Multi-Scale Variance Stabilizing Transform for Multi-Dimensional Image Restoration with Poisson Noise

Jalal M. FADILI[a], Bo ZHANG[b] and Jean-Luc STARCK[c]

[a] GREYC CNRS UMR 6072, Caen France

[b] Institut Pasteur, Paris France

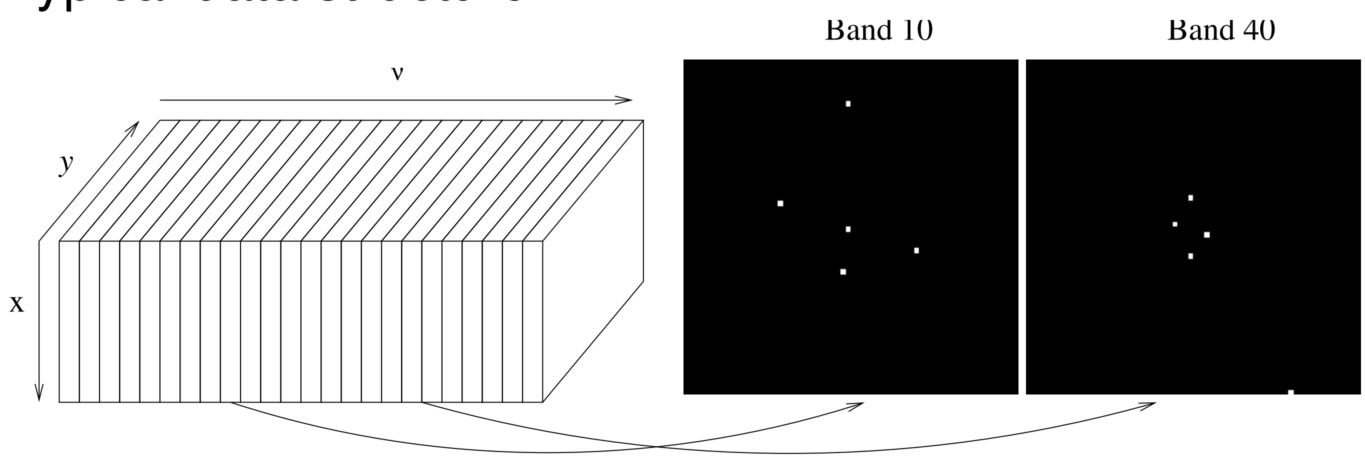
[c] CEA-Saclay, Gif-sur-Yvette France

Outline

- Introduction.
  - Context.
  - Observation model.
  - Goals.
- Poisson image intensity estimation: an overview.
- Variance Stabilizing Transform (VST).
- Asymptotic analysis.
- Application to multi-scale transforms.
- Experimental results.
- Conclusion and extensions.

Introduction (Context)

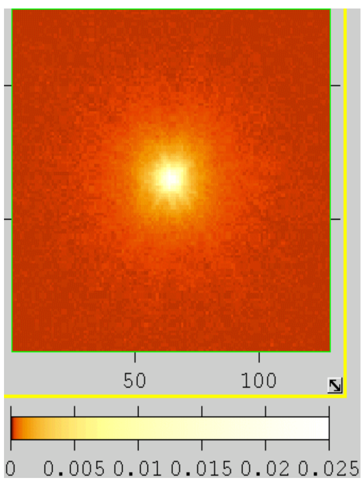
- GLAST (The Gamma Ray Large Area Space Telescope, 6 countries, Stanford Univ.).
- For each photon, we have its position and energy.
- Typical data structure:



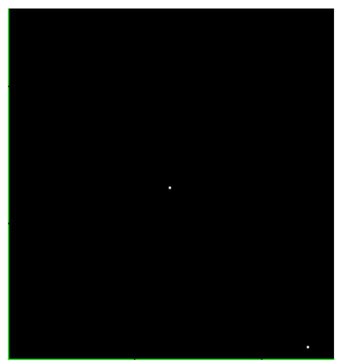
- Photon counts $v(x, y, \nu)$ at (x, y) spatial coordinates and energy bin ν.
- First difficulty: very low count setting (typically 0-1 photons /pixel/energy bin).
- Widely used method: average along energy and then denoise $\Rightarrow$ loss of energy information.

Introduction (Observation model)

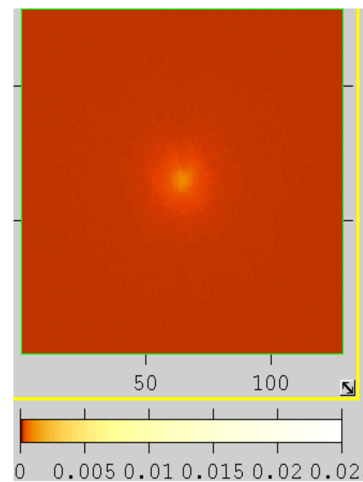
- Observe a discrete image of counts $\mathbf{v} = (v_i)_{i \in I}$ (I the index set).
- The v_i's are iid Poisson distributed $\mathbf{v} \sim \mathcal{P}(\Lambda)$, $\Lambda = (\lambda_i)_{i \in I}$ is the underlying intensity image.
- Examples



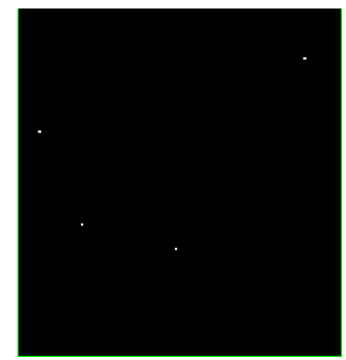
Intensity



Counts



Intensity



Counts

Introduction (Goals)

- Get an estimate of the intensity $\hat{\Lambda} = \mathcal{D}\mathbf{v}$ using statistical estimation/decision theory in a sparse-representation domain (e.g. wavelets, curvelets).
- Must be efficient enough in very low-count setting.
- Adaptive Shape preservation (e.g. isotropic structures, lines, etc).
- Photometry preservation.

