arXiv:submit/2830805 [astro-ph.CO] 4 Sep 2019

Mining for Dark Matter Substructure:
Inferring subhalo population properties from strong lenses with machine learning

Johann Brehmer,[1,2,*] Siddharth Mishra-Sharma,[1,*] Joeri Hermans,[3] Gilles Louppe,[3] and Kyle Cranmer[1,2]

[1]*Center for Cosmology and Particle Physics, Department of Physics, New York University, 726 Broadway, New York, NY 10003, USA*
[2]*Center for Data Science, New York University, 60 Fifth Ave, New York, NY 10011, USA*
[3]*Montefiore Institute, University of Liège, Belgium*

ABSTRACT

The subtle and unique imprint of dark matter substructure on extended arcs in strong lensing systems contains a wealth of information about the properties and distribution of dark matter on small scales and, consequently, about the underlying particle physics. However, teasing out this effect poses a significant challenge since the likelihood function for realistic simulations of population-level parameters is intractable. We apply recently-developed simulation-based inference techniques to the problem of substructure inference in galaxy-galaxy strong lenses. By leveraging additional information extracted from the simulator, neural networks are efficiently trained to estimate likelihood ratios associated with population-level parameters characterizing substructure. Through proof-of-principle application to simulated data, we show that these methods can provide an efficient and principled way to simultaneously analyze an ensemble of strong lenses, and can be used to mine the large sample of lensing images deliverable by near-future surveys for signatures of dark matter substructure. ♡

Contents

Corresponding authors: Siddharth Mishra-Sharma, Johann Brehmer
sm8383@nyu.edu, johann.brehmer@nyu.edu

[*] Both authors contributed equally to this work

1. INTRODUCTION

Dark matter (DM) accounts for nearly a quarter of the energy budget of the Universe, and pinning down its fundamental nature and interactions is one of the most pressing problems in cosmology and particle physics today. Despite an organized effort to do so through terrestrial (e. g., Akerib et al. 2017; Cui et al. 2017; Aprile et al. 2018), astrophysical (e. g., Albert et al. 2017; Chang et al. 2018; Lisanti et al. 2018), and collider (e. g., Sirunyan et al. 2017; Aaboud et al. 2019) searches, no conclusive evidence of interactions between the Standard Model (SM) and dark matter exists to-date.

Meanwhile, dark matter can also be studied directly through its irreducible gravitational interactions. The concordance Λ Cold Dark Matter (ΛCDM) framework of non-relativistic, collisionless dark matter particles provides an excellent description of the observed distribution of matter on large scales. However, many well-motivated models predict deviations from ΛCDM on smaller scales. Fundamental dark matter microphysical properties such as its particle mass and self-interaction cross-section can imprint themselves onto its macroscopic distribution in ways that can be probed by current and future experiments (Buckley & Peter 2018; Drlica-Wagner et al. 2019; Simon et al. 2019). As a motivating example, if dark matter has a significant free-streaming length, this would lead to its early decoupling from the cosmic plasma and an underabundance of lower-mass subhalos today (Bond & Szalay 1983; Bode et al. 2001; Dalcanton & Hogan 2001; Boyanovsky et al. 2008; Boyanovsky & Wu 2011). Dark matter self-interactions (Spergel & Steinhardt 2000; Yoshida et al. 2000; Davé et al. 2001; Colín et al. 2002; Vogelsberger et al. 2012; Peter et al. 2013; Zavala et al. 2013; Kaplinghat et al. 2014, 2016; Kamada et al. 2017; Elbert et al. 2018; Vogelsberger et al. 2019; Kahlhoefer et al. 2019; Nishikawa et al. 2019; Robles et al. 2019) and dissipative dynamics in the dark sector (Fan et al. 2013; Agrawal et al. 2017; Agrawal & Randall 2017; Buckley & DiFranzo 2018) are examples of scenarios that would modify the structure of the subhalo density profiles in addition to possibly depressing the abundance of lower-mass halos as compared to CDM predictions in the latter case (Buckley et al. 2014; Schewtschenko et al. 2015; Vogelsberger et al. 2016).

There exist several avenues for probing the distribution of dark matter on small scales. While the detection of ultrafaint dwarf galaxies through the study of stellar overdensities and kinematics (Koposov et al. 2008, 2015; Bechtol et al. 2015; Drlica-Wagner et al. 2015) can be used to make statements about the underlying dark matter properties, theoretical uncertainties in the connection between stellar and halo properties (Nadler et al. 2019; Wechsler & Tinker 2018) and the effect of baryons on the satellite galaxy population (Errani et al. 2017; Garrison-Kimmel et al. 2017; Brooks 2018; Fitts et al. 2018) pose a challenge. Furthermore, suppressed star formation in smaller halos means that there exists a threshold ($\lesssim 10^8 \, \mathrm{M_\odot}$) below which subhalos are expected to be mostly dark and devoid of baryonic activity (Efstathiou 1992; Fitts et al. 2017; Read et al. 2017). This makes studying the imprint of gravitational interactions the *only* viable avenue for probing substructure at smaller scales. In this spirit, the study of subhalo-

induced perturbations to the kinematic phase-space distribution in cold stellar streams (Johnston et al. 1999; Carlberg 2012; Carlberg & Grillmair 2013; Bonaca & Hogg 2018; Bonaca et al. 2019), and in Galactic stellar fields (Buschmann et al. 2018) have been proposed as methods to look for low-mass subhalos through their gravitational interactions in the Milky Way.

Complementary to the study of locally-induced gravitational effects, gravitational lensing has emerged as an important technique for studying the distribution of matter over a large range of scales. Locally, the use of time-domain astrometry has been proposed as a promising method to measure the distribution of local substructure through correlated, lens-induced motions of background celestial objects due to foreground subhalos (Van Tilburg et al. 2018). In the extragalactic regime, galaxy-scale strong lenses are laboratories for studying dark matter substructure. The typical substructure abundance within galaxy-scale lenses has been constrained through the measurement of positions and flux ratios of multiple images in quasar lenses (Dalal & Kochanek 2002; Hsueh et al. 2019) and lensed images of extended (Vegetti et al. 2010b,a, 2012; Hezaveh et al. 2016b) as well as quasar sources (Fadely & Keeton 2012; Nierenberg et al. 2014, 2017; Gilman et al. 2019a) have been used to set limits on the abundance of or find evidence for individual subhalo clumps with masses $\gtrsim 10^8 \, \mathrm{M_\odot}$. Although these individual high-significance detections can be used to derive constraints on substructure abundance and the subhalo mass function, searches for one (or a general fixed number of) subhalos do not take into account covariances between models with different numbers of subhalos and can leave unexpressed the degeneracies between, e.g., the imprint of several low-mass subhalos and that of a massive subhalo perturber. Additionally, these detections by definition probe the most massive subhalos in the lensing galaxies which, given the particle physics-motivated goal of constraining small-scale structure, is the less interesting regime compared to probing the fainter end of the subhalo mass function.

Another approach relies on probing the collective effect of sub-threshold (i. e., not individually resolvable) subhalos on extended arcs in strongly lensed systems. A particular challenge here is that the properties of the individual subhalos correspond to a high-dimensional space of latent variables, which must be marginalized to compute the likelihood. This complicated marginalization integral makes the likelihood for population-level parameters effectively intractable. Methods based on summary statistics (Birrer et al. 2017a) and studying the amplitude of spatial fluctuations on different

scales through a power spectrum decomposition (Heza-veh et al. 2016a; Cyr-Racine et al. 2016; Diaz Rivero et al. 2018; Chatterjee & Koopmans 2018; Díaz Rivero et al. 2018; Cyr-Racine et al. 2019; Brennan et al. 2019) have been proposed as ways to reduce the dimensionality of the problem and enable substructure inference in a tractable way. This class of methods is well-suited to studying dark matter substructure since they can be sensitive to the *population* properties of low-mass subhalos in strongly lensed galaxies which are directly correlated with the underlying dark matter particle physics.

Particularly promising in this regard are transdimensional techniques like probabilistic cataloging (Brewer et al. 2016; Daylan et al. 2018) that have been proposed to take into account covariances between models with different numbers of subhalos in a principled manner and can efficiently map out the parameter space associated with multiple sub-threshold objects in lensing systems. The output of such analyses is a ensemble of posterior-weighted subhalo catalogs which can be marginalized over to infer higher-level parameters (hyperparameters) characterizing the population properties of subhalos, potentially over multiple lensing images. These results can be highly sensitive to the assumed metamodel complexity however (Daylan et al. 2018) and potentially computationally limited for a large number of lenses as they require running an independent analysis to produce a probabilistic catalog for each image.

