
Celeste: Scalable variational inference for a generative model of astronomical images

Jeffrey Regier¹, Brenton Partridge², Jon McAuliffe¹, Ryan Adams²,
Matt Hoffman³, Dustin Lang⁴, David Schlegel⁵, Prabhat⁵

¹Department of Statistics, University of California, Berkeley

²School of Engineering and Applied Science, Harvard University

³Adobe Research, Adobe Systems Incorporated

⁴McWilliams Center for Cosmology, Carnegie Mellon University

⁵Lawrence Berkeley National Laboratory

1 Introduction

Stars and galaxies radiate photons. An astronomical image records photons—each originating from a particular celestial body or from background atmospheric noise—that pass through a telescope’s lens during an exposure. Multiple celestial bodies may contribute photons to a single image (e.g. Figure 1), and even to a single pixel of an image. Locating and characterizing the imaged celestial bodies is an inference problem central to astronomy. This paper presents Celeste: a generative model of astronomical images accompanied by a fast variational inference procedure.

To date, the algorithms proposed for this inference problem have been primarily heuristic, based on finding bright regions in the images [2, 3]. Though generative models have been developed for measuring galaxy shapes [4]—a subproblem of ours—to our knowledge, generative models for inferring celestial bodies’ locations and characteristics have not yet been published. Photon counts from light sources and background atmospheric noise are well approximated as independent Poisson processes, so a generative model is natural.

Difficulty scaling the inference for expressive generative models may have hampered their development: telescopic sky surveys produce very large amounts of data. For example, the Dark Energy Survey’s 570-megapixel digital camera, mounted on a four-meter telescope in the Andes, captures 300 gigabytes of sky images every night [5]. Once completed, the Large Synoptic Survey Telescope will house a 3200-megapixel camera producing eight terabytes of images nightly [6].

We focus on modeling images of stars and galaxies—celestial bodies whose characteristics are effectively static during human lifetimes. The model we present here applies exclusively to stars; modeling galaxies and stars together is the focus of ongoing work.

2 The model

Celeste is a generative probabilistic model of astronomical images, represented graphically in Figure 2. In this section we describe how Celeste relates stars’ latent characteristics to the observed



Figure 1: An image from the Sloan Digital Sky Survey [1], showing a galaxy from the constellation Serpens, 100 million light years from Earth, along with several other galaxies and many stars from our own galaxy.

pixel intensities in each image. The physics, and the instruments themselves, require us to account for a number of details. We explain in stages.

2.1 Stars

Celeste is hierarchical, with stars atop pixels. We suppose images have been resampled to a common pixel grid. Let A_n denote the pixel grid for image n , with 20 pixels of padding on each border; any stars contributing photons to image n must have coordinates in A_n . For each star $s = 1, \dots, S$, the latent random 2-vector u_s encodes its position, in image-pixel coordinates. The prior distribution on the (u_s) , for image n , is jointly iid uniform:

$$u_s \sim \text{Uniform}(A_n). \quad (1)$$

Inference in Celeste is based on a known value of S ; in Section 3.1 we describe our approach to setting that value.

The scalar parameter γ_s encodes a star’s observed brightness—the total radiation from the star that is expected to pass through a unit area of the Earth’s surface, facing the star, per unit time. The scalar parameter τ_s encodes a star’s temperature in Kelvin. We treat γ_s and τ_s as fixed parameters to estimate, rather than random variables, in order to improve the behavior of our current variational optimization algorithm. Adding posterior uncertainty about γ_s and τ_s is a subject of research.

2.2 Radiation at different wavelengths

In Celeste, the total rate of radiation arriving from a light source s , across all wavelengths, is proportional to its brightness γ_s . The distribution of that total rate across wavelengths depends on the nature of the light source. We model stars as idealized radiation sources, so-called “blackbodies”. The wavelength distribution for a blackbody of temperature τ is given by Planck’s law, a cornerstone of modern physics:

$$I(\tau, \lambda) = \frac{2\pi c^2 h}{\lambda^5} \left[\exp \left\{ \frac{hc}{\lambda k \tau} \right\} - 1 \right]^{-1}. \quad (2)$$

Here λ is a particular radiation wavelength, h is Planck’s constant, k is Boltzmann’s constant, and c is the speed of light. This simplification of the star source function could be replaced with a more realistic alternative if warranted, without complicating the inference.