Wavelet Poisson intensity estimation: Review

- Transformation-based methods
  - Variance stabilizing transforms (e.g. Anscombe) [Donoho 93].
  - Gaussianizing transforms (Fisz) [Frylewizc 04, Fadili 04].
- Modulation estimators [Antoniadis et al. 01].
- Classical direct methods
  - Wavelet histogram auto-convolution [Bijaoui et al. 01].
  - l_1 penalized estimator [Sardy et al. 03].
  - Adaptive Wiener filter [Nowak et al. 99].
- Bayesian direct methods
  - Bayesian mutliscale models [Kolaczyk 99,04].
  - Mixture-based prior [Timmerman 99, Lu 04].
- Hypothesis testing methods [Kolaczyk 99, Zhang et al. 04].

VST of filtered Poisson data

- Let $\{Y\}_n$ a sequence of RVs observed at the output of a FIR filter $h \in l^2(\mathbb{Z})$:

$$y_n = \sum_i h_i x_{n-i}, \ \{X\}_n \sim \mathcal{P}(\lambda) \qquad (1)$$

- If the filter h_i acts as an "averaging" kernel (more generally a low-pass filter), it is expected that stabilizing Y_n would enhance the output SNR.
- Consequence: efficiency in low count settings.
- The shape to be preserved will be conditionned by the choice of h (matched filter).

Delta method

Lemma 1

- *Let h_i be positive valued and $\sum_i h_i = 1$. If G is $C^m, m \geq 3$ and $\|G^{(m)}(\lambda)P(\lambda)\|_\infty \leq M < \infty$, then:*

$$\mathrm{E}\left[G(Y_n)\right] = G(\lambda) + \sum_{k=2}^{m-1} \frac{G^{(k)}(\lambda)}{k!}\mathrm{E}\left[Y_n - \lambda\right]^k + R_m$$

$$\exists K_m > 0 \quad \mathrm{tq} \quad |R_m| < K_m n_h^{-m/2}\frac{M}{m!} \tag{2}$$

$P(\lambda)$ is a polynomial with positive coefficients whose degree is at least 1 and at most $\lfloor m/2 \rfloor$.

- *If $\|G^{(m)}\|_\infty < \infty, 1 \leq m \leq 3$:*

$$\mathrm{Var}\left[G(Y_n)\right] = \lambda\|h\|_2^2\left[G'(\lambda)\right]^2 + O\left(n_h^{-3/2}\right) \tag{3}$$

Delta method

Lemma 1

- *Let h_i be positive valued and $\sum_i h_i = 1$. If G is $C^m, m \geq 3$ and $\|G^{(m)}(\lambda)P(\lambda)\|_\infty \leq M < \infty$, then:*

$$\begin{aligned} \mathrm{E}\left[G(Y_n)\right] &= G(\lambda) + \sum_{k=2}^{m-1} \frac{G^{(k)}(\lambda)}{k!} \mathrm{E}\left[Y_n - \lambda\right]^k + R_m \\ \exists K_m > 0 \quad \text{tq} \quad |R_m| &< K_m n_h^{-m/2} \frac{M}{m!} \end{aligned} \quad (3)$$

  $P(\lambda)$ is a polynomial with positive coefficients whose degree is at least 1 and at most $\lfloor m/2 \rfloor$.

- *If $\|G^{(m)}\|_\infty < \infty, 1 \leq m \leq 3$:*

$$\mathrm{Var}\left[G(Y_n)\right] = \lambda \|h\|_2^2 \left[G'(\lambda)\right]^2 + O\left(n_h^{-3/2}\right) \quad (2)$$

Delta method (cnt'd)

- A Variance Stabilizing Transform (VST) amounts to canceling out this term by solving the ODE:

$$\lambda \left[G'(\lambda)\right]^2 = cst$$

- The general form of the VST $\mathcal{A}$ is arrived at by delta-method argument, giving that:

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

- b is a normalizing factor and c is a constant that controls the convergence speed towards the asymptotic behaviour.

Asymptotic results

Theorem 1

(i) The asymptotic expansion of the mean and variance of the RV Z/b are given by:

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

$$\mathrm{E}\,[Z/b] = \sqrt{\lambda} + \frac{4\,c\ - \tau_2}{8}\left(\frac{1}{\lambda}\right)^{1/2} + O\left[\frac{1}{\lambda}\right]^{3/2} \qquad (4)$$

Asymptotic results

Theorem 1

(i) The asymptotic expansion of the mean and variance of the RV Z/b are given by:

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

$$\mathrm{E}\left[Z/b\right] = \sqrt{\lambda} + \frac{4\,c\; - \tau_2}{8}\left(\frac{1}{\lambda}\right)^{1/2} + O\left[\frac{1}{\lambda}\right]^{3/2} \quad (4)$$

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

$$c = \tfrac{7\tau_2}{8} - \tfrac{\tau_3}{2\tau_2} > 0 \qquad b = \tfrac{2}{\sqrt{\tau_2}} \quad (4)$$

where $\tau_k = \sum_i (h_i)^k$.