Current and near-future observatories like DES (Dark Energy Survey Collaboration et al. 2016), LSST (LSST Science Collaboration et al. 2009; Drlica-Wagner et al. 2019; Verma et al. 2019), and *Euclid* (Refregier et al. 2010) are expected to find hundreds to thousands of galaxy-galaxy strong lenses (Oguri & Marshall 2010; Treu 2010; Collett 2015), making substructure inference in these systems (and high-resolution followups on a subset) one of the key avenues for investigating dark matter substructure and stress-testing the Cold Dark Matter paradigm in the near future. This calls for methods that can efficiently analyze large samples of lensed images to infer the underlying substructure properties with minimal loss of information stemming from dimensional reduction.

In recent years, a large number of methods have been developed that train neural networks to estimate the likelihood function, likelihood ratio function, or posterior (Fan et al. 2012; Dinh et al. 2014; Germain et al. 2015; Jimenez Rezende & Mohamed 2015; Cranmer et al. 2015; Dinh et al. 2016; Paige & Wood 2016; Papamakarios & Murray 2016; Thomas et al. 2016; Uria et al. 2016; van den Oord et al. 2016b,c,a; Tran et al. 2017; Pa-pamakarios et al. 2017; Louppe & Cranmer 2017; Lueck-mann et al. 2017; Gutmann et al. 2017; Chen et al. 2018; Dinev & Gutmann 2018; Grathwohl et al. 2018; Huang et al. 2018; Kingma & Dhariwal 2018; Lueck-mann et al. 2018; Papamakarios et al. 2018; Alsing et al. 2019; Hermans et al. 2019). These techniques can be directly applied to population-level parameters, avoiding an additional marginalization step. In contrast to traditional simulation-based (or "likelihood-free") approaches, namely Approximate Bayesian Computation, they do not rely on summary statistics and instead learn to extract information directly from the full input data, which in our case corresponds to the observed lensed images. Finally, some of these methods let us to amortize the computational cost of the inference—after an upfront simulation and training phase, inference for any observed lensed image is efficient, enabling a simultaneous analysis of a large number of observations.

In this paper, we follow this approach and apply a particularly powerful technique for simulation-based inference introduced in Brehmer et al. (2018a,b,c); Stoye et al. (2018) to the problem of extracting high-level substructure properties from an ensemble of galaxy-galaxy strong lensing images. This method extracts additional information from the simulator, which is then used to train a neural network as a surrogate for the likelihood ratio function. The additional information increases the sample efficiency during training and thus reduces the computational cost. A calibration procedure ensures correct inference results even in the case of imperfectly trained networks. We demonstrate the feasibility of this method on a catalog of simulated lenses. After discussing the information content in individual lensed images, we switch to a simultaneous analysis of multiple observed images and calculate the expected combined constraints on population-level substructure parameters in both a frequentist and a Bayesian setup.

This paper is organized as follows. In Section 2 we briefly review the formalism of gravitational strong lensing and describe our simulation setup, including the assumptions we make about the population of background sources and host galaxies, the substructure population, and observational parameters. In Section 3 we describe the simulation-based analysis technique used and its particular application to the problem of mining dark matter substructure properties from an ensemble of extended lensed arcs. We show a proof-of-principle application to simulated data in Section 4 and comment on how this method can be extended to more realistic scenarios in Section 5. We conclude in Section 6. In the spirit of reproducibility, code associated with this paper is available on GitHub ○ and we provide links below

each figure (⟨/⟩) pointing to the *Jupyter* notebooks used to generate them.

2. STRONG LENSING FORMALISM AND SIMULATION SETUP

In strong lensing systems a background light source is gravitationally lensed by an intervening mass distribution, resulting in multiple localized images on the lens plane (in the case of a point-like quasar source) or an arc-like image (in the case of an extended galaxy source). The latter provides the ability to probe substructure over a relatively larger region on the lens plane. Additionally, young, blue galaxies are ubiquitous in the redshift regime $z \gtrsim 1$ and dominate the faint end of the galaxy luminosity function, resulting in a larger deliverable sample of galaxy-galaxy strong lenses compared to that of multiply-imaged quasars. For these reasons, we focus our method towards galaxy-galaxy lenses— systems with extended background sources producing images with lensed arcs—although the techniques presented here can also be applied to samples of lensed quasars.

We now briefly review the basic mathematical formalism behind strong gravitational lensing before describing in turn the models for the lensing galaxy, background source, and dark matter substructure assumed in this study. We also describe the mock observational parameters assumed for the image sample as well as the population properties of the host lenses. Taken together, these define our lensing forward model.

2.1. *Strong lensing formalism*

Given a mass distribution with dimensionless projected surface mass density $\kappa(\boldsymbol{\theta}) = \Sigma(\boldsymbol{\theta})/\Sigma_{\mathrm{cr}}$, where $\Sigma_{\mathrm{cr}} \equiv \frac{1}{4\pi G_{\mathrm{N}}} \frac{D_{\mathrm{s}}}{D_{\mathrm{ls}} D_{\mathrm{l}}}$ is the critical lensing surface density and D_l, D_s, and D_{ls} are the observer-lens, observer-source, and lens-source angular diameter distances respectively, the two-dimensional projected lensing potential is given by (e.g., Schneider et al. 1992; Bartelmann & Schneider 2001)

$$\psi(\boldsymbol{\theta}) = \frac{1}{\pi} \int \mathrm{d}\boldsymbol{\theta}' \ln|\boldsymbol{\theta} - \boldsymbol{\theta}'| \, \kappa(\boldsymbol{\theta}'). \tag{1}$$

The reduced deflection angle is given by the gradient of the projected lensing potential,

$$\boldsymbol{\phi}(\boldsymbol{\theta}) = \nabla\psi(\boldsymbol{\theta}) = \frac{1}{\pi} \int \mathrm{d}\boldsymbol{\theta}' \, \frac{\boldsymbol{\theta} - \boldsymbol{\theta}'}{|\boldsymbol{\theta} - \boldsymbol{\theta}'|^2} \, \kappa(\boldsymbol{\theta}') \tag{2}$$

and can be used to determine the position of the lensed source $\boldsymbol{\theta}$ through the lens equation,

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\phi}(\boldsymbol{\theta}) \tag{3}$$

where $\boldsymbol{\beta}$ is the position of the source. For an extended source profile f_s, the final lensed image f_s' can be obtained as the source light profile evaluated on the image plane (e.g., Daylan et al. 2018),

$$f_s'(\boldsymbol{\theta}) = f_s\left(\boldsymbol{\theta} - \boldsymbol{\phi}(\boldsymbol{\theta})\right). \tag{4}$$

Given a lens density profile, the deflection vector can be computed using Equation (2), and analytic expressions for many commonly considered profiles are available in the literature (e.g., Keeton 2001). The projected lensing potential and mass density are related through the Poisson equation $\nabla^2 \psi(\boldsymbol{\theta}) = 2\kappa(\boldsymbol{\theta})$, and its linearity implies that the combined projected potential due to multiple perturbers can be written as the sum of individual potentials, and the individual deflections can be superimposed as a consequence. The total deflection can then be used to calculate the lensed image for a given source profile using Equation (4). For more details on the gravitational lensing formalism see, e.g., Schneider et al. (1992); Bartelmann & Schneider (2001); Treu (2010).

2.2. *Lensing host galaxy*

Cosmological N-body simulations suggest that the dark matter distribution in structures at galactic scales can be well-described by a universal, spherically symmetric Navarro-Frenk-White (NFW) profile. However, strong lensing probes a region of the host galaxy much smaller than the typical virial radii of galaxy-scale dark matter halo, and the mass budget here is dominated by the baryonic bulge component of the galaxy. Taking this into account, the total mass budget of the lensing host galaxy, being early-type, can be well-described by a singular isothermal ellipsoid (SIE) profile. Since neither the dark matter nor the baryonic components are individually isothermal, this is sometimes known as the bulge-halo conspiracy (Treu 2010). We consider the spherical simplification of the SIE profile, the singular isothermal sphere (SIS), with the density distribution given by (Kormann et al. 1994; Treu 2010)

$$\rho_{\mathrm{SIS}}(r) = \frac{\sigma_v^2}{2\pi G_{\mathrm{N}} r^2} \tag{5}$$

where σ_v is the central 1-D velocity dispersion of the lens galaxy and q is the ellipsoid axis ratio, with $q = 1$ corresponding to the SIS profile. The Einstein radius for this profile, defining the characteristic lensing scale, is given by (Treu 2010)

$$\theta_E = 4\pi\sigma_v^2 \frac{D_{ls}(z_l, z_s)}{D_s(z_s)}, \tag{6}$$

where z_l and z_s are respectively the lens and source redshifts. We use the cosmology from Planck Collaboration

et al. (2016) to compute cosmological distances throughout this paper.

