Draft version January 31, 2023
Typeset using LaTeX **twocolumn** style in AASTeX631


# *Jolideco* : Joint Likelihood Deconvolution of Astronomical Images in the Presence of Poisson Noise

Axel Donath, Aneta Siemiginowska, Vinay Kashyap, and David van Dyk

## ABSTRACT

We present a new method for (Jo)int (li)kelihood (Deco)nvolution (*Jolideco* ) of a set of astronomical observations of the same sky region in the presence of Poisson noise. The method reconstructs an image from a set of observations by optimizing the a-posteriori joint Poisson likelihood of all observations under an patch based image prior. The patch prior is parameterised by a standard Gaussian Mixture model (GMM) learned from astronomical images at other wavelenghts and adapted to the structures expected to see in the image. By applying the method to simulated data we show that the combination of mutiple observations leads to an improved reconstruction quality in the regime with S/N ratio of. We show also that the method yields superior reconstruction quality to alternative standard methods such as the Richardson-Lucy method.

## 1. INTRODUCTION

The quality of all astronomical images is affected by the limited angular resolution of the instrument or telescope. In addition the images are affected by the presence of noise and non-uniform exposure.

In literature there have been multiple efforts to

Any astronomical observation is affected by the limited angular resolution of the telescope.

- Richardson & Lucy (RL)Richardson (1972) Lucy (1974)
- Joint likelihood RL: Ingaramo et al. (2014) has found that RL take the best out of each image and create a "merged" image
- Problem of RL decomposing structures into point sources: Perry & Reeves (1994) Fish et al. (1995)
- Multi-scale LIRA prior of course: Esch et al. (2004); Connors et al. (2011), advantage of error estimates / Bayesian sampling. However prior is rather "ad-hoc", no physical information / assumptions. Just "smoothness"
- What abput d3po and d4po? Priors based on physical assumptions, decomposition into point and diffuse flux, but in 3d/4d Selig, Marco & Enßlin, Torsten A. (2015)Pumpe, Daniel et al. (2018)

### 1.1. *A note on Deep Learning Methods*

Xu et al. (2014) have introduced a baseline architecture for a deep convolutional neural network for image deblurring. The convolutional nature of the network does not make it suitable for the task. As convolution introduces smoothing. However the network can learn an "inverse kernel", which needs to be larger because of the uncertainty principle (TODO: reference...). The large kernels are introduced by separability assumption... In the limit of Poisson noise it is desirable to trace the full a-postiori likelihood, which is not possibly for most network architectures.

## 2. METHOD

### 2.1. *Poisson Joint Likelihood*

Our goal is to recover an image $\mathbf{x}$ from a set of multiple low counts observations. Most generally we assume our total dataset consists of $M$ individual observations of the same region of the sky, which are jointly modelled. The assumption is that the underlying *true* emission image, we are looking for does not change with time. Under this assumption these datasets can be for example:

- Different observations of one instrument or telescope at different times and observation conditions. E.g., multiple observations of *Chandra* with different offset angles and exposure times.
- Observations of different telescopes, which operate in the same wavelength range. E.g., a *Chandra* and *XMM* observation of the same region in the sky.
- A single observation of one telescope with different data quality categories and different associated instrument response functions. E.g., event classes for *Fermi*-LAT .

Or a even an arbitrary combination of the possibilities listed above.

For each individual observation $m$ the predicted counts can be modelled by forward folding the unknown flux image $\mathbf{x}$ with the individual (per observation) instrument response:

$$\lambda_m = \mathrm{PSF}_m \circledast (\mathbf{E}_m \cdot (\mathbf{x} + \mathbf{B}_m)) \tag{1}$$

Where the expected counts $\lambda_i$ are given by the convolution of the true underlying flux distribution $\mathbf{x}$ with the PSF. Additionally the observation specific exposure $\mathbf{E}$ and background emission $\mathbf{B}$ can be taken into account.

Given a single observation $m$ of a true underlying flux distribution $\mathbf{x}$ and assuming the noise in each pixel $i$ in the recorded counts image $\mathbf{d}$ follows a Poisson distribution, the total likelihood $\mathcal{L}_m$ of obtaining the measured image from a model image of the expected counts $\lambda_i$ with $N$ pixels is given by:

$$\mathcal{L}_m\left(\mathbf{d}|\lambda\right) = \prod_i^N \frac{e^{-d_i}\lambda_i^{d_i}}{d_i!} \tag{2}$$

By taking the logarithm and dropping the constant terms one can transform the product into a sum over pixels, which is also often called the *Cash* (Cash 1979) fit statistics:

$$\mathcal{C}\left(\mathbf{d}|\lambda\right) = \sum_i^N \lambda_i - d_i \log \lambda_i \tag{3}$$