Asymptotic results

Theorem 1

(i) The asymptotic expansion of the mean and variance of the RV Z/b are given by:

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

$$\mathrm{E}\left[Z/b\right] = \sqrt{\lambda} + \frac{4\,c\; - \tau_2}{8}\left(\frac{1}{\lambda}\right)^{1/2} + O\left[\frac{1}{\lambda}\right]^{3/2} \quad (3)$$

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

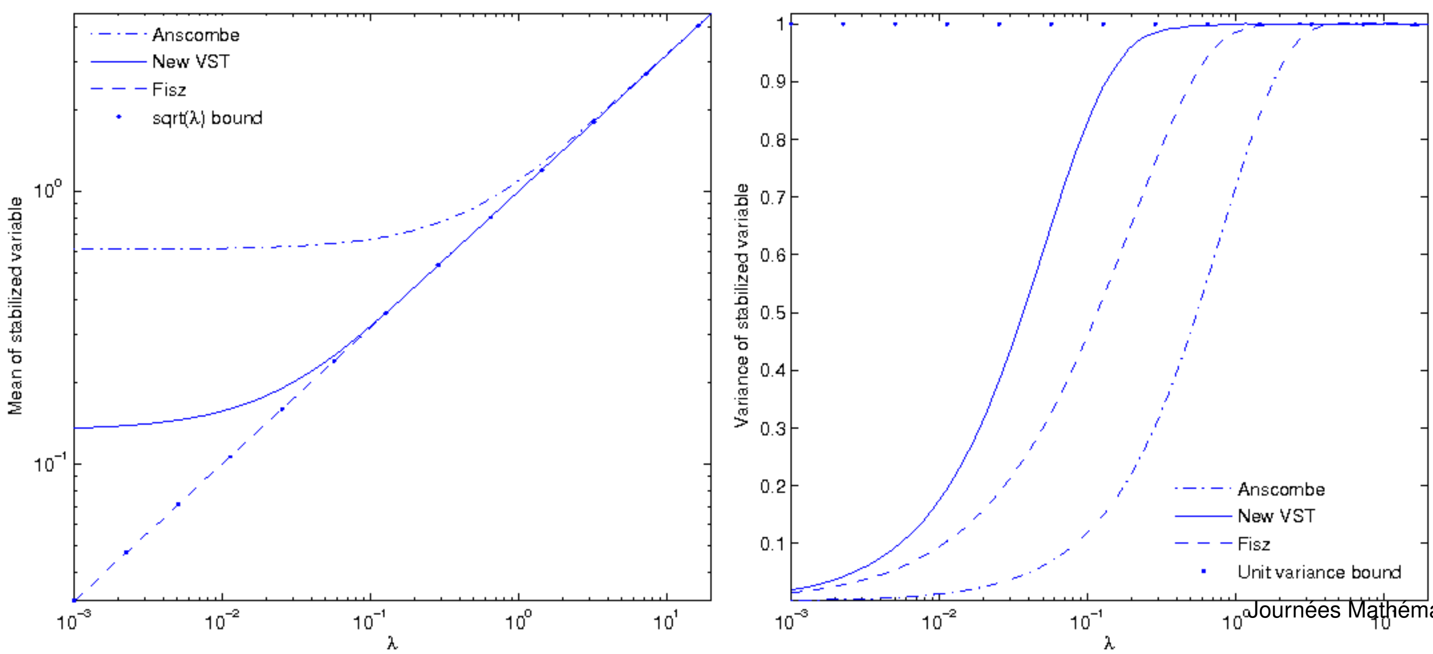
$$c = \tfrac{7\tau_2}{8} - \tfrac{\tau_3}{2\tau_2} > 0 \qquad b = \tfrac{2}{\sqrt{\tau_2}} \quad (3)$$

where $\tau_k = \sum_i (h_i)^k$.

(iii) For b and c as above, $\left(Z - 2\sqrt{\frac{\lambda}{\tau_2}}\right) \xrightarrow[\lambda \to +\infty]{d} \mathcal{N}\left(0, 1\right)$.

Asymptotic results: comments

- The variance is almost constant up to a second order residual term.
- For appropriately chosen h, the convergence rate toward the asymptotic behavior is faster for the new VST than for the Anscombe transform.
- The same reasoning holds for the expansion of $\mathrm{E}[Z]$.
- Example: B_3-spline filter, $c = 0.0177 \ll 3/8$ and $b = 7.3143$.
  - The mean and variance of the RV stabilized using our VST are faster to stick to the asymptotic bounds.
  - The RV stabilized using Anscombe can be reasonably considered as Gaussianized for $\lambda \gtrapprox 4$, using Fisz for $\lambda \gtrapprox 1$ and using our VST for $\lambda \gtrapprox 0.1$.
  - Clearly, the new VST will be better that Anscombe and Fisz in low count situations.