2.3 Filter bands

Cameras for optical astronomy typically use charge-coupled devices (CCDs), which convert light into electrons [5]. A CCD is a grid of millions of pixels. It is designed to sum the radiation arriving at a pixel across all the wavelengths in a designated interval, or “band”—its filter band. Most telescopes have several filters, enabling them to measure radiation in multiple bands. In the Celeste model, the number of measured bands is B . The CCD for filter-band b has a corresponding filter function $\eta_b(\lambda)$. It indicates what fraction of arriving photons at wavelength λ the CCD will successfully detect, for each λ in band b . Thus, the final readout at each pixel is a weighted integral $d\lambda$ of the in-band radiation arriving at that pixel, with weights $\eta_b(\lambda)$.

2.4 Point-spread functions

Astronomical images are blurred by a combination of small-angle scattering in the earth’s atmosphere, the diffraction limit of the telescope, optical distortions in the camera, and charge diffusion within the silicon of the CCD detectors. Together these effects are called the point spread function (PSF) of a given image. Stars (other than our Sun) appear intrinsically as point sources, but, because of the PSF, their photons are spread over typically dozens of adjacent pixels. We model the action of the PSF as a mixture of K Gaussians. Consider pixel m (in band b of image n), having coordinates

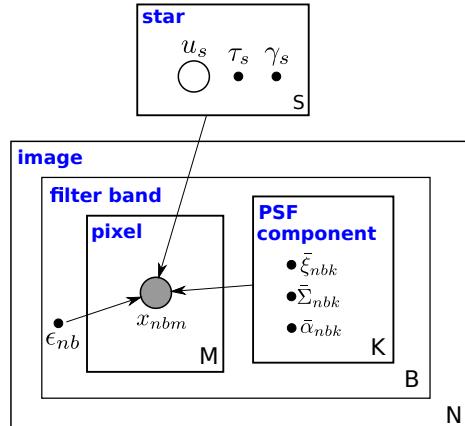


Figure 2: The Celeste graphical model. Shaded vertices represent observed random variables. Empty vertices represent latent random variables. Black dots represent constants. Edges signify conditional dependency. Rectangles are “plates”; they represent independent replication.

c_m . The probability that a photon from star s , positioned at u_s in pixel coordinates, lands on m is obtained from the density

$$f_s(n, b, m) = \sum_{k=1}^K \bar{\alpha}_{nbk} \phi(c_m; u_s + \bar{\xi}_{nbk}, \bar{\Sigma}_{nbk}). \quad (3)$$

Here ϕ is the bivariate normal density. Both K and the $(\bar{\alpha}_{nbk}, \bar{\xi}_{nbk}, \bar{\Sigma}_{nbk})$ are the parameters of the PSF, determined a priori.

2.5 Sky noise

The Earth's atmosphere reflects terrestrial light, creating noise in each telescopic image. We model this noise as a homogeneous spatial Poisson process, independent of all light sources. The sky-noise process produces photons at rate ϵ_{nb} uniformly, for each pixel in band b of image n . The noise rate depends on the image and the band because atmospheric conditions vary over time. Also, the atmospheric effects differ for imaging targets closer to the horizon.

2.6 The likelihood function

We now put the preceding five subsections together to describe the distribution of pixel intensities, conditional on the locations of all light sources. For each band b , we know the values of the filter function $\eta_b(\lambda)$ on a grid Υ_b of λ values—this is a matter of instrument design. If there were no PSF, the fraction of arriving photons detected at a band- b pixel containing a temperature- τ star would be

$$I_b(\tau) = \int I(\tau, \lambda) \eta_b(\lambda) d\lambda \approx \frac{1}{|\Upsilon_b|} \sum_{\lambda \in \Upsilon_b} I(\tau, \lambda) \eta_b(\lambda). \quad (4)$$