Bayes rule:

$$P(\theta|\mathbf{D}) = P(\theta)\frac{\mathcal{C}(\mathbf{D}|\theta)}{P(\mathbf{D})} \tag{4}$$

The total objective function $\mathcal{L}$ is given by:

$$\mathcal{L}\left(\mathbf{d_m}|\mathbf{x}\right) = \sum_m^M \mathcal{C}\left(\mathbf{d_m}|\mathbf{x}\right) - \beta \cdot \mathcal{P}(\mathbf{x}) \tag{5}$$

Where $\mathcal{C}\left(\mathbf{d_m}|\mathbf{x}\right)$ represents the summed log-likelihood for an individual observation $m$. Additionally the function includes a prior term $\mathcal{P}(\mathbf{x})$ and a factor $\beta$ to adjust the weight of the prior with respect to the likelihood term. TODO: To be clarified: the weight does not really make it a prior, more a regularisation term...

### 2.2. *Priors*

Just include patch priors, or LIRA as well, or uniform?

#### 2.2.1. *Patch-Prior*

To improve the reconstruction quality of images for inverse problems, such as denoising, inpainting or deconvolution it is useful to have prior assumptions on the statistics of the image to guide the optimization based image reconstruction process. However images are high dimensional and thus difficult to capture the global image statistics taking into account all pixel to pixel correlations.

By recognising that on smal scale images often contain basic structure such as edges, corners, periodic patterns or region of constant brightness. The idea of patch priors was first introduced by Zoran & Weiss (2011). Their main idea was to learn the statistics of natural images on a small spatial scales instead. For this they proposed to split images from a representative dataset into small patches of size 8x8 pixels.

Then they learned a 64 dimensional Gaussian Mixture Model (GMM) on the extracted patches, treating each pixel as an independent dimension in the model, with $k = 200$ components. They showed that the GMM prior led to much improved image reconstructions, compared to comparable approaches (which ones?).

One of the main advantages of the GMM is the possibility to evaluate its log-likelihood in closed form. It is given by:

$$\ell_{GMM}(\theta) = \sum_{i=1}^{n} \log\left(\sum_{k=1}^{K} \pi_k N(x_i; \mu_k, \sigma_k^2)\right) \tag{6}$$

With parameters $\theta = \{\mu_1, \ldots, \mu_K, \sigma_1, \ldots, \sigma_K, \pi_1, \ldots, \pi_K\}$. To use the learned GMM as prior for the reconstruction of another image, the image is split into overlapping patches. For each of the patches the log-likelihood for each of the GMM components is evaluated. The resulting grid of overlapping patches is illustrated in Figure 1.

For each 8x8 pixel patch the likelihood of all components of the GMM is evaluated. Then

$$\hat{k} = \arg\max \ell_{GMM}(\theta) \tag{7}$$

$$\mathcal{P}(x) = \sum_n \log p_{\hat{k}}(\mathbf{P}_n x) \tag{8}$$

$\mathbf{P}_n$ is a matrix that extracts the $n$-th patch from the image $\mathbf{x}$ to be reconstructed. $p(\mathbf{P}_n x)$ is the probability of that patch under the GMM.

Bouman et al. (2016) later adapted the patch prior reconstruction to be used with radio astronomy data. They have shown that the reconstructed image only weakly depends on the choice of the reference data on which the the patch prior is learned. They found equivalent results for GMMs learned on natural images and specifically simulated images of black hole ring structures.

#### 2.2.2. *Image Normalisation*

Astromomical images typically show a much higher dynamic range compared to natural images. This is

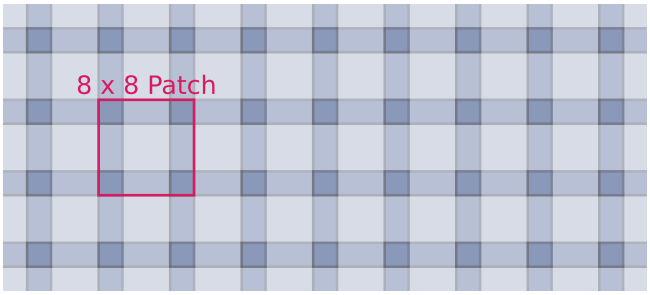


**Figure 1.** Grid of overlapping patches of size 8x8 pixels. The overlap size was chosen to be 2 pixels.

mostly due to the existence of point sources, which rarely occur in natural images, but a very common in astronomical images because of objects at hig distances. The GMM patch prior is learned on normalized images, where the intensity values are constrained between 0 and 1. To evaluate For the likelihood term the image intensity need to be conserved.

The choice of the image normalisation allows to adjust the contrast of the image and e.g. enhance low intensity structures, which also enhances the regularising effect of the patch prior.

TODO: Which normalisation to choose, atan, asinh, log, inverse-cdf?