Application to wavelets: Multi-Scale VST

- The UWT $\mathcal{W}$ using the filter bank (h, g) of a 1D signal $\mathbf{x}$ leads to a set $\{w_1, \ldots, w_J, a_J\}$.
- The passage from one resolution to the next one is obtained using the "à trous" algorithm:

$$a_{j+1,l} = \bar{h}^j * a_j \quad and \quad w_{j+1,l} = \bar{g}^j * a_j \tag{3}$$

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

$$a_j = \frac{1}{2}(\tilde{h}^j * a_{j+1} + \tilde{g}^j * w_{j+1}). \tag{4}$$

- Here, the filter bank $(h, g, \tilde{h}, \tilde{g})$ needs only verify the exact reconstruction condition.

Coupled detection and estimation algorithm

- Let's consider the case of filter banks where $(h, g = \delta - h, \tilde{h} = \delta, \tilde{g} = \delta)$.
- Such filter banks are encountered in many application areas (astronomy, biomedical imaging, etc).
- The coupled MSVST denoising algorithm is as follows:

1: Let $a_0 = \mathbf{x}$. For a given filter h,
2: **for** $j = 0$ to $J - 1$
3: Calculate the approximation coefficients a_{j+1}.
4: $w_{j+1} = \mathcal{A}_j\ a_j - \mathcal{A}_{j+1}\ a_{j+1}$,
$\mathcal{A}_j$ is our VST operator, with a constant c_j associated with the scaling function $\bar{\phi}^{(j)}$.
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 the estimate:

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

- Application to any d-dimensional data is straighforward.

Denoising (coupled scheme)

Theorem 2 *Consider the above filter bank.*

- $W_j = \mathcal{A}_{j-1}\, A_{j-1} - \mathcal{A}_j\, A_j \xrightarrow[\lambda \to +\infty]{d} \mathcal{N}\left(0, \sigma_j^2\right).$
- $\sigma_j^2 = \frac{1}{b_{j-1}^2} + \frac{1}{b_j^2} - \frac{\zeta_j(h)}{2}$ *(independent of λ). Where*
  $\zeta_j(h) = \sum_i \bar{\phi}_i^{(j)} \bar{\phi}_i^{(j-1)}, \bar{\phi}_i^{(0)} = \delta_i.$
-

$$\frac{1}{\sqrt{\tau_2^{(j)}}}\left(\mathcal{A}_j\, A_j - \sqrt{\lambda}\right) \xrightarrow[j \to +\infty]{d} \mathcal{N}\left(0, \frac{1}{4}\right)$$

$$\frac{W_j}{\sqrt{\tau_2^{(j)}}} \xrightarrow[j \to +\infty]{p.s.} 0.$$

$$\tau_2^{(j)} = \|\bar{\phi}^{(j)}\|_2^2.$$

Use your favorite denoising algorithm on the stabilised data.

Separate detection and estimation algorithm

- Arbitrary high-pass g.
- But no closed-form reconstruction operator $\Rightarrow$ iterative reconstruction (HSD).

1: Let $a_0 = \mathbf{x}$. For a given filter h,
2: **for** $j = 0$ to $J - 1$
3: Calculate the approximation coefficients a_j.
4: $w_{j+1} = \bar{g}^j * \mathcal{A}_j\ a_j$,
$\mathcal{A}_j$ is our VST operator, with a constant c_j associated with the scaling function $\bar{\phi}^{(j)}$.
5: Denoise w_{j+1}, assuming a WGN and get the estimates $\hat{w}_{j+1} = \mathcal{D}\ w_{j+1}$.
6: **end for**
7: Reconstruct the estimate using an iterative HSD scheme [Yamada et al. 00, Starck 04].

- Application to any d-dimensional data is straightforward.

Denoising (separate scheme)

