

---

# GalaxyScore: Estimating Local Dark Matter Density in Galaxies with Score Matching

---

Sung Hak Lim<sup>1,2</sup> David Shih<sup>2</sup> Matthew R. Buckley<sup>2</sup> Eric Putney<sup>2</sup>

<sup>1</sup>Particle Theory and Cosmology Group, Center for Theoretical Physics of the Universe,  
Institute for Basic Science (IBS), Yuseong-gu, Daejeon 34126, Republic of Korea

<sup>2</sup>NHETC, Department of Physics and Astronomy,  
Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, USA  
sunghak.lim@ibs.re.kr  
shih,mbuckley,eputney@physics.rutgers.edu

## Abstract

We introduce GalaxyScore, an unsupervised machine learning method for estimating local dark matter density in galaxies using collisionless Boltzmann equation solvers with a stellar kinematics catalog. Current approaches require the derivatives of normalizing flows and have high computational overhead and unnecessary noise. GalaxyScore bypasses the explicit density evaluation by using score matching to directly estimate the log-probability density derivatives from stellar positions and velocities. We anticipate that this direct approach will provide more accurate and computationally efficient solutions compared to existing methods. We demonstrate GalaxyScore on simulated spherical galaxies from the Gaia Challenge Datasets.

## 1 Introduction

Understanding the dark matter distribution in galaxies is crucial for uncovering its fundamental nature [1, 2]. This distribution can be revealed by measuring the total gravitational field acting on within a galaxy (see [3–5] for reviews on the range of available methods). In particular, *collisionless Boltzmann equation* (CBE)-based methods relate the stellar distribution function  $f(\vec{r}, \vec{v}; t)$  – the probability density function of finding a star at position  $\vec{r}$  and velocity  $\vec{v}$  at time  $t$  – to the gravitational acceleration  $\vec{a}(\vec{r}; t)$  through the following differential equation:

$$\frac{\partial f}{\partial t} + v_i \frac{\partial f}{\partial r_i} + a_i \frac{\partial f}{\partial v_i} = 0. \quad (1)$$

As the time derivative of  $f$  is not directly measurable, analyses first choose a stellar population in approximate equilibrium ( $\partial f / \partial t = 0$ ) and solve for the acceleration. This approach yields consistent results within the Milky Way [6] unless the chosen population experienced major interaction events, which can bias the acceleration estimation [7]. The total mass density  $\rho(\vec{r})$  can be found from  $\nabla \cdot \vec{a} = -4\pi G\rho(\vec{r})$  and the visible matter density subtracted to find the dark matter density  $\rho_{\text{DM}}(\vec{r})$ .

Machine learning has been increasingly applied to CBE-based dark matter density estimation, particularly normalizing flows [8, 9]. Normalizing flows use neural networks to learn coordinate transformations from a standard normal distribution to the data. The learned transformation can estimate the probability density function via the change of variables formula. As  $f(\vec{r}, \vec{v})$  is a probability density function, normalizing flows trained on stellar positions and velocities directly model the distribution function, often using  $f(\vec{r}, \vec{v}) = p(\vec{r})p(\vec{v}|\vec{r})$  (where  $p(\vec{r})$  is the stellar number density at  $\vec{r}$  and  $p(\vec{v}|\vec{r})$  is the conditional probability distribution of  $\vec{v}$  at  $\vec{r}$ ).

Following validation on analytic models [10–13] and simulated Milky Way-like galaxies [14, 15], this approach was used to infer the local dark matter density of the Milky Way [6, 16] using Gaia

DR3 catalogs [17]. Extensions and further applications of normalizing flows to dynamical modeling of galaxies include handling dust extinction effects [18], upsampling simulated galaxies [19], and analyses of dwarf spheroidal galaxies with limited observed kinematic information [20, 21].