The deflection field for the SIE profile is given by (Keeton 2001)

$$\phi_x = \frac{\theta_E q}{\sqrt{1-q^2}} \tan^{-1} \left[\frac{\sqrt{1-q^2}\theta_x}{\chi} \right] \qquad (7)$$

$$\phi_y = \frac{\theta_E q}{\sqrt{1-q^2}} \tanh^{-1} \left[\frac{\sqrt{1-q^2}\theta_y}{\chi + q^2} \right] \qquad (8)$$

with $\chi \equiv \sqrt{\theta_x^2 q^2 + \theta_y^2}$ and we explicitly denote our angular coordinates as $\{\theta_x, \theta_y\}$.

Although the total galaxy mass (baryons + dark matter) describe the macro lensing field, for the purposes of describing substructure we require being able to map the measured properties of an SIE lens onto the properties of the host dark matter halo. To do this, we relate the central stellar velocity dispersion σ_v to the mass M_{200} of the host dark matter halo. Zahid et al. (2018) derived a tight correlation between σ_v and M_{200}, modeled as

$$\log \left(\frac{M_{200}}{10^{12}\,\mathrm{M_\odot}} \right) = \alpha + \beta \left(\frac{\sigma_v}{100\,\mathrm{km\,s^{-1}}} \right) \qquad (9)$$

with $\alpha = 0.09$ and $\beta = 3.48$. We model the host dark matter halo with an NFW profile (Navarro et al. 1996, 1997)

$$\rho_{\mathrm{NFW}}(r) = \frac{\rho_s}{(r/r_s)(1 + r/r_s)^2} \qquad (10)$$

where ρ_s and r_s are the scale density and scale radius, respectively. The halo virial mass M_{200} describes the total mass contained with the virial radius r_{200}, defined as the radius within which the mean density is 200 times the critical density of the universe and related to the scale radius through the concentration parameter $c_{200} \equiv r_{200}/r_s$. Thus, an NFW halo is completely described by the parameters $\{M_{200}, c_{200}\}$. We use the concentration model from Sánchez-Conde & Prada (2014) to derive the halo concentration for a given NFW virial mass.

The spherically-symmetric deflection for an NFW perturber is given by (Keeton 2001)

$$\phi_r = 4\kappa_s r_s \frac{\ln(x/2) + \mathcal{F}(x)}{x}, \qquad (11)$$

where $x \equiv r/r_s$, $\kappa_s \equiv \rho_s r_s / \Sigma_{\mathrm{cr}}$, and

$$\mathcal{F}(x) = \begin{cases} \frac{1}{\sqrt{x^2-1}} \tan^{-1}\sqrt{x^2-1} & (x > 1) \\ \frac{1}{\sqrt{1-x^2}} \tanh^{-1}\sqrt{1-x^2} & (x < 1) \\ 1 & (x = 1). \end{cases} \qquad (12)$$

We described the population parameters used to model the host velocity dispersion (and thus its Ein-

stein radius and dark matter halo mass) in Section 2.6 below.

2.3. Background source

We model the emission from background source galaxies using a Sérsic profile, with the surface brightness given by (Sérsic 1963)

$$f_s(\theta_r) = f_e \exp \left\{ -b_n \left[\left(\frac{\theta_r}{\theta_{r,e}} \right)^{1/n} - 1 \right] \right\}, \qquad (13)$$

where $\theta_{r,e}$ is the effective circular half-light radius, n is the Sérsic index, and b_n is a factor depending on n that ensures that $\theta_{r,e}$ contains half the total intensity from the source galaxy, given by (Ciotti & Bertin 1999)

$$b_n \approx 2n - \frac{1}{3} + \frac{4}{405n} + \frac{46}{25515n^2} + \frac{131}{1148175n^3} - \frac{2194697}{30690717750n^4}.$$

We assume $n = 1$ for the source galaxies, corresponding to a flattened exponential profile and consistent with expectation for blue-type galaxies at the relevant redshifts. f_e encodes the flux at half-light radius, which can be inferred from the total flux (or magnitude) associated with a given galaxy as follows. For a detector with zero-point magnitude M_0, which specifies the magnitude of a source giving 1 count s^{-1} in expectation, by definition the total counts are given by $S_{\mathrm{tot}} = 10^{0.4(M-M_0)}$. Requiring the half-light radius to contain half the expected counts, for $n = 1$ we have the relation $f_e \approx 0.526\, t_{\mathrm{exp}} S_{\mathrm{tot}}/(2\pi\theta_{r,e}^2)$ in counts arcsec^{-2}, where t_{exp} is the exposure time.

The treatment of the other Sérsic parameters, in particular the total emission and half-light radius, in the context of population studies is described in Section 2.6 below.

2.4. Lensing substructure

The ultimate goal of our method is to characterize the substructure population in strong lenses. Here we describe our procedure to model the substructure contribution to the lensing signal. Understanding the expected abundance of substructure in galaxies over a large range of epochs is a complex problem and an active ongoing area of research. Properties of individual subhalos (such as their density profiles) as well as those that describe their population (such as the mass and spatial distribution) are strongly affected by their host environment, and accurately modeling all aspects of subhalo evolution and environment is beyond the scope of this paper. Instead, we use a simplified description to model the substructure contribution in order to highlight the broad

methodological points associated with the application of our method.

Λ Cold Dark Matter (ΛCDM), often called the standard model of cosmology, predicts a scale-invariant power spectrum of primordial fluctuations and the existence of substructure over a broad range of masses with approximately equal contribution per logarithmic mass interval. We parameterize the distribution of subhalo masses m_{200} in a given host halo of mass M_{200}—the subhalo mass function—as a power law distribution with a linear dependence on the host halo mass,

$$
\frac{\mathrm{d}n}{\mathrm{d}\log\frac{m_{200}}{m_{200,0}}} = \begin{cases} \alpha \frac{M_{200}}{M_{200,0}} \left(\frac{m_{200}}{m_{200,0}}\right)^{\beta} & (m_{200}^{\min} \leq m_{200} \leq m_{200}^{\max}) \\ 0 & (\text{else}), \end{cases}
$$
(14)

where α encodes the overall substructure abundance, with larger α corresponding to more substructure, and the slope $\beta < 0$ encodes the relative contribution of subhalos at different masses, with more negative β corresponding to a steeper slope with more low-mass subhalos. $m_{200,0}$ and $M_{200,0}$ are arbitrary normalization factors.

Theory and simulations within the framework of ΛCDM predict a slope $\beta \approx -0.9$ (Madau et al. 2008; Springel et al. 2008), resulting in a nearly scale-invariant spectrum of subhalos, which we assume in our fiducial setup. We parameterize the overall subhalo abundance α through the mass fraction within the lensing galaxies contained in subhalos, f_{sub}, defined as the fraction of the total dark matter halo mass contained in bound substructure in a given mass range:

$$
f_{\mathrm{sub}} = \frac{\int_{m_{200,\min}}^{m_{200,\max}} \mathrm{d}m_{200}\, m_{200}\, \frac{\mathrm{d}n}{\mathrm{d}m_{200}}}{M_{200}}.
$$
(15)

For a given $\{f_{\mathrm{sub}}, \beta\}$ and host halo mass M_{200}, this can be used to determine α in Equation (14). The linear scaling of the subhalo mass function with the host halo mass M_{200} in Equation (14) is additionally described in Han et al. (2016); Despali & Vegetti (2017). In our fiducial setups, we take the minimum and maximum subhalo masses to be $m_{200,\min} = 10^6\,\mathrm{M}_\odot$ and $m_{200,\max} = 0.01\,M_{200}$ (Despali & Vegetti 2017; Hiroshima et al. 2018) respectively, and corresponding fiducial substructure mass fraction in this range of 5%, roughly consistent with observations in Dalal & Kochanek (2002); Hiroshima et al. (2018); Hsueh et al. (2019).