Theorem 3 *Consider the above filter bank with an arbitrary high-pass g.*

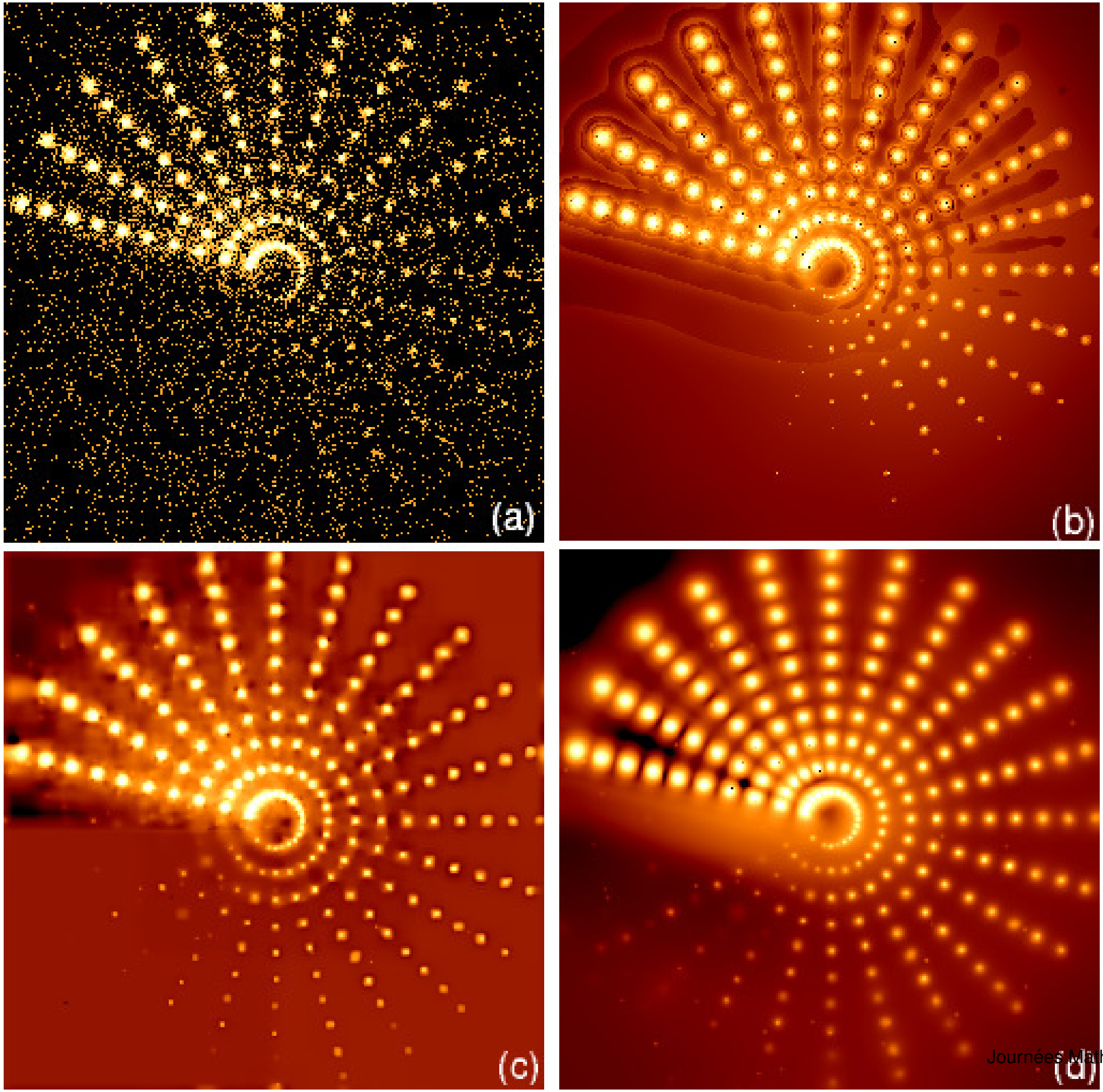
$$W_j = \bar{g}^j * \mathcal{A}_{j-1}\, A_{j-1} \xrightarrow[\lambda \to +\infty]{d} \mathcal{N}\left(0, \sigma_j^2\right)$$

where $\sigma_j^2 = \frac{1}{\tau_2^{(j-1)}} \sum_{m,n} \bar{g}_m^j \bar{g}_n^j \sum_k \bar{\phi}_k^{(j-1)} \bar{\phi}_{n-m+k}^{(j-1)}$ *(independent of λ).*

Again, use your favorite denoising algorithm on the stabilised data.

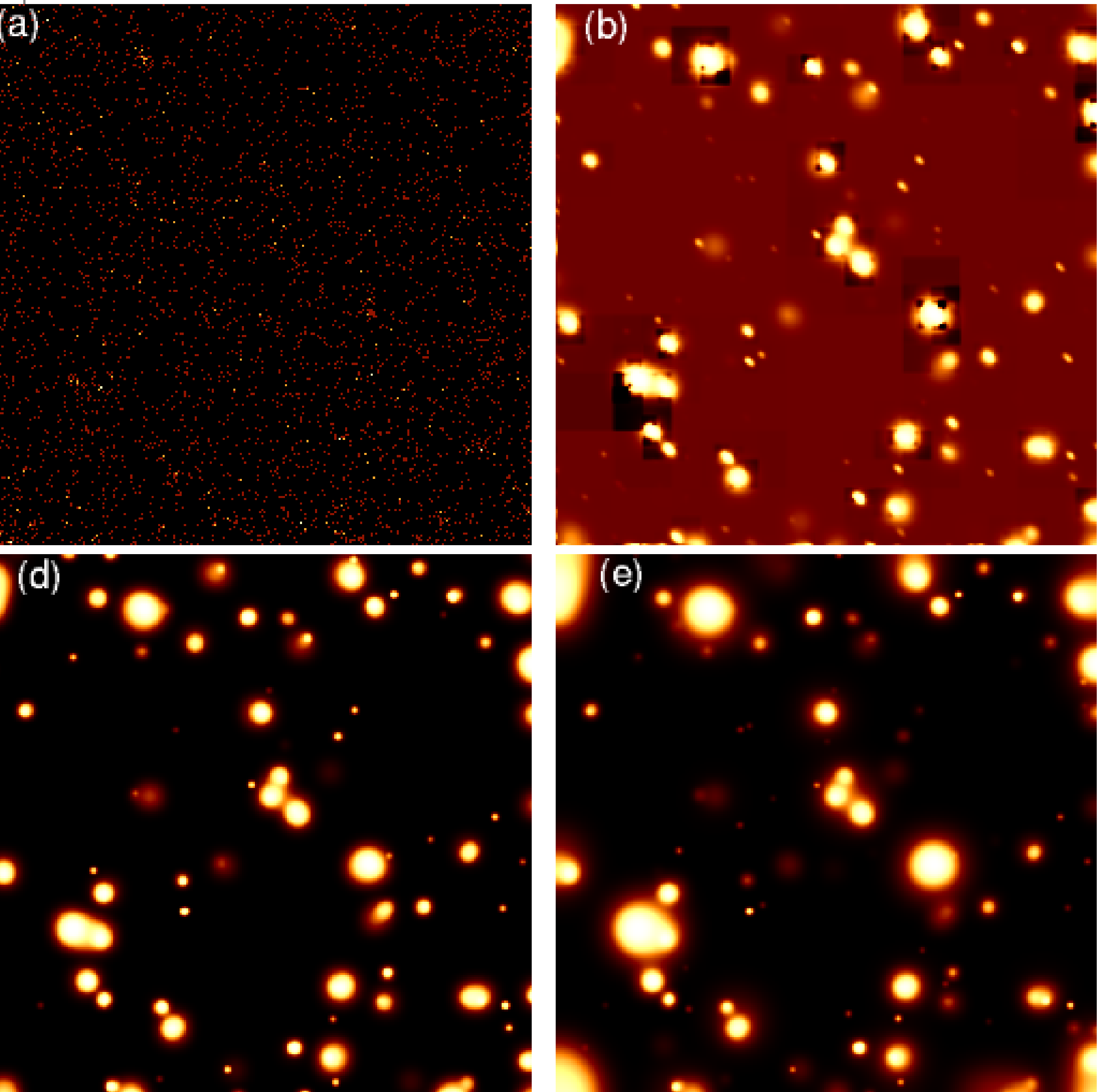
2D example (1)