When computing the acceleration field from density derivatives (as in [12–14]), the acceleration field is found by minimizing the mean square error (MSE) loss of the RHS of Eq. 1 at each point  $\vec{r}$

$$\mathcal{L}_{\text{CBE}, \vec{r}}(\vec{a}(\vec{r})) = \int d\vec{v} p(\vec{v}|\vec{r}) \left| v_i \frac{\partial \log f}{\partial r_i} + a_i(\vec{r}) \frac{\partial \log f}{\partial v_i} \right|^2. \quad (2)$$

Minimizing  $\mathcal{L}_{\text{CBE}, \vec{r}}$  translates into solving CBEs with velocity samples drawn from  $p(\vec{v}|\vec{r})$ . The analytic solution is the least squares solution of the overdetermined system of CBEs [12–14]:

$$\vec{a}(\vec{r}) = A^{-1}B, \quad A_{ij}(\vec{r}) = E_{p(\vec{v}|\vec{r})} \left[ \frac{\partial \log f}{\partial v_i} \frac{\partial \log f}{\partial v_j} \right], \quad B_i(\vec{r}) = -E_{p(\vec{v}|\vec{r})} \left[ \frac{\partial \log f}{\partial v_i} v_j \frac{\partial \log f}{\partial r_j} \right]. \quad (3)$$

Thus, accelerations can be estimated by solving the CBEs if we 1) have estimated the log-density derivatives  $\frac{\partial \log f}{\partial v_i}$  and  $\frac{\partial \log f}{\partial r_j}$  and 2) sampled from  $p(\vec{v}|\vec{r})$  to compute  $A_{ij}$  and  $B_i$  with sample means.

Normalizing flows find the log-density derivatives to solve the CBE by first computing the probability density and then differentiating. This amplifies fitting errors through derivative computation and has high computational costs, as backpropagation through the entire network is required for each gradient evaluation. Ensemble averaging can limit noise, but this exacerbates the cost issue.

Crucially, what appears in the CBE is actually the *score function* and not the phase space densities themselves. Therefore, the CBE should be an amenable target for *score matching* [22] based approaches, which estimate log-density derivatives without requiring explicit density computation. Additionally, score matching can serve as a generative model via Langevin dynamics, enabling us to generate the required samples from  $p(\vec{v}|\vec{r})$  within a unified framework. In the following sections, we show how this approach can be applied to CBE-solvers for gravitational acceleration and dark matter density estimation. We dub our approach **GalaxyScore**.

## 2 Score Matching

### 2.1 Log-Density Derivative Estimation

Score matching [22] uses the *explicit score matching loss*, which directly compares the score model  $\vec{s}(\vec{x}; \theta_s)$  parametrized by  $\theta_s$  with the true score function  $\frac{\partial}{\partial x_i} \log p(\vec{x})$  for a given data point  $\vec{x}$ :

$$\mathcal{L}_{\text{ESM}}(\theta_s) = \frac{1}{2} \int d\vec{x} p(\vec{x}) \left| \vec{s}(\vec{x}; \theta_s) - \frac{\partial}{\partial x_i} \log p(\vec{x}) \right|^2. \quad (4)$$

The experimentally inaccessible  $\frac{\partial}{\partial x_i} \log p(\vec{x})$  is removed from the loss using the integration-by-parts approach of [22], resulting in the *implicit score matching loss*:

$$\mathcal{L}_{\text{ISM}}(\theta_s) = \int d\vec{x} p(\vec{x}) \left[ \frac{\partial}{\partial x_i} s_i(\vec{x}; \theta_s) + \frac{1}{2} |\vec{s}(\vec{x}; \theta_s)|^2 \right] \quad (5)$$

This derivation uses integration by parts and assumes boundary terms vanish as  $p(\vec{x}) \rightarrow 0$  when  $|\vec{x}|^2 \rightarrow \infty$ . Since this loss function is simply an expectation of a function of the score model  $\vec{s}(\vec{x}; \theta_s)$ , we can evaluate it using sample expectations from training data.

We directly use the score function  $\vec{s}(\vec{r}, \vec{v})$  learned from the 6D kinematic dataset  $\{(\vec{r}, \vec{v})\}$  to evaluate the log density derivatives  $\frac{\partial}{\partial r_i} \log f$  and  $\frac{\partial}{\partial v_i} \log f$  needed in CBE solvers. We use a multilayer perceptron (MLP) to model  $\vec{s}(\vec{r}, \vec{v})$ , and the details are described in Appendix A.

