variance are given by:*

$$Z/b = \sqrt{\tau_1}\sqrt{\lambda} + \frac{4\, c\, \tau_1 - \tau_2}{8\, \tau_1^{3/2}} \lambda^{-1/2} + O(\lambda^{-3/2}) \tag{3}$$

$$\mathrm{Var}\,[Z/b] = \frac{\tau_2}{4\,\tau_1} + \left(\frac{7\,\tau_2^2}{32\,\tau_1^3} - \frac{c\,\tau_2}{4\tau_1^2} - \frac{\tau_3}{8\,\tau_1^2}\right)\lambda^{-1} +$$

$$\left(\frac{16c^2\,\tau_2 + 16c\,\tau_3 + 5\,\tau_4}{64\tau_1^3} - \frac{21\,c\,\tau_2^2}{32\tau_1^4}\right)\lambda^{-2} + O(\lambda^{-3}) \tag{4}$$

*where $\tau_k = \sum_i (h[i])^k$.*

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

$$b = 2\sqrt{\frac{\tau_1}{\tau_2}}, \quad c = \frac{7\tau_2}{8\tau_1} - \frac{\tau_3}{2\tau_2} \geq 0 \tag{5}$$

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

This result tells us that for the chosen value of $c$, the first order term in the expansion disappears, and the variance is almost constant up to a second order residual term. Thus, for appropriately chosen filter $h$, this residual is expected to be smaller than for the case of the Anscombe transform applied on the original samples $X_n$. The same reasoning holds for the expansion of $Z$ in (3). Let's consider the case of a 2D $B_3$-spline filter[2] (cf. (8)), which gives $c = 0.0177$ and $b = 7.3143$. With asymptotic unit variance and $\sqrt{\lambda}$ for the expectation, the coefficient before the term $O(\lambda^{-2})$ in (4) is $1.72 \times 10^{-3}$ and that before $O(\lambda^{-1/2})$ in (3) is $-4.94 \times 10^{-4}$, while these coefficients associated with the Anscombe VST are respectively $-1.56 \times 10^{-1}$ and $6.25 \times 10^{-2}$. Therefore, for this kind of filter, the convergence rate towards the asymptotic behavior of Lemma 1 is much faster (about 100 times) for the new VST than for the Anscombe VST. Clearly, the new VST will outperforms the Anscombe VST in low count situations.

This is confirmed by the simulations depicted in Fig.1, where the estimates of $Z$ (resp. $\mathrm{Var}\,[Z]$) obtained from 500 realizations of $Z$ are plotted as a function of the intensity $\lambda$ for both Anscombe (dashed-dotted), Fisz (dashed) and our VST (solid). The theoretical bounds from expansions of Lemma 1 limited to the first term (i.e. 1 for the variance and $\sqrt{\lambda}$ for the expectation) are also plotted. The saliency of our new VST is obvious. The variance of the RV stabilized using our VST is faster in sticking to the asymptotic bounds. Consequently, the RV transformed using Anscombe can be reasonably considered as stabilized for $\lambda \gtrsim 10$, using Fisz for $\lambda \gtrsim 1$ and using our VST (after low-pass filtering with $h$) for $\lambda \gtrsim 0.1$.

**Fig. 1**. $Z$ and $\mathrm{Var}\,[Z]$ for the Anscombe, Fisz and our (using the $B_3$-spline filter) VST. Notice that the RV transformed using Anscombe can be reasonably considered as stabilized for $\lambda \gtrsim 10$, using Fisz for $\lambda \gtrsim 1$ and using our VST for $\lambda \gtrsim 0.1$.