(a) Original, (b) Anscombe, (c) Fisz, (d) MSVST (coupled).

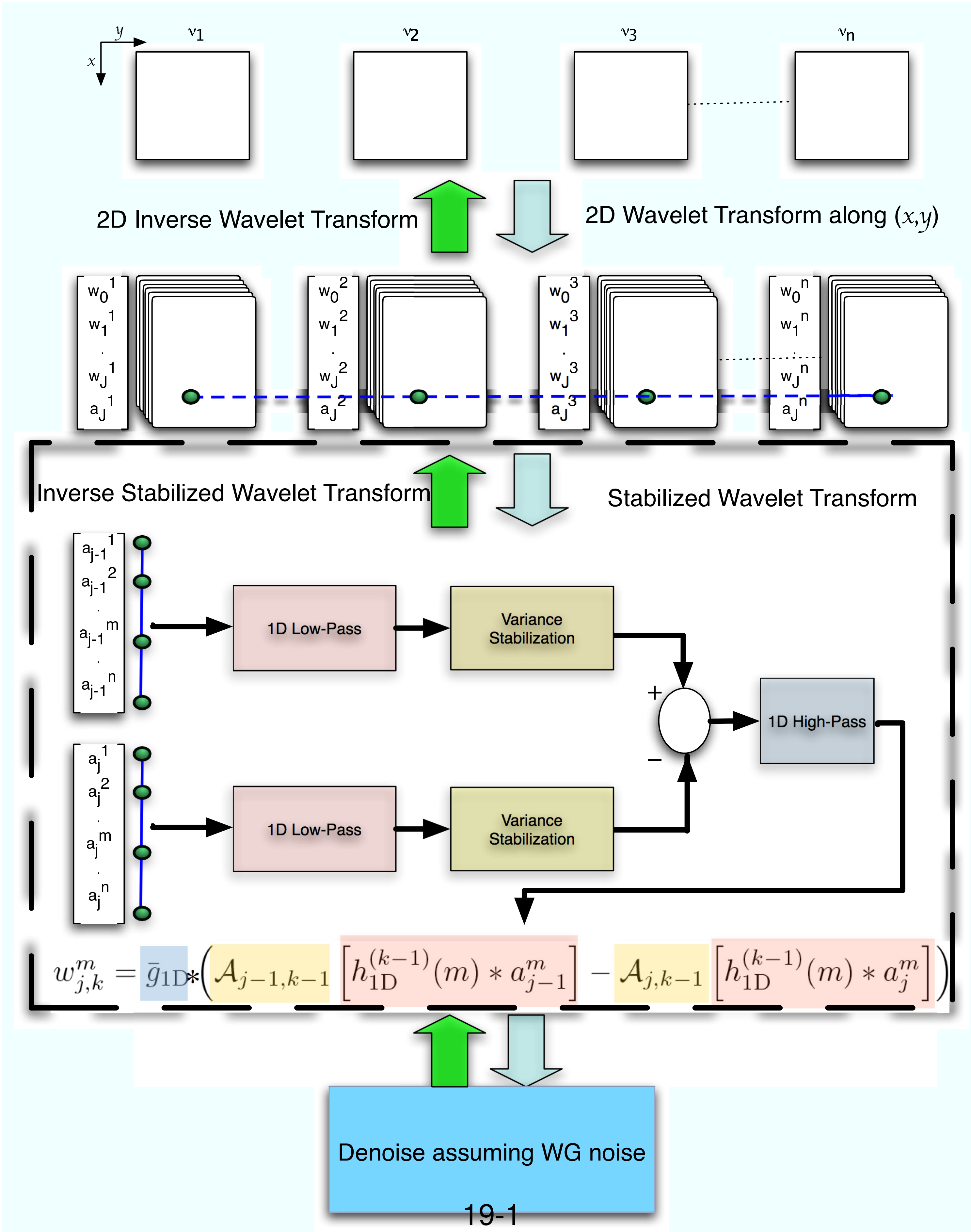


2D example (2)

(a) Original, (b) Fisz, (c) Anscombe, (d) Auto-Convolution (exact), (e) MSVST (coupled).