### 2.2 Sampling with Langevin Dynamics

To evaluate expectations in the above loss functions and solutions of CBE, we need to draw samples from  $p(\vec{v}|\vec{r})$ . While various generative models could be used for this sampling task, such as continuous

normalizing flows [23, 24] and diffusion models [25, 26], we employ Langevin dynamics to draw samples directly from score functions to maintain a unified score-based framework.<sup>1</sup>

Langevin dynamics is a process describing the damping motion of a particle experiencing stochastic forces. The equation of motion for Langevin dynamics is a stochastic differential equation (SDE):

$$dx_i = \frac{\partial}{\partial x_i} \log p(\vec{x}) dt + \sqrt{2} dW_i, \quad (6)$$

where  $dW_i$  is the Wiener process with zero mean and variance  $dt$ . With this SDE, the probability density function  $q(\vec{x}; t)$  at time  $t$  evolves according to the Fokker-Planck equation:

$$\frac{\partial q(\vec{x}; t)}{\partial t} = -\frac{\partial}{\partial x_i} \left[ \left( \frac{\partial}{\partial x_i} \log p(\vec{x}) \right) q(\vec{x}; t) \right] + \frac{\partial^2 q(\vec{x}; t)}{\partial x_i^2}. \quad (7)$$

This equation has a stationary solution  $q(\vec{x}) = p(\vec{x})$ ; therefore, if we evolve samples from any initial distribution using the SDE for sufficient time, the resulting distribution converges to  $p(\vec{x})$ . Thus, we can draw samples from the score functions by evolving some initial distribution with the above Langevin dynamics for a sufficient time.

To generate samples from  $p(\vec{v}|\vec{r})$  specifically, we train a conditional score function  $\vec{s}(\vec{v}|\vec{r})$  to estimate  $\frac{\partial}{\partial v_i} \log p(\vec{v}|\vec{r})$ . Details of the sampling setup are described in Appendix A.

### 3 Acceleration Matching

With score matching, we can now compute the solutions of CBE in Eq. 3 without computing derivatives of neural networks. However, to compute mass density, we must take derivatives of the estimated acceleration field. For regularizing the estimated acceleration field, we use an MLP  $\vec{a}(\vec{r}; \theta_a)$  parameterized by  $\theta_a$  that directly regresses the CBE solution. We train this model by minimizing the following MSE loss that matches neural network outputs to the estimated acceleration:

$$\mathcal{L}_{\text{AccM}}(\theta_a) = \int d\vec{r} p(\vec{r}) |\vec{a}(\vec{r}; \theta_a) - A^{-1} B|^2. \quad (8)$$

The position samples for evaluating  $\int d\vec{r} p(\vec{r})$  are drawn from  $\vec{s}(\vec{r}, \vec{v})$  using Langevin dynamics, and we simply drop generated velocity components. After training, the mass density is estimated from  $\nabla \cdot \vec{a}(\vec{r}; \theta_a)$ . One may further introduce a potential model  $\phi(\vec{r}; \theta_a)$  to impose conservative force constraints as in the Deep Potential approaches [10, 11, 15, 18], but for this proof-of-concept demonstration, we do not impose such constraints for simplicity.

### 4 Experiment

We demonstrate the GalaxyScore method on the NonplumCoreIso and NonplumCuspIso datasets from the Gaia Challenge dataset suite [27, 28]. These datasets contain simulated stellar populations tracing gravity sourced by dark matter halos with the following mass density:

$$\rho_{\text{DM}}(\vec{r}) = \rho_0 \left( \frac{r}{r_{\text{DM}}} \right)^{-\gamma_{\text{DM}}} \left[ 1 + \left( \frac{r}{r_{\text{DM}}} \right) \right]^{\gamma_{\text{DM}}-3}, \quad r_{\text{DM}} = 1 \text{ kpc}. \quad (9)$$

The Navarro-Frenk-White (NFW) profile [29–31] has  $\gamma_{\text{DM}} = 1$ . Stars are distributed in a modified Plummer profile [32]:

$$p(\vec{r}) \propto (r/r_*)^{-1} \left[ 1 + (r/r_*)^2 \right]^{-2}. \quad (10)$$

Velocity distributions are set to be isotropic and kinematically consistent with the given gravitational field and stable stellar population. Each dataset uses the following parameters:

- NonplumCoreIso:  $\gamma_{\text{DM}} = 0$ ,  $\rho_0 = 0.4 \text{ M}_\odot/\text{pc}^3$ ,  $r_* = 1 \text{ kpc}$
- NonplumCuspIso:  $\gamma_{\text{DM}} = 1$ ,  $\rho_0 = 0.064 \text{ M}_\odot/\text{pc}^3$ ,  $r_* = 0.25 \text{ kpc}$

---

<sup>1</sup>This approach uses similar mathematical foundations as diffusion models [25, 26], both leveraging score functions for sampling, but avoids the complexity of choosing noising strategy.

