Deep Potential: Recovering the gravitational potential from a snapshot of phase space

Gregory M. Green
Max Planck Institute for Astronomy
D-69117 Heidelberg, Germany
green@mpia.de

Yuan-Sen Ting
Institute for Advanced Study, Princeton
Princeton, NJ 08540, United States
ting@ias.edu

Abstract

One of the major goals of the field of Milky Way dynamics is to recover the gravitational potential field. Mapping the potential would allow us to determine the spatial distribution of matter – both baryonic and dark – throughout the Galaxy. We present a novel method for determining the gravitational field from a snapshot of the phase-space positions of stars, based only on minimal physical assumptions. We first train a normalizing flow on a sample of observed phase-space positions, obtaining a smooth, differentiable approximation of the phase-space distribution function. Using the collisionless Boltzmann equation, we then find the gravitational potential – represented by a feed-forward neural network – that renders this distribution function stationary. This method is far more flexible than previous parametric methods, which fit narrow classes of analytic models to the data. This is a promising approach to uncovering the density structure of the Milky Way, using rich datasets of stellar kinematics that will soon become available.

1 Introduction

To know the gravitational potential of the Milky Way is to know the three-dimensional distribution of matter. Stars and gas make up most of the baryonic mass of the Galaxy. However, dark matter is only detectable through its gravitational influence. Mapping the gravitational potential in 3D is therefore key to mapping the distribution of matter – both baryonic and dark – throughout the Galaxy.

The trajectories of stars orbiting in the Milky Way are guided by gravitational forces. Were it possible to directly measure accelerations of individual stars due to the Galaxy's gravitational field, then each star's acceleration would indicate the local gradient of the gravitational potential [12, 13]. This would allow us to use Hamiltonian neural network approaches to learn the gravitational potential [6]. However, the scale of these gravitational accelerations – on the order of $1\,\mathrm{cm\,s^{-1}\,yr^{-1}}$ – is beyond the ability of current spectroscopic and astrometric instruments to measure [14]. We instead observe a frozen snapshot of stellar positions and velocities. The gravitational potential determines how the phase-space density, the "distribution function," evolves in time. Unless one invokes further assumptions, any gravitational potential is consistent with any snapshot of the distribution function, as the potential only determines the *time evolution* of the distribution function. A critical assumption of most dynamical modeling of the Milky Way is therefore that the Galaxy is in a steady state, meaning that its distribution function does not vary in time [1, 2].

State-of-the-art dynamics modeling techniques generally work with simplified analytic models of the distribution function and gravitational potential. The results produced by such techniques can only be as good as the models that are assumed. This motivates us to go beyond simple parametric models. Here, we demonstrate a technique that learns highly flexible representations of both the distribution function and potential. Our method makes only minimal assumptions about the underlying physics:

Third Workshop on Machine Learning and the Physical Sciences (NeurIPS 2020), Vancouver, Canada.

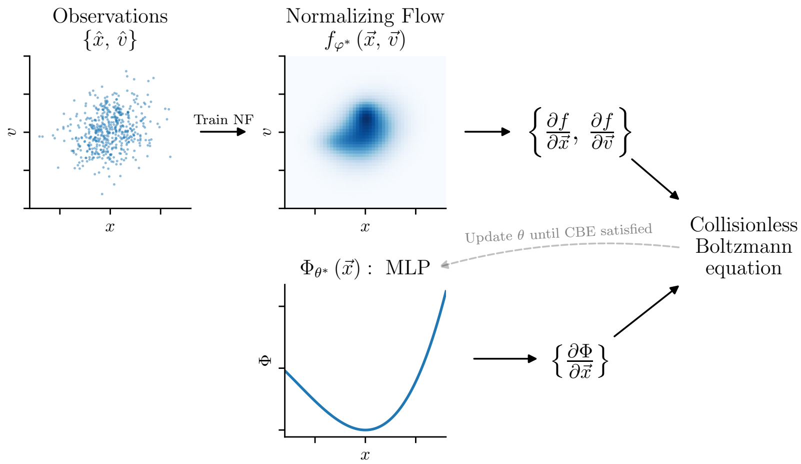

Figure 1: Overview of our method. Using observed phase-space information, we train a normalizing flow to represent the distribution function, $f\left(\vec{x},\,\vec{v}\right)$. We represent the gravitational potential, $\Phi\left(\vec{x}\right)$, by a feed-forward neural network. We update the neural network until the gradients of the potential and the distribution function satisfy the collisionless Boltzmann equation for a stationary system.