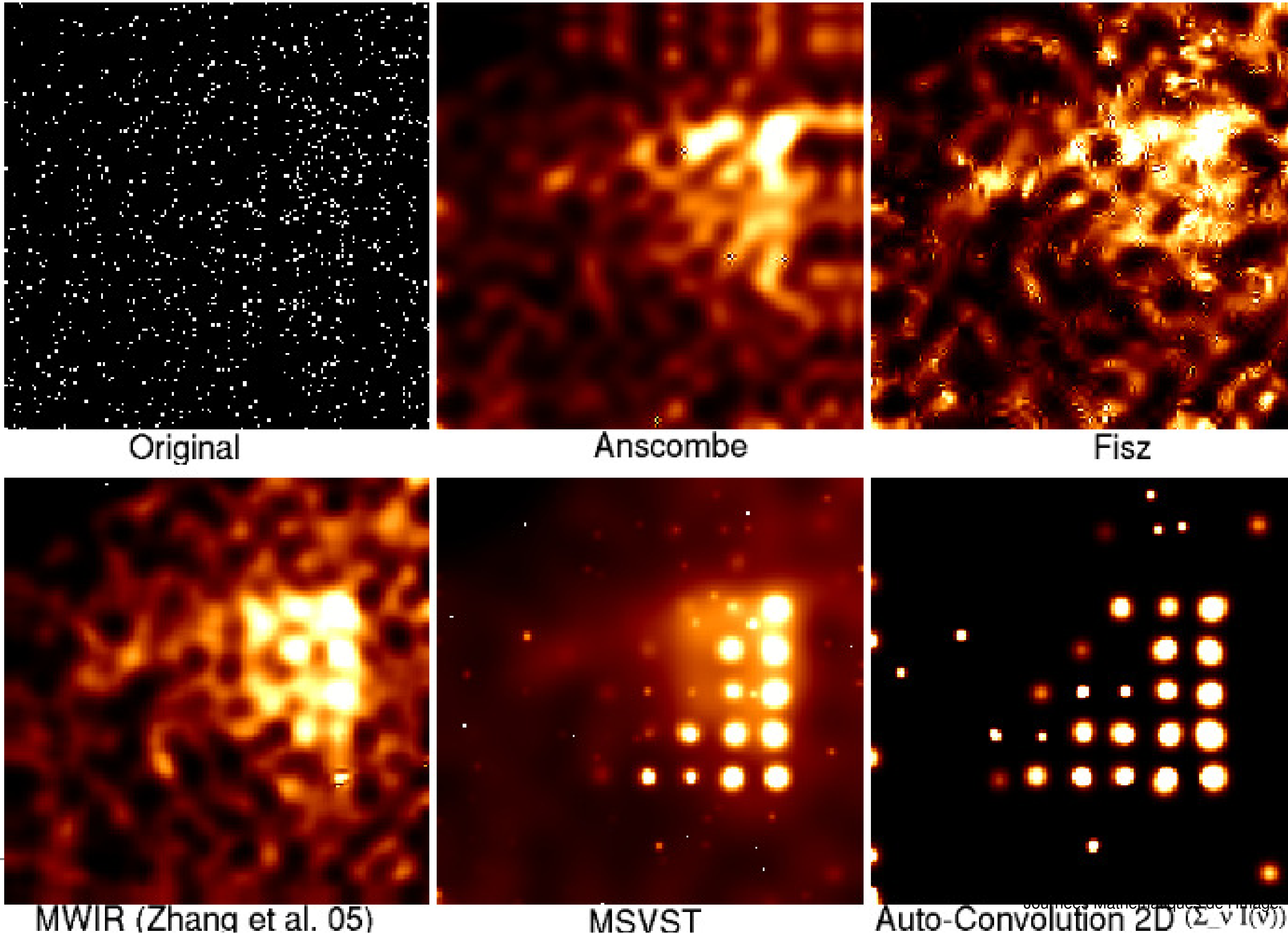
Extension to multispectral data



y
x
ν1
ν2
ν3
νn
2D Inverse Wavelet Transform
2D Wavelet Transform along (x,y)
w_0^1 w_1^1 . w_J^1 a_J^1
w_0^2 w_1^2 . w_J^2 a_J^2
w_0^3 w_1^3 . w_J^3 a_J^3
w_0^n w_1^n . w_J^n a_J^n
Inverse Stabilized Wavelet Transform
Stabilized Wavelet Transform
a_{j-1}^1 a_{j-1}^2 . a_{j-1}^m . a_{j-1}^n
a_j^1 a_j^2 . a_j^m . a_j^n
1D Low-Pass
Variance Stabilization
1D Low-Pass
Variance Stabilization
+
−
1D High-Pass
$w_{j,k}^m = \bar{g}_{1D} * \left(\mathcal{A}_{j-1,k-1} \left[h_{1D}^{(k-1)}(m) * a_{j-1}^m \right] - \mathcal{A}_{j,k-1} \left[h_{1D}^{(k-1)}(m) * a_j^m \right] \right)$
Denoise assuming WG noise