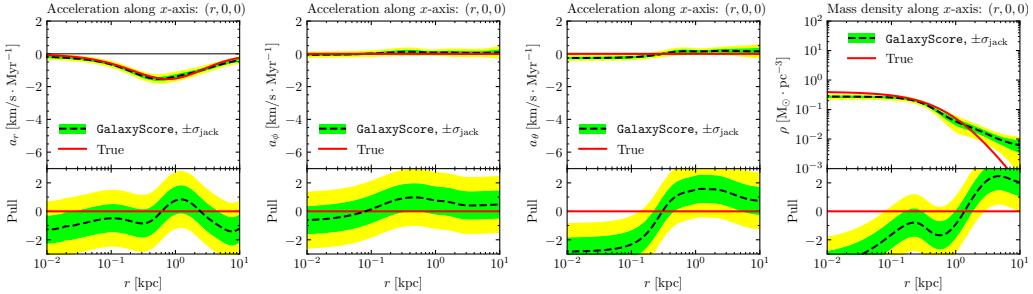


Figure 1: True (red) and GalaxyScore (black) acceleration and dark matter density along the  $x$ -axis for the NonplumCoreIso dataset, along with  $1\sigma$  (green) and  $2\sigma$  (yellow) statistical uncertainties. The lower panel of each figure shows the pulls, defined as the deviation from the true value divided by the uncertainty  $\sigma$ .

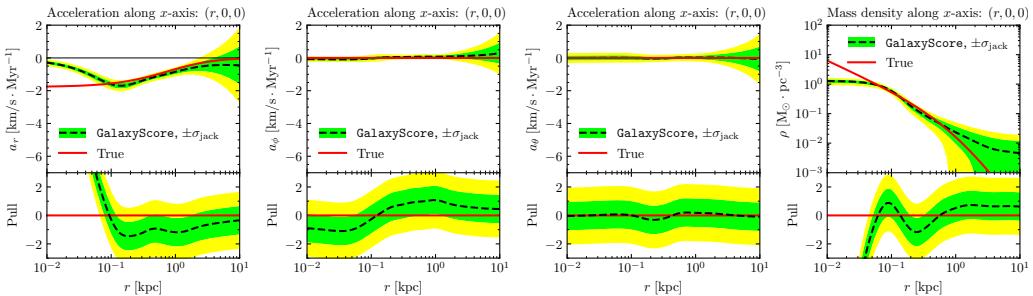


Figure 2: Acceleration and dark matter density along the  $x$ -axis for the NonplumCuspIso dataset. Simulation-truth, GalaxyScore, and uncertainties are displayed as in Figure 1. The lower panel of each figure shows the pulls, defined as the deviation from the true value divided by the uncertainty  $\sigma$ .

We apply GalaxyScore to datasets containing 10,000 samples each, using 25% for validation. These datasets were originally designed as benchmarks for methods analyzing dwarf spheroidal galaxies; hence, the sample size is quite low compared to those for the Milky Way analyses [6, 16]. Though we anticipate that the local dark matter density estimation on these datasets suffers from large statistical uncertainties, we use these datasets to test CBE-solvers in the low statistics regime. We estimate statistical uncertainties using a jackknife method [33] with 20 subsampled datasets. Neural network architectures used for modeling score and acceleration functions are explained in Appendix A, and details of the uncertainty quantification is described in Appendix B.

Figure 1 shows the results on the NonplumCoreIso dataset. GalaxyScore results match the true values well for both acceleration and mass density in the region well-supported by data,  $r < r_* = 1$  kpc. Due to the limited sampling beyond  $r > 10$  kpc, the uncertainty in this region grows rapidly, so we clipped them out. However, about three  $\sigma$  deviations are observed for  $a_\theta$  and  $\rho$  for  $r < 0.1$  kpc. Nevertheless, the same deviation appears in well-established flow-based methods (see Appendix C), suggesting that the bias may be due to the systematic bias in the dataset.