1. Stars orbit in a time-independent gravitational potential $\Phi\left(\vec{x}\right)$.
2. We have observed the phase-space coordinates of a population of stars that are statistically stationary (*i.e.*, whose phase-space distribution does not change in time).
3. The gravitational potential is related to the matter density, $\rho\left(\vec{x}\right)$, by Poisson's equation: $\nabla^2\Phi = 4\pi G\rho\left(\vec{x}\right)$. Matter density is non-negative everywhere. Thus, $\nabla^2\Phi \geq 0$.

We represent the distribution function using a normalizing flow, and the gravitational potential using a densely connected feed-forward neural network. We train the normalizing flow to represent the distribution of the observed phase-space coordinates of the stars, and then train the gravitational potential to render this distribution stationary, subject to the constraint that matter density must be positive. We thus use highly flexible representations for the distribution function and gravitational potential, and apply only minimal physical assumptions.

2 Method

Our first assumption is that stars orbit in a time-independent gravitational potential, $\Phi\left(\vec{x}\right)$. The density of an ensemble of stars in six-dimensional phase space (position $\vec{x}$ and velocity $\vec{v}$) is referred to as the *distribution function*, $f(\vec{x}, \vec{v})$. Liouville's theorem states that the total derivative of the distribution function of a collisionless system (in which the stars are not scattered by close-range interactions) is zero. For particles orbiting in a gravitational potential, this implies that

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial t} + \sum_{\text{dimension } i} \left(v_i \frac{\partial f}{\partial x_i} - \frac{\partial \Phi}{\partial x_i} \frac{\partial f}{\partial v_i} \right) = 0\,. \tag{1}$$

This is known as the "collisionless Boltzmann equation." Our second assumption, that the distribution function is stationary, implies that the density in any given region of phase space is constant in time: $\frac{\partial f}{\partial t} = 0$. This assumption links gradients of the distribution function to gradients of the gravitational potential. *Once we can describe the distribution function of a stationary system, in almost all cases, the gravitational potential can be uniquely determined by solving the collisionless Boltzmann equation* (Eq. 1). Note that we do not assume that the gravitational potential is sourced by the observed stellar population alone. Accordingly, we do not impose $\nabla^2\Phi = 4\pi G \int f\left(\vec{x}, \vec{v}\right) \mathrm{d}^3\vec{v}$. Additional mass components, such as unobserved stars or dark matter, also contribute to the gravitational field.

In practice, when we observe stellar populations, we obtain a discrete sample of points in phase space, which we will refer to as $\{\hat{x}, \hat{v}\}$. We do not directly observe the smooth distribution function,

$f(\vec{x}, \vec{v})$. A typical way of surmounting this difficulty is to fit a simple parametric model of the distribution function to the observed sample of phase-space locations of stars (e.g., [3, 10, 11, 16]). Here, we instead represent the distribution function as a normalizing flow, trained on an ensemble of stars with measured phase-space information. We refer to the normalizing flow as $f_\varphi(\vec{x}, \vec{v})$, where φ refers to the trainable parameters in the flow. The best-fit flow parameters are those that maximize the Poisson likelihood of the phase-space coordinates of the stars:

$$\varphi^* = \arg\max_\varphi \left[\ln p\left(\{\hat{x}, \hat{v}\} \mid \varphi\right) \right] = \arg\max_\varphi \left[\sum_{\text{star } k} \ln f_\varphi(\hat{x}_k, \hat{v}_k) \right] . \tag{2}$$

We thus obtain an approximation to the distribution function, f_{φ^*}. The great advantage of using a normalizing flow is that our representation is both highly flexible and auto-differentiable.