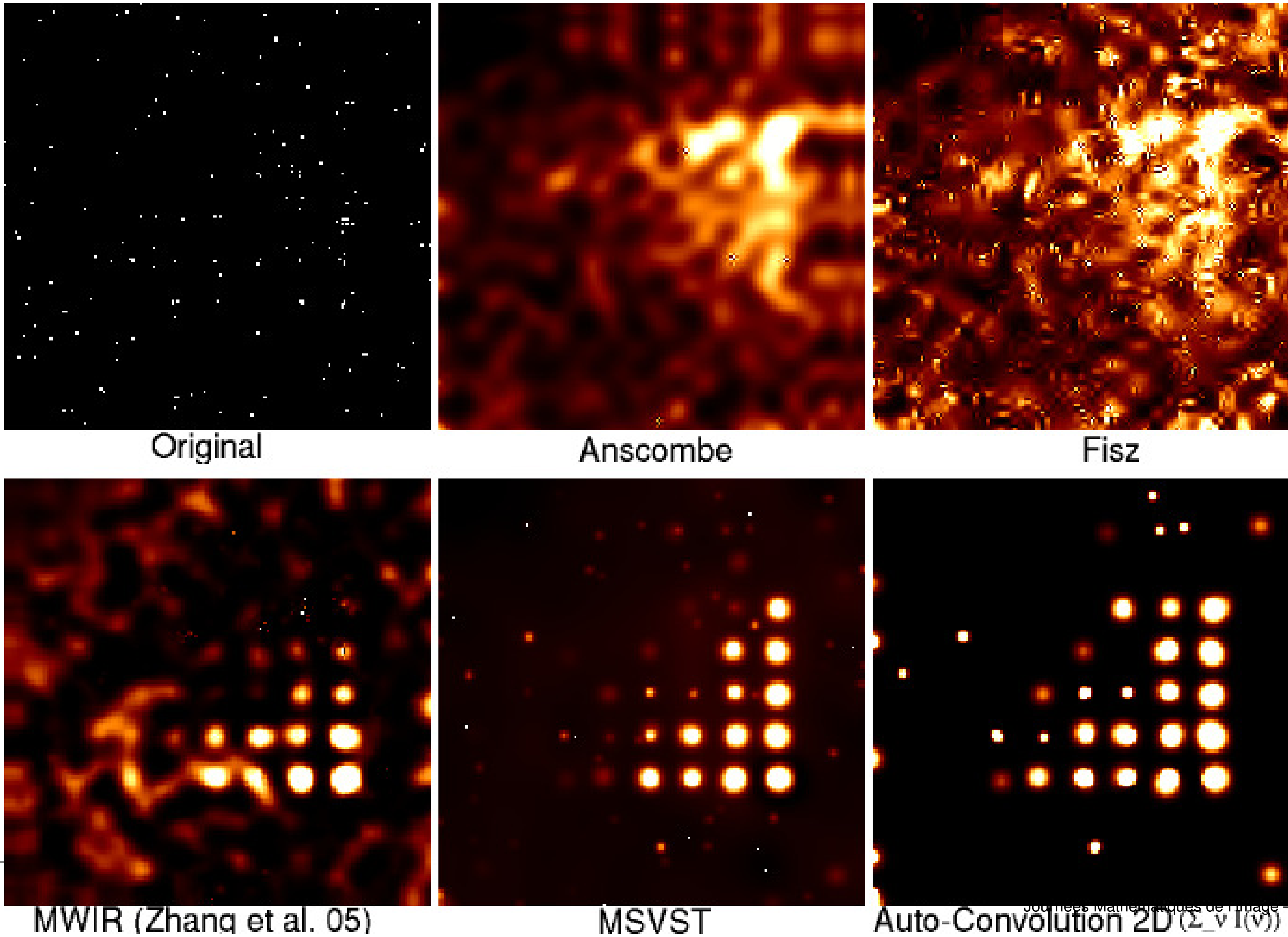
Multispectral example

GLAST data, 161×161 images and 31 (energy bins): Frame 8



Multispectral example

GLAST data, 161×161 images and 31 (energy bins): Frame 16



Multispectral example (GLAST)
Animated cube.

Let’s summarize: Take away messages

- Both theoretical and experimental evidence show that MSVST is better than Ascombe and Fisz in low intensity settings.
- Results are very close to the wavelet histogram auto-convolution method (an exact method).
- MSVST is easy and fast to implement in any dimension (which is not the case for the exact method).
- Flexible enough to use your favorite denoising operator.
- Data (shape) adaptive with the proper filter bank.

Conclusion and Perspectives

- An original MSVST was proposed for multi-dimensional data restoration with Poisson noise.
- Many difficulties were solved (specificity vs sensitivity, regularity, photometry).
- The methods extends easily to multispectral data.
- More validation and comparison.
- Our current work is focusing on extending the method to other transforms such as curvelets.

More details

http://www.greyc.ensicaen.fr/~jfadili

http://jstarck.free.fr

Any questions ?