For the NonplumCuspIso results in Figure 2, the range with small statistical uncertainties is much narrower than in Figure 1. This is because the stellar distribution of NonplumCuspIso is more confined than that of NonplumCoreIso, and the precision falloff starts earlier.

The other notable observation in the NonplumCuspIso results is the non-negligible bias from the true values at small galactocentric radii  $r < 0.1$  kpc due to the inductive bias of neural networks in interpolations [34]. The expected minimum galactocentric distance is 0.029 kpc and 0.046 kpc for sample sizes of 7,500 (training) and 2,500 (validation), respectively. Since neural networks smoothly interpolate in data-sparse regions, the cuspy halo profile in NonplumCuspIso becomes artificially flattened, while the cored profile in NonplumCoreIso aligns with this smoothness bias.

One could explicitly model the singularity as done in the JFlow method [21], but combining it with score matching is beyond the scope of this paper. This limitation in learning singularities should be considered when applying the technique to systems expected to have cuspy density profiles or other singularities.

Note that accurate mass density estimation in the data-rich region indicates that score matching provides higher-quality second-order density derivatives. In the case of flows, the higher-order derivatives are less reliable, so we must use kernel smoothing [6, 14] or train a network to learn the solution of the CBE [10, 11] to infer the dark matter density using only first-order density derivatives. This reliability is one of the main precision advantages of score matching over flows. We discuss results from flow-based approaches in Appendix C.

## 5 Conclusion

In this paper, we introduce **GalaxyScore**, a score matching-based collisionless Boltzmann equation (CBE) solver for gravitational acceleration and dark matter density estimation. This is a fast and highly precise method for directly estimating log-stellar phase space distribution function derivatives from data; improving dark matter density estimation compared to methods using normalizing flows. Note that **GalaxyScore** does not change the fundamental CBE-solver problem but provides a better way to solve it, so we can apply this approach to other CBE applications, for example, estimating stellar selection efficiencies due to dust extinction [18]. We anticipate that this method will help future missions for dynamical modeling of galaxies and understanding galactic dark matter distributions.

## Acknowledgments and Disclosure of Funding

This work was supported by IBS under the project code, IBS-R018-D1. This work was also supported by the DOE under Award Number DOE-SC0010008. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 using NERSC award HEP-ERCAP0027491. This work was also performed in part at Aspen Center for Physics, which is supported by National Science Foundation grant PHY-2210452. The authors acknowledge the Office of Advanced Research Computing (OARC) at Rutgers, The State University of New Jersey for providing access to the Amarel cluster and associated research computing resources that have contributed to the results reported here (URL: <https://oarc.rutgers.edu>).

## References

- [1] J. L. Feng, “Dark Matter Candidates from Particle Physics and Methods of Detection,” *Ann. Rev. Astron. Astrophys.* **48**, 495–545 (2010), arXiv:1003.0904 [astro-ph.CO].
- [2] L. E. Strigari, “Galactic Searches for Dark Matter,” *Phys. Rept.* **531**, 1–88 (2013), arXiv:1211.7090 [astro-ph.CO].
- [3] W. Wang, J. Han, M. Cautun, Z. Li, and M. N. Ishigaki, “The mass of our Milky Way,” *Sci. China Phys. Mech. Astron.* **63**, 109801 (2020), arXiv:1912.02599 [astro-ph.GA].
- [4] P. F. de Salas and A. Widmark, “Dark matter local density determination: recent observations and future prospects,” *Rept. Prog. Phys.* **84**, 104901 (2021), arXiv:2012.11477 [astro-ph.GA].
- [5] J. A. Hunt and E. Vasiliev, “Milky way dynamics in light of gaia,” *New Astronomy Reviews* **100**, 101721 (2025), arXiv:2501.04075 [astro-ph.GA].
- [6] S. H. Lim, E. Putney, M. R. Buckley, and D. Shih, “Mapping dark matter in the Milky Way using normalizing flows and Gaia DR3,” *JCAP* **01**, 021 (2025), arXiv:2305.13358 [astro-ph.GA].
- [7] T. Haines, E. D’Onghia, B. Famaey, C. Laporte, and L. Hernquist, “Implications of a Time-varying Galactic Potential for Determinations of the Dynamical Surface Density,” *ApJ* **879**, L15 (2019), arXiv:1903.00607 [astro-ph.GA].
- [8] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, “Normalizing flows for probabilistic modeling and inference,” *Journal of Machine Learning Research* **22**, 1–64 (2021).
- [9] I. Kobyzhev, S. J. Prince, and M. A. Brubaker, “Normalizing flows: An introduction and review of current methods,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **43**, 3964–3979 (2021).