After learning the distribution function, we find the gravitational potential $\Phi(\vec{x})$ that renders the distribution function stationary. The distribution is stationary if the sum in Eq. (1) evaluates to zero everywhere in phase space. We also require that the matter density be non-negative everywhere in space. By Poisson's equation, which links the potential to the density, this implies that $\nabla^2\Phi \geq 0$. We parameterize the gravitational potential as a feed-forward neural network, which takes a 3-vector, $\vec{x}$, and returns a scalar, Φ. We denote the trainable parameters of this network by θ, and the resulting approximation function as $\Phi_\theta(\vec{x})$. We design a loss function that penalizes both non-stationarity and negative mass at phase-space points drawn from our approximation to the distribution function:

$$\theta^* = \arg\min_\theta \left\langle \sinh^{-1} \left| \frac{\partial f_{\varphi^*}}{\partial t} \right| + \lambda \sinh^{-1} \left(\max\left\{ -\nabla^2 \Phi_\theta, 0 \right\} \right) \right\rangle_{\vec{x}, \vec{v} \sim f_{\varphi^*}} , \tag{3}$$

where λ is a hyperparameter that sets the relative weight given to the stationarity and positive-mass conditions ($\lambda = 2$ in this work), and ${\partial f_{\varphi^*}}/{\partial t}$ is calculated by plugging ${\partial \Phi_\theta}/{\partial \vec{x}}$, ${\partial f_{\varphi^*}}/{\partial \vec{x}}$ and ${\partial f_{\varphi^*}}/{\partial \vec{v}}$ into Eq. (1). The $\sinh^{-1}$ function is linear for small values, but lessens the influence of outliers.

In the following demonstration, we use a normalizing flow architecture similar to `Glow` [7], but using rational-quadratic-spline coupling transforms [4] instead of affine coupling layers, and using an independent rescaling of each axis instead of `ActNorm` layers. One step of our flow architecture thus consists of an independent rescaling of each axis, an invertible 1×1 convolution and a rational-quadratic-spline coupling transform. Our flow architecture contains four steps. We find that we more accurately recover the distribution function with an *ensemble* of 960 normalizing flows trained on identical data than with a single normalizing flow. The feed-forward neural network with which we represent the gravitational potential has three densely connected hidden layers, each with 128 neurons, and produces a scalar output. We implement these models in PyTorch,[1] and obtain φ^* and θ^* through stochastic gradient descent, using the Rectified Adam optimizer [9].

3 Demonstration

We demonstrate our method on a toy physical system, in which both the potential and distribution function can be expressed analytically. This constitutes a ground truth on which we can verify our method. We choose the Plummer sphere, a self-gravitating system with a spherically symmetric density and gravitational potential, given by

$$\rho(r) = \frac{3}{4\pi} \left(1 + r^2\right)^{-5/2} , \qquad \Phi(r) = -\left(1 + r^2\right)^{-1/2} . \tag{4}$$

The Plummer sphere admits a stationary distribution function with an isotropic velocity distribution, in which the distribution function is only a function of the energy E of the tracer particle (for simplicity, we set mass $m = 1$ for all the particles):

$$f(\vec{r}, \vec{v}) \propto \begin{cases} \left[-E\left(\vec{r}, \vec{v}\right)\right]^{7/2}, & E < 0 \\ 0, & E \geq 0 \end{cases} , \qquad \text{where } E = \frac{1}{2} v^2 + \Phi(r) . \tag{5}$$

We generate mock data by sampling 2^{17} (131,072) phase-space points drawn from the above distribution function. Using these points as input data, we first fit the distribution function using an

[1] https://github.com/tingyuansen/deep-potential. A Tensorflow 2.3 implementation can be found at https://github.com/gregreen/deep-potential.

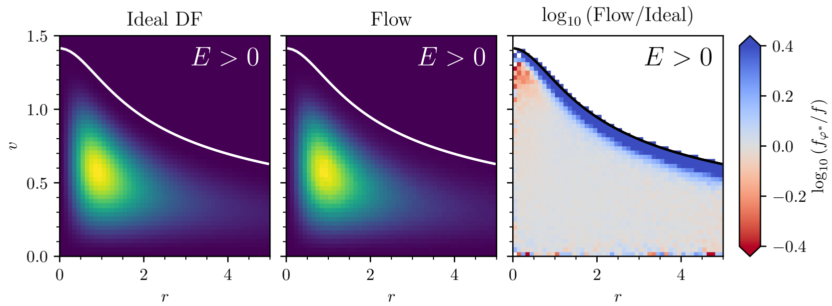

Figure 2: The ideal Plummer sphere distribution function (left panel), our trained ensemble of normalizing flows (middle panel), and a comparison of the two (right panel). We depict phase space in terms of radius and velocity, integrating over the four angular dimensions. Our ensemble of normalizing flows performs well in regions of non-negligible density.