## 3. MULTI-SCALE VST AND WAVELET-DOMAIN DENOISING

Although this section focuses on the (not necessarily separable in nD, $n \geq 2$) undecimated wavelet transform (UWT), our arguments can be extended similarly to other multi-scale transforms (e.g. the curvelet transform). The UWT $\mathcal{W}$ using the filter bank $(h, g)$ of a 1D signal $x \in \ell_2(\mathbb{Z})$ leads to a set $\mathcal{W} = \{w_1, \ldots, w_J, a_J\}$ where $w_j$ are the wavelet coefficients at scale $j$ and $a_J$ are the coefficients at the coarsest resolution. The passage from one resolution to the next one is obtained using the "à trous" algorithm:

$$a_{j+1} = \bar{h}^{(j)} * a_j \quad and \quad w_{j+1} = \bar{g}^{(j)} * a_j \tag{6}$$

where $\bar{h}[k] = h[-k]$ and similarly for $g$. The reconstruction is obtained by:

$$a_j = \frac{1}{2}(\tilde{h}^{(j)} * a_{j+1} + \tilde{g}^{(j)} * w_{j+1}). \tag{7}$$

The filter bank $(h, g, \tilde{h}, \tilde{g})$ needs only verify the exact reconstruction condition. This provides us a high degree of freedom when designing the synthesis prototype filter bank.

Because astronomical images contain mostly isotropic sources (stars, galaxies, etc.), astronomers generally prefer using another transform, the Isotropic Undecimated Wavelet Transform (IUWT) [18]. Requirements for a good analysis of such data are: the filters do not need to be orthogonal or bi-orthogonal, but they must be symmetric, and more importantly, $h$, $g$, the scaling function $\phi$ and the wavelet function $\psi$ must be nearly isotropic. For example, the following filter bank satisfies these requirements:

$$
\begin{aligned}
h^{1D} &= [1, 4, 6, 4, 1]/16 \\
h[k, l] &= h^{1D}[k] \cdot h^{1D}[l], \quad g = \delta - h
\end{aligned} \tag{8}
$$

From the structure of $g$, it is easily seen that the wavelet coefficients are obtained just by taking the difference between two resolutions:

$$
w_{j+1} = a_j - a_{j+1} \tag{9}
$$

In 2D and higher dimensions, at each scale $j$, we obtain one set $\{w_j\}$ (and not three as in the 2D UWT) which has the same number of samples as the input data. The reconstruction is obtained by a simple co-addition of all wavelet scales and the final smoothed array, namely:

$$
a_0 = a_J + \sum_{j=1}^{J} w_j \tag{10}
$$

That is, the synthesis filters are $\tilde{h} = \delta$ and $\tilde{g} = \delta$. Both the analysis and synthesis filter banks implement a frame expansion.

We are now ready to introduce the MSVST-based denoising: a multiscale algorithm for denoising after stabilizing the variance of $w_j$ and bringing its distribution closer to normality, when the input signal $\mathbf{x}$ is a sequence of independent Poisson RVs. The main steps of this general MSVST denoising algorithm are as follows:

1: Let $a_0 = \mathbf{x}$. For a given filter bank $(h, g = \delta - h, \tilde{h} = \delta, \tilde{g} = \delta)$,
2: **for** $j = 0$ to $J - 1$ **do**
3:    Calculate the approximation coefficients $a_{j+1}$ using (6).
4:    Calculate

$$
w_{j+1} = \mathcal{A}_j\, a_j - \mathcal{A}_{j+1}\, a_{j+1} \tag{11}
$$

   where $\mathcal{A}_j\, a_j = \tau_1^{-1/2}\sqrt{a_j + c_j}$ is our VST, and $c_j$ is the constant in (5) obtained when the scaling function $\bar{\phi}^{(j)}$ plays the role of the low-pass filter considered in the previous section.
5:    Apply the denoising operator $\mathcal{D}$ to $w_{j+1}$, assuming that they are contaminated by an (almost) zero-mean Gaussian noise, to get the estimates $\hat{w}_{j+1} = \mathcal{D}\, w_{j+1}$.
6: **end for**

7: Reconstruct $\hat{a}_0$:

$$
\hat{a}_0 = \mathcal{A}_0^{-1}\left(\mathcal{A}_J a_J + \sum_{j=1}^{J} \hat{w}_j\right). \tag{12}
$$

### 3.1. Denoising step

This subsection is devoted to some details on the denoising step of the algorithm above. We aim at designing a hypothesis testing-based denoiser in the same vein as in [17]. To do so, one must access the distributional properties of the stabilized detail coefficients $w_j$, under the null hypothesis that the intensity $\lambda$ underlying the Poisson process is constant. Thus, Let us consider the RV $w_j$ under the null hypothesis. The following proposition gives the asymptotic expansions of $w_j$ and $\mathrm{Var}\,[w_j]$.