- [10] G. M. Green and Y.-S. Ting, “Deep Potential: Recovering the gravitational potential from a snapshot of phase space,” Machine Learning and the Physical Sciences, Workshop at the 34th Conference on Neural Information Processing Systems (NeurIPS) (2020), arXiv:2011.04673 [astro-ph.GA].
- [11] G. M. Green, Y.-S. Ting, and H. Kamdar, “Deep Potential: Recovering the Gravitational Potential from a Snapshot of Phase Space,” *ApJ* **942**, 26 (2023), arXiv:2205.02244 [astro-ph.GA].
- [12] J. An, A. P. Naik, N. W. Evans, and C. Burrage, “Charting galactic accelerations: when and how to extract a unique potential from the distribution function,” *Monthly Notices of the Royal Astronomical Society* **506**, 5721–5730 (2021), <https://academic.oup.com/mnras/article-pdf/506/4/5721/42537089/stab2049.pdf>.
- [13] A. P. Naik, J. An, C. Burrage, and N. W. Evans, “Charting galactic accelerations - II. How to ‘learn’ accelerations in the solar neighbourhood,” *Monthly Notices of the Royal Astronomical Society* **511**, 1609–1621 (2022), arXiv:2112.07657 [astro-ph.GA].
- [14] M. R. Buckley, S. H. Lim, E. Putney, and D. Shih, “Measuring Galactic dark matter through unsupervised machine learning,” *Mon. Not. Roy. Astron. Soc.* **521**, 5100–5119 (2023), arXiv:2205.01129 [astro-ph.GA].
- [15] T. Kalda, G. M. Green, and S. Ghosh, “Recovering the gravitational potential in a rotating frame: Deep Potential applied to a simulated barred galaxy,” *MNRAS* **527**, 12284–12297 (2024), arXiv:2310.00040 [astro-ph.GA].
- [16] T. Kalda and G. M. Green, “Deep Potential: Recovering the gravitational potential and local pattern speed in the solar neighborhood with GDR3 using normalizing flows,” arXiv e-prints , arXiv:2507.03742 (2025), arXiv:2507.03742 [astro-ph.GA].
- [17] Vallenari, A. *et al.* (Gaia Collaboration), “Gaia data release 3 - summary of the content and survey properties,” *A&A* **674**, A1 (2023), arXiv:2208.00211 [astro.GA].
- [18] E. Putney, D. Shih, S. H. Lim, and M. R. Buckley, “Mapping Dark Matter Through the Dust of the Milky Way Part I: Dust Correction and Phase Space Density,” (2024), arXiv:2412.14236 [astro-ph.GA].
- [19] S. H. Lim, K. A. Raman, M. R. Buckley, and D. Shih, “GalaxyFlow: upsampling hydrodynamical simulations for realistic mock stellar catalogues,” *Mon. Not. Roy. Astron. Soc.* **533**, 143–164 (2024), arXiv:2211.11765 [astro-ph.GA].
- [20] S. H. Lim, K. Hayashi, S. Horigome, S. Matsumoto, and M. M. Nojiri, “Generative adversarial network for identifying the dark matter distribution of a dwarf spheroidal galaxy,” 4th Inter-experiment Machine Learning Workshop (2020), contributed presentation.
- [21] S. H. Lim, K. Hayashi, S. Horigome, S. Matsumoto, and M. M. Nojiri, “JFlow: Model-Independent Spherical Jeans Analysis using Equivariant Continuous Normalizing Flows,” (2025), arXiv:2505.00763 [astro-ph.GA].
- [22] A. Hyvärinen, “Estimation of non-normalized statistical models by score matching,” *Journal of Machine Learning Research* **6**, 695–709 (2005).
- [23] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, “Neural ordinary differential equations,” in *Advances in Neural Information Processing Systems*, Vol. 31, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Curran Associates, Inc., 2018) arXiv:1806.07366 [cs.LG].
- [24] W. Grathwohl, R. T. Q. Chen, J. Bettencourt, and D. Duvenaud, “FFJORD: Free-form continuous dynamics for scalable reversible generative models,” in *International Conference on Learning Representations* (2019) arXiv:1810.01367 [cs.LG].
- [25] Y. Song and S. Ermon, “Generative modeling by estimating gradients of the data distribution,” *Advances in Neural Information Processing Systems (NeurIPS 2019)*, **32** (2019), arXiv:1907.05600 [cs.LG].
- [26] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, “Score-based generative modeling through stochastic differential equations,” in *International Conference on Learning Representations* (2021) arXiv:2011.13456 [cs.LG].
- [27] J. Reed, “Spherical & triaxial,” in *3rd Gaia Challenge* (2015).
- [28] M. G. Walker and J. Peñarrubia, “A Method for Measuring (Slopes of) the Mass Profiles of Dwarf Spheroidal Galaxies,” *ApJ* **742**, 20 (2011), arXiv:1108.2404 [astro-ph.CO].
- [29] J. F. Navarro, C. S. Frenk, and S. D. M. White, “Simulations of x-ray clusters,” *Mon. Not. Roy. Astron. Soc.* **275**, 720–740 (1995), arXiv:astro-ph/9408069 .
- [30] J. F. Navarro, C. S. Frenk, and S. D. M. White, “The Structure of cold dark matter halos,” *Astrophys. J.* **462**, 563–575 (1996), arXiv:astro-ph/9508025 .
- [31] J. F. Navarro, C. S. Frenk, and S. D. M. White, “A Universal density profile from hierarchical clustering,” *Astrophys. J.* **490**, 493–508 (1997), arXiv:astro-ph/9611107 .