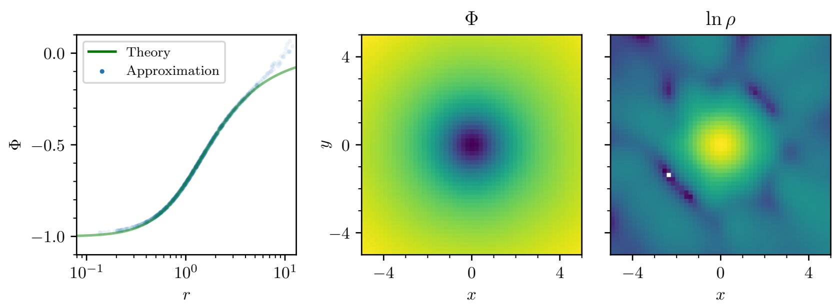

Figure 3: Left panel: comparison of the theoretical Plummer sphere potential with our result at random points drawn from phase space. The middle and right panels show our recovered potential and matter density in a 2D slice of space with $z = 0$.

ensemble of 960 normalizing flows (see Eq. 2). Our results are shown in Fig. 2. Using our ensemble of flows, we then draw 2^{15} (32,768) phase-space points, and calculate the gradients $\partial f_{\varphi^*}/\partial\vec{x}$, $\partial f_{\varphi^*}/\partial\vec{v}$. We use these samples to fit the gravitational potential (see Eq. 3). Our results are shown in Fig. 3. We accurately recover the potential over a wide range of radii.

4 Conclusions

In this paper, we have shown that it is possible to accurately recover the gravitational potential of a stationary system using a snapshot of a sample of phase-mixed tracers. Auto-differentiable tensor frameworks are ideal tools for accomplishing this task, because of the need for smooth and differentiable – yet highly flexible – representations of the distribution function and gravitational potential. We have demonstrated that our method works on ideal mock data. Future work will examine how to take observational selection functions and errors into account. An additional avenue of research is how to apply our framework to physical systems that are not entirely phase-mixed, and therefore not completely stationary.

The *Gaia* space telescope is currently surveying the parallaxes and proper motions of over a billion stars, and is additionally determining radial velocities of tens of millions of stars [5]. Ground-based spectroscopic surveys are set to deliver millions more high-precision radial velocities over the coming years [15, 17, 8]. We thus will soon have access to precise six-dimensional phase-space coordinates of tens of millions of stars throughout the Galaxy. The method that we present in this paper provides a means of extracting the gravitational potential – and therefore the three-dimensional distribution of baryonic and dark matter in the Galaxy – from these rich datasets, starting only from minimal physical assumptions (steady-state dynamics in a background gravitational field that corresponds to a positive matter density), and without resorting to restricted analytical models. This method therefore has the potential to reveal the full, unseen mass distribution of the Galaxy, using only a set of visible kinematic tracers – the stars.

Broader Impact

This work will advance our understanding of the dynamics of the Milky Way, and has the potential to uncover the distribution of both baryonic and dark matter throughout the Galaxy. The distribution of dark matter within a large spiral galaxy, such as the Milky Way, is of great interest to astronomers, particularly because it may help constrain models of the physical nature of dark matter. A great many people in the broader public are passionate about astronomy, and any advances in our understanding of the distribution of dark matter – and anything we learn about the nature of dark matter from its spatial distribution – will interest them. The authors do not foresee broader ethical concerns arising from this work.

Acknowledgments and Disclosure of Funding

Gregory Green acknowledges funding from the Alexander von Humboldt Foundation, through the Sofja Kovalevskaja Award.

Yuan-Sen Ting is grateful to be supported by the NASA Hubble Fellowship grant HST-HF2-51425.001 awarded by the Space Telescope Science Institute.