#### 2.2.3. *Cycle Spinning*

To avoid artifact due to the choice of the grid of overlapping patches we propose three variations:

- Cycle spining by randomly shifting the image by a given number of pixels in x and y direction
- Sub pixel cycle spinning, by randomly distributing the bightness of a given pixel to a 4 pixel grid
- A randomly chosen patch grid, such that each pixel is at least covered once. The random grid of patches was proposed by Parameswaran et al. (2018).

### 2.3. *Cross Validation*

The use of multiple dataset allows to sue cross-validation to prevent overfitting. One of the standard techniques in "machine learning"...

### 2.4. *Calibration of Predicted Counts*

Each dataset comes with systematic errors on absolute position and background normalisation

- allow for systematic shifts between observations
- allow for background scaling per observation $\alpha_m$

$$\lambda_m = \mathrm{PSF}_m \circledast (\mathbf{E}_m \cdot (\phi_m(\mathbf{x}|\delta_x, \delta_y) + \alpha_m \cdot \mathbf{B}_m)) \quad (9)$$

Show example residual images before and after calibration...

## 3. IMPLEMENTATION

### 3.1. *Jolideco Framework*

The goal of the reconstruction process is to optimize the a-posteriori likelihood defined by Equation 5. Given that each pixel in the reconstructed image represents an independent parameter this represents a high dimensional optimization problem. Differentiable programming frameworks such as `PyTorch` allow for solving these kind of high dimensional modeling problems, by using back-propagation and adapted optimization methods such as stochastic gradient decent.

The *Jolideco* method was implemented as an independent Python package, based on `PyTorch` as optimization back-end. We tried to define a modular, object oriented code structure, to allow parts of the algorithm to be flexible and interchangeable, such as choice of image normalisation scales, the choice of GMM models, optimization methods and serilisation formats for the reconstruction results as well as corresponding diagnostic information such as the trace of the posterior and prior likelihood values.

The package is available at https://github.com/jolideco/jolideco

- direct maximum a posteriori optimization using Pytorch etc.
- auto gradient computation
- ML optimizers such as ADAM
- GPU support

In addition we use `Scikit-Learn` (van der Walt et al. 2014) to learn the GMM.

We use `Astropy` (Astropy Collaboration 2018) for FITS serialisation and handling WCS transforms.

We use `Numpy` (Harris et al. 2020) for handling data arrays and `Matplotlib` (Hunter 2007) for plotting.

For optimization and internal data handling we use `PyTorch` (Paszke et al. 2019).

### 3.2. *Jolideco GMM Library*

We provide a selection of learned GMMs to be used with the patch prior for reproducibility and download. Convenience for future users. https://github.com/jolideco/jolideco-gmm-library

#### 3.2.1. *Zoran & Weiss*

As baseline reference GMM we use the original model provided by Zoran & Weiss (2011). The model was learned from the Berkley image database, by splitting the images into 8x8 patches.

#### 3.2.2. *GLEAM data*

High signal to noise Radio data from Hurley-Walker et al. (2022)

#### 3.2.3. *NRAO Jets, Jets, Jets*

For the specific application of jet analysis we learned a GMM patch prior from jet images. It encodes the prior knowledge of the preferred direction of the jet.

- One with horizontal direction, where the image to be analysed should be rotated. E.g. if prior knowledge from radio data is available.
- One with randomised direction, if now prior knowledge is available.

### 3.3. *Computational Performance*

We conduct as as series of performance benchmarks (see Appendix?) to asses the scalability of the method to a large number of observations as well as large images. Results are available in https://github.com/jolideco/jolideco-benchmarks

## 4. EXPERIMENTS

### 4.1. *Test Datasets*

TODO: Vinay writes this? [VLK: yes]

[VLK: We have devised four distinct arrangements of combinations of point and extended sources of differents shapes to test the algorithms. The arrangements are

"Gauss": A set of four point sources arranged around an extended Gaussian source

"Points": Point sources arranged with different separations

"Shield": Disk shaped flat extended source with superposed point sources and linear jet-like features

"Spiral": Extended thin double-spiral with a flat disk at the center and point sources adjacent to it

The arrangements are shown in Figure TBD, and the details of the locations and brightness of each component is listed in Table TBD. ]

We use the test datasets provided by Vinay et al. First we evaluate the performance of the method on a set of simulated observations. For this we assume:

- An instrument with good angular resolution, but low effective area (e.g. like Chandra)
- An instrument with worse angularr resolution, bur higher effective area (e.g. like XMM)

For both scenarios we assume a Gaussian PSF of sizes $\sigma = 2$ pixels and $\sigma = 5$. As true flux we consider the following scenarios:

- Point source with varying distances and brightness
- Disk and point sources of varying brightness
- Spiral and point sources of varying brightness