When we further account for the PSF, the sky noise, and all stars, the total number of photons arriving at pixel m (in band b of image n) becomes

$$x_{nbm}|u, \tau, \gamma \sim \text{Poisson}(F(n, b, m)). \quad (5)$$

Here the Poisson rate parameter $F(n, b, m)$ is the superposition of all the relevant independent Poisson rates, with light-source photons passing through the PSF:

$$F(n, b, m) = \epsilon_n + \sum_{s=1}^S \gamma_s I_b(\tau_s) f_s(n, b, m). \quad (6)$$

3 Variational inference

Let $u = (u_1, \dots, u_S)$ be the stars' positions and let $x = (x_{111}, \dots, x_{NBM})$ be the pixel intensities. Computing the posterior $p(u|x)$ is intractable: to apply Bayes' rule exactly, we need to evaluate

$$p(x) = \int \prod_{n=1}^N \prod_{b=1}^B \prod_{m=1}^M p(x_{nbm}|u) \prod_{s=1}^S p(u_s) du. \quad (7)$$

But the S -dimensional integral in (7) does not generally decompose into a product of low-dimensional integrals, and therefore cannot be computed.

Instead, we find a distribution that approximates the posterior well. For any distribution q over u ,

$$\log p(x) \geq \mathbb{E}_q [\log p(x, u)] - \mathbb{E}_q [\log q(u)] =: \mathcal{L}(q). \quad (8)$$

We call \mathcal{L} the evidence lower bound (ELBO). To find a distribution q^* that approximates the true posterior well, we maximize the ELBO over a set of candidate q 's. To facilitate optimization, we restrict this set to distributions in which the (u_s) are independent (the “mean field” assumption) and

$$q(u_s) \sim \text{Normal}(\Lambda_s, \Xi_s). \quad (9)$$

Then finding q^* is equivalent to finding the optimal (Λ_s, Ξ_s) for every star. By design, the expectations in \mathcal{L} can all be evaluated analytically. Space constraints prevent us from showing the full derivation here. Key results include, for a particular image n , band b , and pixel m ,

$$\mathbb{E}_q [F(n, b, m)] = \epsilon_n + \sum_{s=1}^S \gamma_s I_b(\tau_s) \phi(c_m; \bar{\xi}_{nbk} + \Lambda_s, \bar{\Sigma}_{nbk} + \Xi_s) \quad (10)$$

and

$$\mathbb{E}_q [\log F(n, b, m)] = \log \left\{ \epsilon_n + \sum_{s=1}^S \sum_{k=1}^K \exp(\mathbb{E}_q [\log y_{sk}]) \right\}, \quad \text{where} \quad (11)$$

$$y_{sk} = \gamma_s I_b(\tau_s) \bar{\alpha}_{bnk} \phi(c_m; u_s + \bar{\xi}_{nbk}, \bar{\Sigma}_{nbk}). \quad (12)$$

3.1 Optimization

To initialize our variational optimization, we convolve the images with matched filters to increase the signal-to-noise ratio [7]. We find each pixel whose value exceeds both the values of its neighboring pixels and an upper bound on the number of photons that could come from sky noise. We initialize the center of each such pixel as a light source. The number of such pixels determines S , the number of light sources we will assume are present in the image. In future work we will perform posterior inference about S as well.

Next, we minimize \mathcal{L} from equation (8) with L-BFGS [8] and automatic differentiation [9]. Because \mathcal{L} is not globally convex, the optimization path matters. For a particular filter band b_0 , we reparameterize \mathcal{L} using $\gamma'_s := \gamma_s I_{b_0}(\tau_s)$. We run L-BFGS several times, optimizing different blocks of coordinates. First, we optimize the (γ'_s) simultaneously for all stars, with the remaining parameters fixed, using just the pixels from band b_0 . For this band, because $I_{b_0}(\tau_s)$ appears in both the numerator and the denominator, \mathcal{L} is constant with respect to the (τ_s) . In other words, we do not need an estimate of the (τ_s) to estimate the (γ'_s) from this band. Next, we optimize the (τ_s) with L-BFGS for all stars and images simultaneously, with the other parameters fixed. Finally, we simultaneously optimize all parameters over all stars and images.

4 Experiments

We validated our procedure on synthetic data. To generate images, we set the (τ_s) and (u_s) to the temperatures and locations of actual stars and used the point spread functions from actual images, as approximated by a mixture of $K = 3$ Gaussians. For $N = 1$ image and $B = 3$ filter bands, we drew pixel values x_{nbm} from (5).