With all parameters of the subhalo mass function specified, the total number of subhalos $\bar{n}_{\mathrm{tot}}$ expected within the virial radius R_{200} of the host halo can be inferred as $\int_{m_{200,\min}}^{m_{200,\max}} \mathrm{d}m_{200} \frac{\mathrm{d}n}{\mathrm{d}m_{200}}$. Strong lensing probes a region much smaller than this scale—the typical Einstein radii for the host deflector are much smaller than

the virial radius of the host dark matter halos. In order to obtain the expected number of subhalos within the lensing observation's region of interest (ROI), we scale the total number of subhalos obtained from the above procedure by the ratio of projected mass within our region of interest θ_{ROI} and the host halo mass M_{200} as follows. We assume the subhalos to be distributed in number density following the host NFW dark matter profile. In this case, the enclosed mass function is $M_{\mathrm{enc}}(x) = M_{200}\left[\ln(x/2) + \mathcal{F}(x)\right]$ (e.g., Keeton 2001), where x is the angular radius in units of the scale radius, $x \equiv \theta/\theta_s$ and $\mathcal{F}(x)$ is given by Equation (12) above. The expected number of subhalos within our ROI is thus obtained as $\bar{n}_{\mathrm{ROI}} = \bar{n}_{\mathrm{tot}}\left[\ln(x_{\mathrm{ROI}}/2) + \mathcal{F}(x_{\mathrm{ROI}})\right]$. We conservatively take the lensing ROI to enclose a region of angular size twice the Einstein radius of the host halo, $\theta_{\mathrm{ROI}} = 2 \cdot \theta_E$.

Since strong lensing probes the line-of-sight distribution of subhalos within the host, their projected spatial distribution is approximately uniform within the lensing ROI (Despali & Vegetti 2017). We thus distribute subhalos uniformly within our ROI. The density profile of subhalos is assumed to be NFW and given by Equation (10), with associated lensing properties as described and the concentration inferred using the model in Sánchez-Conde & Prada (2014).

We finally emphasize that we do not intent to capture all of the intricacies of the subhalo distribution, such as the effects of baryonic physics, tidal disruption of subhalos in proximity to the center of the host and redshift evolution of host as well as substructure properties. Although our description can be extended to take these effects into account (see Section 5), their precise characterization is still subject to large uncertainties, and our simple model above captures the essential physics for demonstration purposes.

2.5. Observational considerations

A noted above, our method is best-suited to analyzing a statistical sample of strong lenses, such as those that are expected to be obtained in the near future with optical telescopes like *Euclid* and LSST, to quantify the effect of substructure. Given the challenges associated with the precise characterization of such a sample at the present time, we describe here the observational characteristics we assume in order to build up training and testing samples to validate our inference techniques.

We largely follow the description of Collett (2015) and use the associated *LensPop* package to characterize our mock observations. In particular, we use the nominal detector configuration for *Euclid*, assuming a zero-point magnitude $m_{\mathrm{AB}} = 25.5$ in the single opti-

cal VIS passband, a 64×64 pixel grid with pixel size 0.1 arcsec, a Gaussian point spread function (PSF) with FWHM 0.18 arcsec, individual exposures with exposure time 1610 s, and an isotropic sky background with magnitude $22.8 \, \mathrm{arcsec}^{-2}$ in the detector passband.

These properties, in particular the exposure, sky background, and PSF shape, are expected to vary somewhat across the lens sample. Additionally, a given region may be imaged by multiple exposures over a range of color bands. Although such variations can be incorporated into our analysis, modeling these features is beyond the scope of this study. We comment on these extensions in Section 5.

2.6. *Population properties of the lens and source samples*

The fact that the strong lens population is expected to be dominated by higher-redshift ($z_s \gtrsim 1$) blue source galaxies lensed by intermediate-redshift ($z_l \sim 0.5$–1) elliptical galaxies presents significant challenges for quantifying the lens population obtainable with future observations. Specifically, planned ground-based surveys like LSST and space telescopes like *Euclid* present complementary challenges for delivering images of strong lensing systems suitable for substructure studies. LSST is expected to image in six bands, allowing for efficient separation between source and lens emission, but at the cost of lower resolution by virtue of being a ground-based instrument. *Euclid* imaging is expected be higher in resolution but with a single optical passband (VIS). Near-IR imaging from WFIRST may deliver a high-resolution, multi-wavelength dataset that is more suitable for substructure studies, although potentially with different lens and source samples from those deliverable by optical telescopes.

In light of these uncertainties, we confine ourselves to a setting where the main methodological points can be made without detailed modeling of the detector capabilities and the deliverable lensing dataset, which is outside of the scope of the current paper. For concreteness, we simulate a sample of lenses with a simplified subset of host galaxy properties consistent with those deliverable by *Euclid* as modeled by Collett (2015). In particular, we assume spherical lenses, with ellipticity parameter $q = 1$ in Equation (5). We draw the central 1-D velocity dispersions σ_v of host galaxies from a normal distribution with mean $225 \, \mathrm{km \, s}^{-1}$ and standard deviation $50 \, \mathrm{km \, s}^{-1}$. Following Zahid et al. (2018), Equation (9) is used to map the drawn σ_v to a dark matter halo mass M_{200}, and the host Einstein radius is analytically inferred with Equation (6).

We draw the lens redshifts z_l from a log-normal distribution with mean 0.56 and scatter 0.25 dex, discarding lenses with $z_l > 1$ as these tend to have a small angular size over which substructure perturbations are relevant. The source redshift is fixed at $z_s = 1.5$, its offsets $\Delta\theta_x$ and $\Delta\theta_y$ are drawn from a normal distribution with zero mean and standard deviation 0.2. These choices are consistent with the lens sample generated from the *LensPop* code packaged with Collett (2015). We show a sample of simulated lensed images with these settings in Figure 1.

3. STATISTICAL FORMALISM AND SIMULATION-BASED INFERENCE

Our goal is to infer the subhalo mass function parameters from a catalog of images of observed lenses. In this section we will describe the challenges of this inference problem and our approach of simulation-based inference. For simplicity, we will use a more abstract notation, distinguishing between three sets of quantities in the lensing system:

Parameters of interest ϑ**:** The vector $\vartheta = (f_{\mathrm{sub}}, \beta)^T$ parameterizes the subhalo mass function given, and our goal is to infer their values.

Latent variables z**:** A vector of all other unobservable random variables in the simulator. These include the mass M_{200}, source-host offset ($\Delta\theta_x, \Delta\theta_y$), and redshift z_l of the lens, the number of subhalos in the region of interest n_{ROI}, the position $\mathbf{r}$ and mass m_{200} of each subhalo, and the random variables related to the point spread function and Poisson fluctuations.

Observables x**:** The observed lens images.

Unfortunately, the same symbols are used with different meanings in astrophysics and statistics: note the difference between the parameters ϑ and the angular positions θ_x, θ_y and the Einstein radius θ_E; between the latent variables z and the redshifts z_s, z_l; and between the observed image x and the argument of the NFW profile $M_{\mathrm{enc}}(x)$ and $\mathcal{F}(x)$ used in the last section.

As described above, we have implemented a simulator for the lensing process in the "forward" direction: for given parameters ϑ, the simulator samples latent variables z and finally observed images $x \sim p(x|\vartheta)$. Here $p(x|\vartheta)$ is the probability density or likelihood function of observing a lens image x given parameters ϑ. It can be schematically written as