**Proposition 1** *We have the following asymptotic expansions:*

$$
w_j = 0 + O(\lambda^{-1/2})
$$

$$
\mathrm{Var}\,[w_j] = \frac{1}{\tau_1^{(j)}b_j^2} + \frac{1}{\tau_1^{(j+1)}b_{j+1}^2} - \frac{\zeta_{j+1}(h)}{2\tau^{(j)}{}_1\tau^{(j+1)}{}_1} + O(\lambda^{-1/2})
$$

*where $\zeta_{j+1}(h) = \sum_{k,l} H^{j+1}(k, l)H^j(k, l)$, and:*

$$
H^{j+1}(k, l) = \left(H^j * \bar{h}^{(j)}(k)\bar{h}^{(j)}(l)\right)(k, l),\ H^0 = \delta \tag{13}
$$

where $b_j$ (resp. $b_{j+1}$) is defined as in (5) associated to $\bar{\phi}^{(j)}$ (resp. $\bar{\phi}^{(j+1)}$) and similar for $\tau_1^{(j)}$ and $\tau_1^{(j+1)}$. The results above clearly prove that, to a good approximation, the Gaussianized version of the detail coefficients have a zero mean and a variance that depends only on the scale and the chosen filter bank.

Thus, exploiting these results, the hypothesis testing-based denoiser only requires pre-calculating these variances before applying the MSVST, which is computationally simple and fast. Moreover, as the wavelet coefficients are tested simultaneously (multiple testing) and are dependent, we used the FDR strategy under dependency [19] to control the global statistical error.

### 4. RESULTS AND DISCUSSION

For illustrative purposes, we have simulated an image with circle-like X-ray sources on a constant background for XMM-Newton telescope. This image can be seen as a model for celestial objects of different size and flux.[3] Each source along any radial branch has the same flux and has a more and more extended support as we go farther from the center. The flux reduces as the branches turn in the clockwise direction. Denoising such an image is highly challenging. The observed image of counts is Fig. 2(a), the restored intensity using our approach in (d) and those restored using some of our competitors Anscombe in (b) and Fisz in (c). For all the methods, the FDR level was fixed at 0.2. Cycle-spinning was used for the

---

[3]Defined as the integral of the source intensity over its support.

Fisz transform in order to not bias the comparison of estimators in favor of our approach.

As revealed by this figure, all estimators perform comparatively well at high intensity levels (image center). However, compared with (b), the relative merits (sensitivity) of the MSVST estimator become increasingly salient as we go farther from the center, and as the branches turn clockwise, i.e. as the intensity becomes low. Even the sources of very low counts were detected by our estimator (see the last branches clockwise in (d) and compare to (b)). Fisz method also proves relatively sensitive. However, it still exhibits a clear staircasing artifact despite the cycle-spinning. Cycle-spinning has the drawback of making the Fisz approach much slower than our method (typically $10 - 50$ shifts are used in cycle-spinning for each axis of an $512 \times 512$ image, hence $100 - 2500$ times slower).

**Fig. 2**. (a) Noisy; (b) Anscombe VST; (c) Fisz approach [3]; (d) Our method.

## 5. CONCLUSION

In this paper, a multi-scale VST is proposed for intensity estimation of a sequence of independent Poisson RVs. A general wavelet-domain denoiser was also described and has proven very efficient in estimating Poisson noise contaminated data. The algorithm can be easily extended to any dimension and can also be generalized to other multi-scale transforms such as those integrating directionality, e.g. curvelet transform. Another potential application of the MSVST method is the denoising of multi-spectral data, e.g. 2D + energy. Our current research is focusing on these aspects.

## 6. REFERENCES

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

[2] 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.

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

[4] 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.

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

[6] P. Besbeas, I. De 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.

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

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

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

[10] 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.

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

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

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

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

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

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

[17] B. Zhang, M. J. Fadili, and J-.L. Starck, "Poisson intensity estimation based on the wavelet domain hypothesis testing," Tech. Rep., CEA-Saclay, 2005.

[18] J.-L. Starck and F. Murtagh, *Astronomical Image and Data Analysis*, Springer-Verlag, 2002.

[19] 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.