- [32] H. C. Plummer, “On the problem of distribution in globular star clusters,” *Monthly Notices of the Royal Astronomical Society* **71**, 460–470 (1911).
- [33] J. Shao, “Consistency of jackknife variance estimators,” *Statistics: A Journal of Theoretical and Applied Statistics* **22**, 49–57 (1991).
- [34] N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. Hamprecht, Y. Bengio, and A. Courville, “On the spectral bias of neural networks,” in *Proceedings of the 36th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 97, edited by K. Chaudhuri and R. Salakhutdinov (PMLR, 2019) pp. 5301–5310, arXiv:1806.08734 [stat.ML].
- [35] D. Hendrycks and K. Gimpel, “Gaussian error linear units (GELUs),” (2016), arXiv:1606.08415 [cs.LG].
- [36] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (2015) arXiv:1502.01852 [cs.CV].
- [37] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *the 3rd International Conference for Learning Representations (ICLR 2015)* (2014) arXiv:1412.6980 [cs.LG].
- [38] G. Papamakarios, T. Pavlakou, and I. Murray, “Masked autoregressive flow for density estimation,” in *Advances in Neural Information Processing Systems*, Vol. 30, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc., 2017) arXiv:1705.07057 [stat.ML].

## A Implementation Details

For the score models  $\vec{s}(\vec{r}, \vec{v})$  and  $\vec{s}(\vec{v}|\vec{r})$  and the acceleration model  $\vec{a}(\vec{r})$ , we use multilayer perceptrons with 4 hidden layers with width 64, and with GELU activations [35]. The weights of the neural network are initialized with Kaiming uniform initialization [36], and the biases are initialized with a uniform distribution in the range [-0.001, 0.001]. For model training, we use the ADAM optimizer [37] with a learning rate 0.001.

For computing expectations  $E_{p(\vec{v}|\vec{r})}$  in Eq. 3 by sample mean, we first draw samples from a truncated multivariate Gaussian at  $|\vec{x}| < 3$  and evolve them using the Euler-Maruyama method with the score model  $\vec{s}(\vec{v}|\vec{r})$  regressing  $\frac{\partial}{\partial v_i} \log p(\vec{v}|\vec{r})$ . We use a time step of  $dt = 0.01$  with 1000 steps, which provides sufficient accuracy for our purposes. Note that the computational complexity of sample generation remains similar to the continuous normalizing flows [23, 24] used in [10, 11, 15, 19], as both methods require solving time evolutions.