On synthetic data, our procedure converged consistently and quickly, for a wide range of initial estimates of the variational parameters and the unknown constants. Moreover, it recovered the true locations and temperatures: the average distance of the posterior mean star locations Λ_s from the truth—approximately 0.40 pixels upon initialization—decreased to 0.01 pixels. For simulated stars with temperatures ranging from 4000°K to 9000°K, the average discrepancy of estimated temperature was 55°K.

We also applied our procedure to a collection of 61 images from telescopes, each captured through five different filter bands, containing a total of 113 stars and approximately 150,000 pixels. These images do not overlap. We are planning an analysis of a data set five orders of magnitude larger, whose images have little overlap. These images will be processed largely in parallel, on a supercomputer, with minimal communication overhead.

Plots like Figure 3 reveal that the approximate posterior predictive means of x_{nbm} match the observed data. Our approximation of the point spread function is accurate, and our initialization routine has found the right number of stars. Though the true locations of the stars in the images are unknown, we validate our procedure by comparing our estimates to those from existing star catalogs. At initialization, the star locations Λ_s differed from the existing star catalogs by 0.40 pixels on average. After optimization, they differed by 0.22 pixels on average. (This ≈50% error reduction is significant to astronomers.) Thus, the variational inference improves on the initialization method.

Nine stars in our dataset have been observed with a spectrograph. From the resulting spectra, we estimated the temperature of each star from its peak wavelength. These estimates—though not highly accurate—are independent of the estimates from the Celeste model. According to the spectra, the star temperatures range from 3200°–7500°K. The median and mean differences between the spectral estimates and Celeste’s estimates are 435°K and 735°K, respectively. These differences are larger than Celeste’s misfit on synthetic data: our modeling assumptions are inexact on real data, and ground truth was only available for the synthetic data. Still, Celeste’s temperature predictions agree closely with the independent spectral estimates: a no-intercept regression of them on Celeste’s predictions has slope 0.95. The F -test for addition of that slope to an intercept has p -value 0.0024.

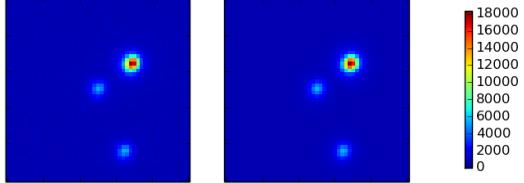


Figure 3: Left is a 51 pixel \times 51 pixel sub-region of an astronomical image, captured through the r band filter. Each pixel’s value corresponds to the number of photons that hit it. The right panel shows $\mathbb{E}_{q^*}[F(n, b, m)]$, the mean of x_{nbm} with respect to our posterior approximation.

Acknowledgments

This work was supported by the Director, Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

We thank the anonymous referees for their helpful comments.

References

- [1] Sky images observed by the SDSS telescope. http://classic.sdss.org/gallery/gal_data.html, 2014. [Online; accessed October 9, 2014].
- [2] R. H. Lupton, Z. Ivezić, et al. SDSS image processing II: The photo pipelines. Technical report, 2005.
- [3] C. Stoughton, R. H. Lupton, M. Bernardi, et al. Sloan Digital Sky Survey: Early Data Release. *The Astronomical Journal*, January 2002.
- [4] L. Miller, T. D. Kitching, et al. Bayesian galaxy shape measurement for weak lensing surveys - I. Methodology and a fast-fitting algorithm. *Monthly Notices of the Royal Astronomical Society*, 2007.
- [5] The dark energy survey. <http://www.darkenergysurvey.org/>, 2014. [Online; accessed October 9, 2014].
- [6] Large synoptic survey telescope. <http://www.lsst.org/lsst/about>, 2014. [Online; accessed October 7, 2014].
- [7] D. O. North. An analysis of the factors which determine signal/noise discrimination in pulsed-carrier systems. *Proceedings of the IEEE*, 1963.
- [8] D. C. Liu and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical programming*, 1989.
- [9] J. Rauch et al. ForwardDiff.jl: Julia package for performing forward mode automatic differentiation. <https://github.com/JuliaDiff/ForwardDiff.jl>, 2014.