$$p(x|\vartheta) = \int \mathrm{d}z \, p(x, z|\vartheta), \qquad (16)$$

where we integrate over the latent variables z and $p(x, z|\vartheta)$ is the joint likelihood of observables and latent
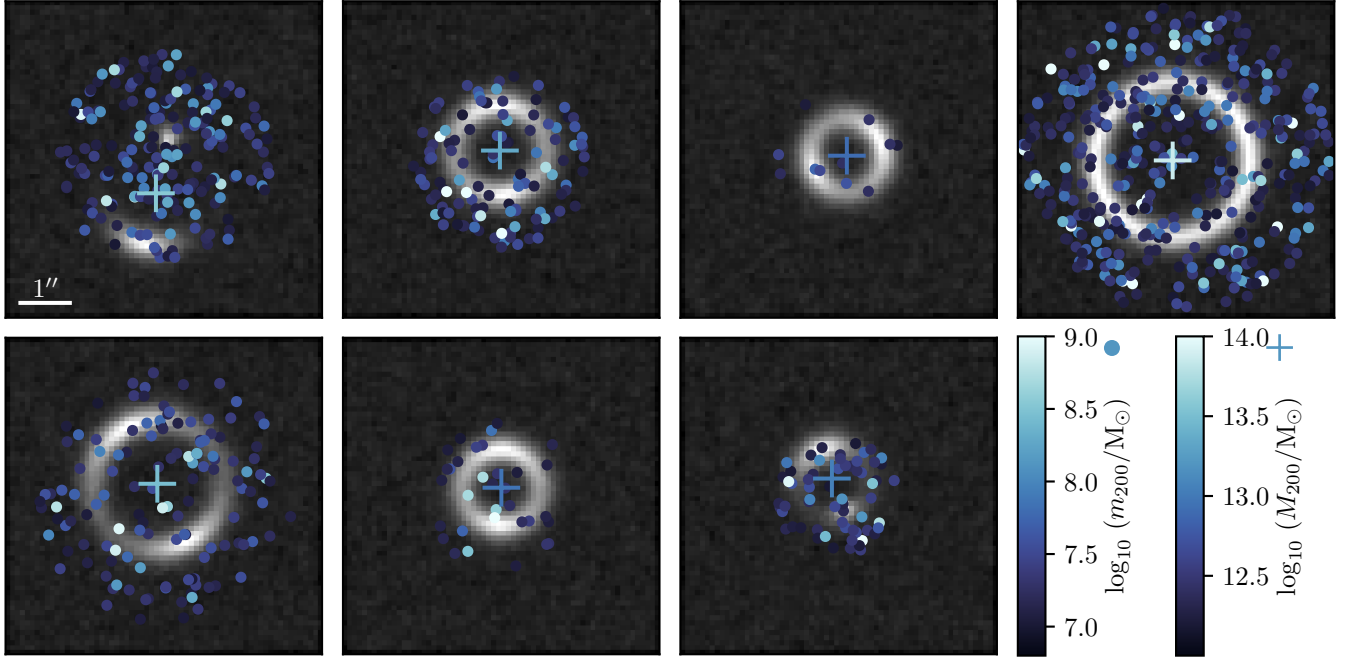
Figure 1. A sample of simulated lenses. The cross markers show the offset between the host and source centers, the color corresponding to the virial mass of the host dark matter halo. The simulated subhalos are shown as dots, the color again indicates their masses. The greyscale images show the corresponding observed images. We show seven images randomly generated for $f_{\rm sub} = 0.05$ and $\beta = -0.9$. 🔗

variables:

$$p(x, z|\vartheta) = p_{\rm host}(M_{200}, \Delta\theta_x, \Delta\theta_y, z_l)$$

$$\times \, {\rm Pois}(n_{\rm ROI}|\overline{n}_{\rm ROI}(\vartheta)) \prod_i^{n_{\rm ROI}} \left[p_m(m_{200,i}|\vartheta) \; {\rm Uniform}\,(\mathbf{r}_i) \right]$$

$$\times \, p_{\rm obs}(x|f(M_{200}, \Delta\theta_x, \Delta\theta_y, z_l; \{(m_{200,i}, \mathbf{r}_i)\})) . \quad (17)$$

Here $p_{\rm host}(M_{200}, \Delta\theta_x, \Delta\theta_y, z_l)$ is the distribution of the host halo parameters; $\overline{n}_{\rm ROI}(\vartheta)$ is the mean number of subhalos in the region of interest as a function of the parameters $\vartheta = (f_{\rm sub}, \beta)^T$, while $n_{\rm ROI}$ is the actually realized number in the simulation; $m_{200,i}$ and $\mathbf{r}_i$ are the subhalo masses and positions; $p_m(m|\vartheta) = 1/n \, {\rm d}n/{\rm d}m_{200}$ is the normalized subhalo mass function given in Equation (14); and in the last line $p_{\rm obs}$ is the probability of observing an image x based on the true lensed image $f(z_l, \{(m_{200,i}, r_i)\})$ taking into account Poisson fluctuations and detector response through the point spread function.

Standard frequentist and Bayesian inference methods rely on evaluating the likelihood function $p(x|\vartheta)$. Unfortunately, even in our somewhat simplified simulator each run of the simulation easily involves hundreds to thousands of latent variables, the integral in Equation (16) over this enormous space clearly cannot be computed explicitly. The likelihood function $p(x|\vartheta)$ is thus intractable, providing a major challenge for both frequen-

tist and Bayesian inference. Similarly, inference with Markov Chain Monte Carlo (MCMC) methods based directly on the joint likelihood function $p(x, z|\vartheta)$ requires unfeasibly many samples before converging because the latent space is so large. Systems defined through a forward simulator that does not admit a tractable likelihood are known as "implicit models", inference techniques for this case as "simulation-based inference" or "likelihood-free inference".

One way to tackle this issue is to estimate the density for observables x from samples from the simulator, where the latent variables z are marginalized by the sampling procedure. But traditional density estimation techniques require reducing the dimensionality of x with summary statistics $v(x)$, for instance based on power spectra (Hezaveh et al. 2016a; Cyr-Racine et al. 2016; Diaz Rivero et al. 2018; Chatterjee & Koopmans 2018; Díaz Rivero et al. 2018; Cyr-Racine et al. 2019; Brennan et al. 2019). The likelihood $p(v|\vartheta)$ in the space of summary statistics can either be explicitly estimated through density estimation techniques such as histograms, kernel density estimation, or Gaussian processes, or replaced by a rejection probability in an Approximate Bayesian Computation (ABC) technique (Rubin 1984). Substructure inference in quasar and extended-arc lenses using ABC techniques was explored in Gilman et al. (2018) and Birrer et al. (2017a),

respectively. While the compression to summary statistics makes the analysis tractable, it typically loses information and hence reduces the statistical power of the analysis.

Instead, the likelihood function or density can be approximated without any compression to summary statistics with a neural network, which has to be trained only once and can be evaluated efficiently for any parameter point and observed image. Similarly, one can train a neural network to estimate the likelihood *ratio* for a fixed observation x between two different hypotheses or parameter points. We will show how this turns the intractable integral in Equation (16) into a tractable minimization problem and amortizes the marginalization over z. This approach scales well to the expected large number of lenses expected in upcoming surveys (Oguri & Marshall 2010; Treu 2010; Collett 2015). Since the full image is used as input, no information is lost due to dimensionality reduction.

We use a simulation-based inference technique introduced in Brehmer et al. (2018a,b,c) that extracts additional information from the simulation and uses it to improve the sample efficiency of the training of the neural network. Our inference strategy consists of four steps:

1. During each run of the simulator, additional information that characterizes the subhalo population and lensing process is stored together with the simulated observed image.

2. This information is used to train a neural network to approximate the likelihood ratio function.

3. The neural network output is calibrated, ensuring that errors during training do not lead to incorrect inference results.

4. The calibrated network output is then used in either a frequentist or Bayesian setting to perform inference.

In the remainder of this section, we will explain these four steps in detail.

3.1. *Extracting additional information from the simulator*

In a first step, we generate training data by simulating a large number of observed lenses. For each lens, we first draw two parameter points from a proposal distribution, $\vartheta, \vartheta' \sim \pi(\vartheta)$. This proposal distribution should cover the region of interest in the parameter space, but does not have to be identical to the prior in a Bayesian inference setting, which allows us to be agnostic about the inference setup at this stage. Note that we use the

term "proposal distribution" to avoid confusion with the prior, even though it is not a proposal distribution in the MCMC sense.

Next, the simulator is run for the parameter point ϑ, generating an observed image $x \sim p(x|\vartheta)$. In addition, we calculate and save two quantities: the joint likelihood ratio

$$r(x, z|\vartheta) = \frac{p(x, z|\vartheta)}{p_{\rm ref}(x, z)} \qquad (18)$$

and the joint score

$$t(x, z|\vartheta) = \nabla_\vartheta \log p(x, z|\vartheta). \qquad (19)$$

The joint likelihood ratio quantifies how much more or less likely a particular simulation chain including the latent variables z is for the parameter point ϑ compared to a reference distribution

$$p_{\rm ref}(x, z) = \int {\rm d}\vartheta' \, \pi(\vartheta') \, p(x, z|\vartheta'), \qquad (20)$$

where we choose the marginal distribution of latent variables and observables corresponding to the proposal distribution $\pi(\vartheta)$. Unlike the distribution for a single reference parameter point, this marginal model has support for every potential outcome of the simulation (Hermans et al. 2019). The joint score is the gradient of the joint log likelihood in model parameter space and quantifies if a particular simulation chain becomes more or less likely under infinitesimal changes of the parameters of interest. Both quantities depend on the latent variables of the simulation chain.

We compute the joint likelihood ratio and joint score with Equation (17). Conveniently, the first and third line of that equation do not explicitly depend on the parameters of interest ϑ and cancel in the joint likelihood ratio and joint score; the remaining terms can be evaluated with little overhead to the simulation code. We also calculate the joint likelihood ratio $r(x, z|\vartheta')$ and the joint score $t(x, z|\vartheta')$ for the second parameter point ϑ' and store the parameter points ϑ and ϑ', the simulated image x, as well as the joint likelihood ratios and joint scores.