## B Uncertainty Estimation

We estimate the statistical uncertainty of a prediction  $\hat{\theta}$  using a variant of the remove- $d$  jackknife estimator [33], which works as follows. Let  $\hat{\theta}_{(j)}$  be the  $j$ -th estimated result from a subsampled dataset after randomly removing  $d$  samples. The variance of the prediction  $\hat{\theta}$  is related to the variance of  $\hat{\theta}_{(j)}$  for the ensemble of subsampled datasets by a simple scaling formula:

$$\sigma_{\text{jack}}^2 = \frac{N - d}{d} \text{Var}_j [\hat{\theta}_{(j)}] \quad (11)$$

In our analysis, we randomly remove 10% of the data, i.e.,  $N = 10000$  and  $d = 1000$ , and use 20 subsampled datasets.

Bootstrap methods are also alternative approaches for quantifying uncertainty in the local dark matter density estimations [6, 14–16]; however, it is less compatible with score-based methods. As score matching is a method for estimating density derivatives, and hence, it is more sensitive to duplicated samples. In terms of the empirical distribution, these duplicated samples can be interpreted as a local bump, and the learned density derivative around there may fluctuate. As a result, the bootstrap-estimated uncertainty additionally includes variance from these density fluctuations due to duplicated samples, and the predicted uncertainty increases.

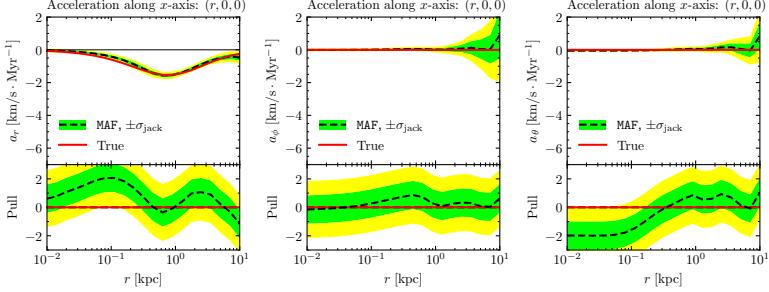


Figure 3: True (red) and MAF-estimated (black) acceleration along the  $x$ -axis for the NonplumCoreIso dataset, along with  $1\sigma$  (green) and  $2\sigma$  (yellow) statistical uncertainties. The lower panel of each figure shows the pulls, defined as the deviation from the true value divided by the uncertainty  $\sigma$ .

### C Comparison with a Flow-based Method

In this appendix, we present results using masked autoregressive flows (MAFs) [38]. For direct comparison with GalaxyScore results, we use the standardized datasets to train the flows and compare the acceleration explicitly computed using Eq. 3. This comparison reveals the precision of density derivatives between the two methods without hiding it behind acceleration or potential networks. Other architectural details are the same as those in [6, 14, 18], and training details are the same as those for score matching.

Figure 3 shows the MAF-based acceleration estimation for the NonplumCoreIso dataset. Note that directly taking the divergence of Eq. 3 is not a reliable way to estimate mass density for flows, so we do not show the mass density plot. In contrast, for score matching, directly taking the divergence gives the same results as in Figure 1, demonstrating that the second-order density derivatives are highly reliable. Comparing the acceleration plots shows that score matching provides acceleration estimates with precision comparable to flows.

Flows are also slower computationally. For acceleration evaluation at 100 locations simultaneously, score matching takes 1.5 seconds while flows take 7.3 seconds on an NVIDIA L40S GPU. Both methods use approximately 700 MB of GPU memory for this inference. However, for mass density evaluation by taking the divergence of the estimated acceleration, score matching completes the computation in 8 seconds while flows cannot perform this computation as the 48 GB GPU memory is insufficient. We conclude that score matching is much faster and more memory-efficient than flows. This computational efficiency makes GalaxyScore particularly well-suited for studying large stellar populations such as main sequence stars in the disk and galactic center of the Milky Way.