The authors thank Harshil Kamdar, Stephen Portillo, Hans-Walter Rix, Douglas Finkbeiner and Joshua Speagle for insightful discussions during the development of this method.

References

[1] J. Binney. Dynamics for galactic archaeology. *New Astronomy Review*, 57(3-4):29–51, Sept. 2013. doi: 10.1016/j.newar.2013.08.001.

[2] J. Bland-Hawthorn and O. Gerhard. The Galaxy in Context: Structural, Kinematic, and Integrated Properties. *Annual Review of Astronomy and Astrophysics*, 54:529–596, Sept. 2016. doi: 10.1146/annurev-astro-081915-023441.

[3] J. Bovy and H.-W. Rix. A Direct Dynamical Measurement of the Milky Way's Disk Surface Density Profile, Disk Scale Length, and Dark Matter Profile at 4 kpc < R < 9 kpc. *Astrophysical Journal*, 779(2):115, Dec. 2013. doi: 10.1088/0004-637X/779/2/115.

[4] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios. Neural Spline Flows. *arXiv e-prints*, art. arXiv:1906.04032, June 2019.

[5] Gaia Collaboration, T. Prusti, J. H. J. de Bruijne, et al. The Gaia mission. *Astronomy and Astrophysics*, 595:A1, Nov. 2016. doi: 10.1051/0004-6361/201629272.

[6] S. Greydanus, M. Dzamba, and J. Yosinski. Hamiltonian Neural Networks. *arXiv e-prints*, art. arXiv:1906.01563, June 2019.

[7] D. P. Kingma and P. Dhariwal. Glow: Generative Flow with Invertible 1x1 Convolutions. *arXiv e-prints*, art. arXiv:1807.03039, July 2018.

[8] J. A. Kollmeier, G. Zasowski, H.-W. Rix, et al. SDSS-V: Pioneering Panoptic Spectroscopy. *arXiv e-prints*, art. arXiv:1711.03234, Nov. 2017.

[9] L. Liu, H. Jiang, P. He, et al. On the Variance of the Adaptive Learning Rate and Beyond. *arXiv e-prints*, art. arXiv:1908.03265, Aug. 2019.

[10] P. J. McMillan and J. J. Binney. Analysing surveys of our Galaxy - II. Determining the potential. *Monthly Notices of the Royal Astronomical Society*, 433(2):1411–1424, Aug. 2013. doi: 10.1093/mnras/stt814.

[11] T. Piffl, J. Binney, P. J. McMillan, et al. Constraining the Galaxy's dark halo with RAVE stars. *Monthly Notices of the Royal Astronomical Society*, 445(3):3133–3151, Dec. 2014. doi: 10.1093/mnras/stu1948.

[12] C. Quercellini, L. Amendola, and A. Balbi. Mapping the galactic gravitational potential with peculiar acceleration. *Monthly Notices of the Royal Astronomical Society*, 391(3):1308–1314, Dec. 2008. doi: 10.1111/j.1365-2966.2008.13968.x.

[13] A. Ravi, N. Langellier, D. F. Phillips, et al. Probing Dark Matter Using Precision Measurements of Stellar Accelerations. *Physical Review Letters*, 123(9):091101, Aug. 2019. doi: 10.1103/PhysRevLett.123.091101.

[14] H. Silverwood and R. Easther. Stellar accelerations and the galactic gravitational field. *Publications of the Astronomical Society of Australia*, 36:e038, Oct. 2019. doi: 10.1017/pasa.2019.25.

[15] M. Steinmetz, T. Zwitter, A. Siebert, et al. The Radial Velocity Experiment (RAVE): First Data Release. *Astronomical Journal*, 132(4):1645–1668, Oct. 2006. doi: 10.1086/506564.

[16] W. H. Trick, J. Bovy, and H.-W. Rix. Action-Based Dynamical Modeling for the Milky Way Disk. *Astrophysical Journal*, 830(2):97, Oct. 2016. doi: 10.3847/0004-637X/830/2/97.

[17] G. Zhao, Y.-H. Zhao, Y.-Q. Chu, et al. LAMOST spectral survey — An overview. *Research in Astronomy and Astrophysics*, 12(7):723–734, July 2012. doi: 10.1088/1674-4527/12/7/002.