Background levels of $\lambda_{Bkg} = 0.001, 0.01 \text{and} 0.1 \text{counts/pixel}$.
Show results of experiments on simulated toy datasets

### 4.2. *Results*

The package is available at https://github.com/jolideco/jolideco-comparison

## 5. APPLICATION EXAMPLES

Show results from real observations

### 5.1. *Deep Chandra Observation*

### 5.2. *Combined XMM and Chandra Observation*

### 5.3. *Fermi-LAT Event Classes*

## 6. REPRODUCIBILITY

The paper is available at https://github.com/jolideco/jolideco-paper

## 7. SUMMARY & CONCLUSIONS

In this work we presented a new method for image deconvolution and denoising in the presence of Poisoon noise. Jolideco is great...

Extend *Jolideco* to handl spectral dimension at the same time.

## ACKNOWLEDGEMENTS

This work was conducted under the auspices of the CHASC International Astrostatistics Center. CHASC is supported by NSF grants DMS-21-13615, DMS-21-13397, and DMS-21-13605; by the UK Engineering and Physical Sciences Research Council [EP/W015080/1]; and by NASA 18-APRA18-0019. We thank CHASC members for many helpful discussions, especially Xiao-Li Meng and Katy McKeough. DvD was also supported in part by a Marie-Skodowska-Curie RISE Grant (H2020-MSCA-RISE-2019-873089) provided by the European Commission. Aneta Siemiginowska, Vinay Kashyap, and Doug Burke further acknowledge support from NASA contract to the Chandra X-ray Center NAS8-03060.

## REFERENCES


Astropy Collaboration. 2018, The Astrophysical Journal, 156, 123, doi: 10.3847/1538-3881/aabc4f

Bouman, K. L., Johnson, M. D., Zoran, D., et al. 2016, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

Cash, W. 1979, The Astrophysical Journal, 228, 939, doi: 10.1086/156922

Connors, A., Stein, N. M., van Dyk, D., Kashyap, V., & Siemiginowska, A. 2011, in Astronomical Society of the Pacific Conference Series, Vol. 442, Astronomical Data Analysis Software and Systems XX, ed. I. N. Evans, A. Accomazzi, D. J. Mink, & A. H. Rots, 463

Esch, D. N., Connors, A., Karovska, M., & van Dyk, D. A. 2004, The Astrophysical Journal, 610, 1213, doi: 10.1086/421761

Fish, D. A., Brinicombe, A. M., Pike, E. R., & Walker, J. G. 1995, J. Opt. Soc. Am. A, 12, 58, doi: 10.1364/JOSAA.12.000058

Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357, doi: 10.1038/s41586-020-2649-2

Hunter, J. D. 2007, Computing in Science & Engineering, 9, 90, doi: 10.1109/MCSE.2007.55

Hurley-Walker, N., Galvin, T. J., Duchesne, S. W., et al. 2022, PASA, 39, e035, doi: 10.1017/pasa.2022.17

Ingaramo, M., York, A., Hoogendoorn, E., et al. 2014, Chemphyschem : a European journal of chemical physics and physical chemistry, 15, doi: 10.1002/cphc.201300831

Lucy, L. B. 1974, Astronomical Journal, 79, 745, doi: 10.1086/111605

Parameswaran, S., Deledalle, C.-A., Denis, L., & Nguyen, T. Q. 2018, IEEE Trans. Image Process., 28, 687

Paszke, A., Gross, S., Massa, F., et al. 2019, in Advances in Neural Information Processing Systems 32, ed. H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, & R. Garnett (Curran Associates, Inc.), 8024–8035. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-libr pdf

Perry, K. M., & Reeves, S. J. 1994, in The Restoration of HST Images and Spectra - II, ed. R. J. Hanisch & R. L. White, 97

Pumpe, Daniel, Reinecke, Martin, & Enßlin, Torsten A. 2018, A&A, 619, A119, doi: 10.1051/0004-6361/201832781

Richardson, W. H. 1972, Journal of the Optical Society of America (1917-1983), 62, 55

Selig, Marco, & Enßlin, Torsten A. 2015, A&A, 574, A74, doi: 10.1051/0004-6361/201323006

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., et al. 2014, PeerJ, 2, e453, doi: 10.7717/peerj.453

Xu, L., Ren, J. S., Liu, C., & Jia, J. 2014, in Advances in Neural Information Processing Systems, ed. Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, & K. Weinberger, Vol. 27 (Curran Associates, Inc.). https://proceedings.neurips.cc/paper/2014/file/1c1d4df596d01da60385f0bb17a4a9e0-Paper.pdf

Zoran, D., & Weiss, Y. 2011, in 2011 International Conference on Computer Vision, 479–486, doi: 10.1109/ICCV.2011.6126278