Our training samples consist of 10^6 images, with parameter points chosen from a uniform range in $0.001 < f_{\rm sub} < 0.2$ and $-1.5 < \beta < -0.5$.

3.2. *Machine learning*

How are the joint likelihood ratio and joint score, which are dependent on the latent variables z, useful for inference based on the likelihood function $p(x|\vartheta)$, which only depends on the observed lens images and the pa-

rameters of interest? Consider the functional

$$L[g(x,\vartheta)] = \int d\vartheta \int d\vartheta' \int dx \int dz \; \pi(\vartheta) \; \pi(\vartheta') \; p(x,z|\vartheta)$$

$$\times \Bigg[-s \log g - (1-s) \log(1-g) - s' \log g' - (1-s') \log(1-g')$$

$$+ \alpha \Big\{ \Big| t - \nabla_\vartheta \log \tfrac{1-g}{g} \Big|_\vartheta \Big|^2 + \Big| t' - \nabla_\vartheta \log \tfrac{1-g}{g} \Big|_{\vartheta'} \Big|^2 \Big\} \Bigg],$$

$$\tag{21}$$

where we abbreviate $s \equiv s(x,z|\vartheta) \equiv 1/(1 + r(x,z|\vartheta))$, $s' \equiv s(x,z|\vartheta') \equiv 1/(1 + r(x,z|\vartheta'))$, $g \equiv g(x,\vartheta)$, $g' \equiv g(x,\vartheta')$, $t = t(x,z|\vartheta)$, and $t' \equiv t(x,z|\vartheta')$ for readability. Note that the test function $g(x,\vartheta)$ is a function of x and ϑ only. The first two lines of Equation (21) are an improved version of the cross-entropy loss, in which the joint likelihood ratio is used to decrease the variance compared to the canonical cross-entropy (Stoye et al. 2018). The last line adds gradient information, weighted by a hyperparameter α.

As shown in Stoye et al. (2018), this "ALICES" loss functional is minimized by the function

$$g^*(x,\vartheta) \equiv \underset{g}{\arg\min} \, L[g(x,\vartheta)] = \frac{1}{1 + r(x|\vartheta)}, \tag{22}$$

one-to-one with the likelihood ratio function

$$r(x|\vartheta) \equiv \frac{p(x|\vartheta)}{p_{\mathrm{ref}}(x)} = \frac{1 - g^*(x,\vartheta)}{g^*(x,\vartheta)}. \tag{23}$$

We demonstrate the minimization of this functional explicitly in Appendix A. This means that if we can construct the functional in Equation (21) with the joint likelihood ratio and joint score extracted from the simulator and numerically minimize it, the resulting function lets us reconstruct the (otherwise intractable) likelihood ratio function $r(x|\vartheta)$! Essentially, this step lets us integrate out the dependence on latent variables z from the joint likelihood ratio and score, but in a general, functional form that does not depend on a set of observed images.

This is why extraction of the joint likelihood ratio and joint score has been described with the analogy of "mining gold" from the simulator (Brehmer et al. 2018c)—while calculating these quantities may require some effort and changes to the simulator code, through the minimization of a suitable functional they allow us to calculate the otherwise intractable likelihood ratio function.

In practice, we implement this minimization with machine learning. A neural network plays the role of the test function $g(x,\vartheta)$, the integrals in Equation (21) are approximated with a sum over training data sampled according to $\pi(\vartheta)\pi(\vartheta')p(x,z|\vartheta)$, and we minimize the loss

numerically through a stochastic gradient descent algorithm. The neural network trained in this way provides an estimator $\hat{r}(x|\vartheta)$ of the likelihood ratio function that is exact in the limit of infinite training samples, sufficient network capacity, and efficient minimization. Note the "parameterized" structure of the network, in which a single neural network is trained to estimate the likelihood ratio over all of the parameter space, with the tested parameter point ϑ being an input to the network (Cranmer et al. 2015; Baldi et al. 2016). This approach is more efficient than a point-by-point analysis of a grid of parameter points: it allows the network to "borrow" information from neighboring parameter points, benefiting from the typically smooth structure of the parameter space.

Given the image nature of the lensing data, we choose a convolutional network architecture based on the ResNet-18 (He et al. 2016) implementation in *PyTorch* (Paszke et al. 2017). The parameters ϑ enter as additional inputs in the fully connected layers of the network. Compared to the original ResNet-18 architecture, we add another fully connected layer at the end to ensure that the relation between parameters of interest and image data can be modeled. All inputs are normalized to zero mean and unit variance. We train the networks by minimizing the loss in Equation (21) with $\alpha = 2 \cdot 10^{-3}$ over 100 epochs with a batch size of 128 using stochastic gradient descent with momentum (Qian 1999), exponentially decaying the learning rate from 0.01 to 0.0001 with early stopping. We pretrain the model on data generated from a simplified version of the simulator, namely the "fix" scenario described in Appendix B. This architecture and hyperparameter configuration performed best during a rough hyperparameter scan, though for this proof-of-concept study we have not performed an exhaustive optimization.

3.3. *Calibration*

In reality, the neural network might not learn the likelihood ratio function $r(x|\vartheta)$ exactly, for instance due to limited training data or inefficient training. To make sure that our inference results are correct even in this case, we calibrate the network output with histograms (Cranmer et al. 2015; Brehmer et al. 2018b). For every parameter point ϑ that we want to test, we simulate a set of images $\{x\} \sim p(x|\vartheta)$ from this parameter point and calculate the network prediction $\hat{r} \equiv \hat{r}(x|\vartheta)$ for each image. We also simulate a set of images $\{x\} \sim p_{\mathrm{ref}}(x)$ from the reference model, again calculating the network prediction $\hat{r}$ for each lens. The calibrated likelihood ratio is then calculated from his-

tograms of the network predictions as

$$\hat{r}_{\rm cal}(x|\vartheta) = \frac{\hat{p}(\hat{r}|\vartheta)}{\hat{p}_{\rm ref}(\hat{r})} \qquad (24)$$

where the $\hat{p}(\cdot)$ denote probability densities estimated with univariate histograms.

This additional calibration stage comes with a certain computational cost that increases linearly with the number of evaluated parameter points. However, it guarantees that as long as the simulator accurately models the process, the inference results will be perfect or conservative, but not too optimistic, even if the neural network output is substantially different from the true likelihood ratio.

We will show results both without and with calibration. Where calibration is used, it is based on histograms with 50 bins, with bin boundaries determined automatically to match the distribution of likelihood ratios.

3.4. *Inference*

After a neural network has been trained (and optionally calibrated) to estimate the likelihood ratio function, it provides the basic ingredient to both frequentist and Bayesian inference. Multiple observations can be combined in a straightforward way: since all lens images are assumed as identically distributed and independent (except for the common dependence on the population-level parameters), the combined likelihood of a set of images is given by the product of likelihood ratios for each individual lens,

$$p_{\rm combined}(\{x\}|\theta) = \prod_i p(x_i|\theta). \qquad (25)$$

For frequentist hypothesis tests, the most powerful test statistic to distinguish two parameter points θ_0 and θ_1 is the likelihood ratio (Neyman & Pearson 1933)

$$\frac{p_{\rm combined}(\{x\}|\theta_0)}{p_{\rm combined}(\{x\}|\theta_1)} = \prod_i \frac{r(x_i|\theta_0)}{r(x_i|\theta_1)} \approx \prod_i \frac{\hat{r}(x_i|\theta_0)}{\hat{r}(x_i|\theta_1)}, \quad (26)$$

where in the last step we have replaced the exact likelihood ratio with the estimation from the (calibrated) neural network. In addition, the asymptotic properties of the likelihood ratio allow us in many cases to directly translate a value of the likelihood ratio into a p-value and thus into exclusion limits at a given confidence level (Wilks 1938; Wald 1943; Cowan et al. 2011).

For Bayesian inference, note that we can write Bayes' theorem as

$$p(\vartheta|\{x_i\}) = \frac{p(\vartheta) \prod_i p(x_i|\vartheta)}{\int {\rm d}\vartheta' \, p(\vartheta') \prod_i p(x_i|\vartheta')}$$

$$= p(\vartheta) \left[\int {\rm d}\vartheta' \, p(\vartheta') \prod_i \frac{p(x_i|\vartheta')}{p(x_i|\vartheta)} \right]^{-1}$$

$$\approx p(\vartheta) \left[\int {\rm d}\vartheta' \, p(\vartheta') \prod_i \frac{\hat{r}(x_i|\vartheta')}{\hat{r}(x_i|\vartheta)} \right]^{-1}, \quad (27)$$

where $\{x_i\}$ is the set of observed lens images and $p(\vartheta)$ is the prior on the parameters of interest, which may be different from the proposal distribution $\pi(\vartheta)$ used during the generation of training data. The posterior can thus be directly calculated given an estimator $\hat{r}$, provided that the space of the parameters of interest is low-dimensional enough to calculate the integral, or with MCMC (Hermans et al. 2019) or variational inference techniques otherwise.

While our approach to inference is strongly based on the ideas in Brehmer et al. (2018a,b,c); Stoye et al. (2018), there are some novel features in our analysis that we would like to highlight briefly. Unlike in those earlier papers, we use a marginal model based on the proposal distribution $\pi(\vartheta)$ as reference model in the denominator of the likelihood ratio, which substantially improves the numerical stability of the algorithm. This choice also allows us to include the "flipped" terms with s' and g' in the loss function in Equation (21); we found that this new, improved version of the ALICES loss improves the sample efficiency of our algorithms. Both of these improvements are inspired by Hermans et al. (2019). Finally, this is the first application of the "gold mining" idea to image data, the first combination with a convolutional network architecture, and the first use for Bayesian inference. Although machine learning-based methods have previously been proposed for inferring strong lensing host parameters (Hezaveh et al. 2017; Perreault Levasseur et al. 2017; Morningstar et al. 2018) and for lensed source reconstruction (Morningstar et al. 2019), this paper represents the first proposed application of machine learning for dark matter substructure inference in strong lenses and, as far as we are aware, for substructure inference in general.

4. RESULTS

After training the neural network using the simulations described in Section 2 and the formalism described in Section 3, we can run the inference step on a given set of images to extract the likelihood ratio estimates $\hat{r}(x|\vartheta)$ associated with the substructure parameters of
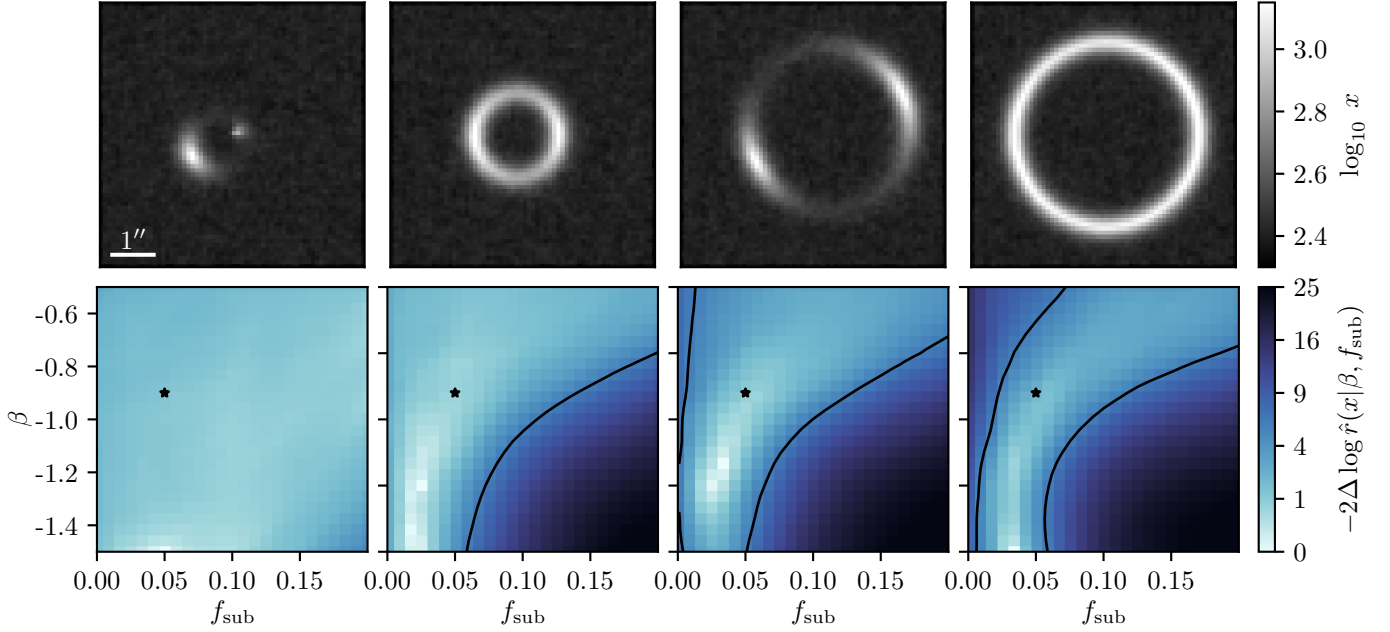
Figure 2. Four simulated lens images (upper panels) and the corresponding likelihood ratio maps estimated by the network (lower panels, without calibration). The star marks the true point used to generate the images, the black line shows 95% CL contours in parameter space based on each image.

interest $\{f_{\mathrm{sub}}, \beta\}$. We start by illustrating in Figure 2 inference on individual simulated lensed images realizing substructure corresponding to benchmark parameters $\beta = -0.9$ and $f_{\mathrm{sub}} = 0.05$. The top row shows example simulated images, with the corresponding inferred 2-D likelihood surfaces shown in the bottom row. The true parameter point is marked with a star and the 95% confidence level (CL) contours are shown.

Several interesting features can already be seen in these results. The 95% CL contours contain the true parameter point, with the overall likelihood surface being strongly correlated with the corresponding image. A smaller projected surface area of the lensed arc, resulting from a smaller host halo or a larger offset between the host and source centers, generally results in a flatter likelihood surface. This is expected, since a smaller host galaxy will contain relatively less substructure, and a smaller host or larger relative offset will result in a smaller effective arc area over which the substructure can imprint itself. The first column of Figure 2 shows an example of such a system. In contrast, the last columns show a system with a relatively massive host and a small offset, producing a symmetric image with a larger effective arc surface area over which the effects of substructure can be discerned. This results in a "peakier" inferred likelihood surface, corresponding to a higher sensitivity to f_{sub} and β. The second and third columns of Figure 2 correspond to systems with a small, centered

and a large, offset halo respectively, and show intermediate sensitivity to substructure properties.

In the spirit of stacking multiple observations, we next consider a simultaneous analysis of multiple lensed images. As discussed in Section 3.4, the product of the likelihood maps of the individual images defines the appropriate test statistic. For the purpose of population-level inference, these two-dimensional likelihood maps are hence a good alternative way to define a probabilistic catalog over individual observations, avoiding the complications of prior dependence and of communicating a complicated trans-dimensional posterior. In the left panel of Figure 3, we show the expected log likelihood ratio surface per-image in the asymptotic limit, with the 1-D slice corresponding to $\beta = -0.9$ shown in the right panel. The 95% CL expected exclusion limits for 5, 20, and 100 lenses are shown using the dotted, dashed, and solid lines respectively. The procedure can easily be extended to an arbitrarily large collection of lenses.

We find that, at least within the simplifying assumptions of our simulator, an analysis of a few tens of lenses is already sensitive to the overall substructure abundance parameterized by f_{sub}. A larger observed lens sample provides a tighter constraint on substructure properties. Approximately 100 lens images are required to begin resolving β. The expected exclusion contours are centered around the true values, confirming that our inference methods yield an unbiased estimate of the underlying substructure properties. Note
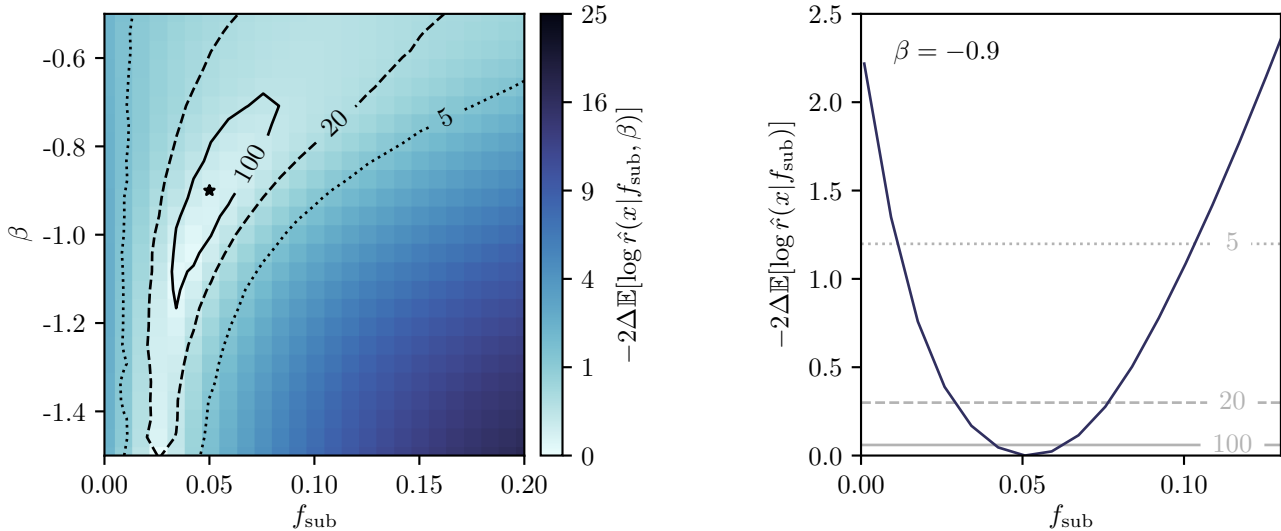
Figure 3. The expected per-lens likelihood ratio map assuming $\beta = -0.9$ and $f_{\mathrm{sub}} = 0.05$ in the two-dimensional parameter space (left) and along a one-dimensional slice at $\beta = -0.9$ (right). The lines show expected 95% CL exclusion limits for 5 (dotted), 20 (dashed), and 100 (solid) observed lenses. While the colormap shows the network output without calibration, the lines include the calibration procedure described in Section 3.3.

the "banana" shape of the expected exclusion limits, which approximately traces the total deflection contributed by substructure. We demonstrate this in Figure 4, where we show a proxy for the total subhalo-induced deflection, $\sum_{\mathrm{subhalos}} 4\kappa_s r_s$, equal to the space-independent part of Equation (11), and compare it to the expected exclusion limits. In our particular substructure scenario, this proxy can be shown to approximately scale like $\sum_{\mathrm{subhalos}} m_s^{2/3}$. We note that this comparison is schematic, as the subtle effects of substructure over a wide range of masses cannot be quantified through a single number (here, the total deflection).

With the likelihood ratio in hand, Equation (27) easily admits a Bayesian interpretation. In the left panel of Figure 5 we show the posterior for 100 lenses derived from the expected likelihood ratio results, assuming a Gaussian prior with mean -0.9 and standard deviation 0.1 on the slope β. This choice is intended to capture a prior expectation on the subhalo mass function slope consistent with the Cold Dark Matter scenario (e.g., Madau et al. 2008; Springel et al. 2008). As expected from the likelihood maps, we find a posterior density peaked around the true point.

The corresponding inferred subhalo mass function (SHMF) per host halo mass, marginalized over the host halo properties, is shown in the right panel of Figure 5. We show the point-wise mean (solid line) and point-wise 68 / 95% credible intervals (cyan and blue bands). A comparison with the true simulated subhalo mass function (dotted line, also marginalized over the host halo properties) shows excellent agreement.

5. EXTENSIONS

For the proof-of-concept analysis presented here our lensing simulation makes a number of simplifying assumptions in order to highlight the broad methodological points in a computationally tractable setting. An application of our method to real lensing data will invariably require modifications to our simulation and inference pipelines to account for the vast physical diversity in host and source galaxy morphologies, as well as ways to deal with more realistic detector response. Modeling substructure in a more involved setting than presented here (e.g., to account for tidal evolution and/or suppression of small-scale structure), and accounting for substructure along the line of sight is also desired. We will now discuss these features and comment on how they might affect our pipeline and the results presented here, leaving implementation and application to real lensing data to future work.

First, we currently fix all properties of the background source as described in Section 2.3. It is straightforward to instead draw and marginalize over the parameters associated with a chosen parameterization for the source light distribution, with Gaussian and Sérsic (Sérsic 1963) profile models being common choices. For high-fidelity images (e.g., those obtainable by targeted followups or interferometric imaging) more complicated features in the background galaxies such as outflows may not be adequately captured by such a parameterization and could introduce degeneracies with the effects of substructure. Alternative parameterizations using shapelet basis sets (Birrer et al. 2015; Tagore & Jackson 2016;
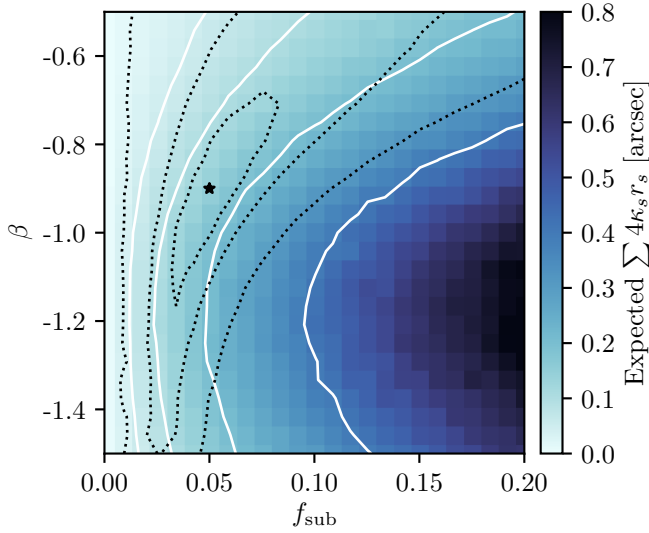
Figure 4. Expected proxy for the total subhalo-induced deflection (see Equation (11)) as a function of $f_{\rm sub}$ and β. The solid white lines show contours of constant deflection, while the dotted black lines show the expected exclusion limits from the left panel of Figure 3. ⟨⋔⟩

Birrer & Amara 2018), and methods based on regularized linear inversion on grids (Warren & Dye 2003; Suyu et al. 2006; Tagore & Keeton 2014; Nightingale et al. 2018) have been introduced as ways to model more complicated source features. For our purposes, generative / data-driven modeling of background galaxies could easily be interfaced with our pipeline to account for the variation in structure of the background sources (Morningstar et al. 2019).

Similarly, the host lens (and associated host dark matter halo) model can be made more realistic by relaxing the restriction to spherical host halos and including more complicated profiles than the Singular Isothermal Sphere considered here, drawing and marginalizing over additional host parameters as required. External shear, which models the fact that the local large-scale structure environment of the host galaxy can induce an additional overall deflection field in a preferred direction, can similarly be parameterized (e.g., Keeton et al. 1997; Schneider 1997) and marginalized over.

A realistic simulator should also model the dynamical evolution of subhalos (Despali & Vegetti 2017). Effects associated with tidal disruption due to the large gradient of the galactic potential towards the center of the host galaxy are expected to deplete the fraction of mass bound in substructures there, leading to a depressed overall subhalo abundance (Han et al. 2016) with profile properties (e.g., concentration (Moliné et al. 2017) and a truncation radius (Baltz et al. 2009)) that depend on the distance from the host center. This could

easily be implemented within our framework by drawing 3-D positions for the subhalos from the host center and assigning properties consistent with more involved modeling. Our subhalo mass function in Equation (14) is independent of the lens redshift, but can easily be extended to include this dependency (Despali & Vegetti 2017; Hiroshima et al. 2018). A more complicated dependence on the host halo than the linear one assumed in Equation (14) is also easily admitted.

All of these effects are straightforward to implement in our setup and only require modifications to the simulation code. The inference algorithm is unaffected; since these extensions do not explicitly depend on the parameters of interest, the likelihood terms associated with them cancel in the calculation of the joint likelihood ratio and the joint score. Nevertheless, these changes affect the final observed image and therefore also the true likelihood function; the variance of the joint likelihood ratio and score could therefore increase, requiring larger training samples before the network converges to the correct likelihood ratio function.

With these extensions, the redshift of the background source and the lens will play a more important role. Since these redshifts can potentially be measured through spectroscopic follow-up observations, it is likely that we can improve the performance of the inference algorithm by using this information. We can treat both the source and lens redshift, potentially with added uncertainty to model measurement noise, as additional observables. The input to the neural networks then consists of the observed lens image, the measured (potentially noisy) redshifts, and the tested parameter point. Except for a simple modification of the network architecture, the inference algorithm remains unchanged.

Including line-of-sight substructure can be somewhat more involved, since it necessitates the introduction of a separate line-of-sight halo mass function (Birrer et al. 2017b; Despali et al. 2018; Gilman et al. 2019b; Hsueh et al. 2019). Depending on the specific model (and whether foreground substructure is treated as a nuisance effect or additional signal to be leveraged) its parameters could depend on the parameters of interest, which would require a modification of the calculation of the joint likelihood ratio and joint score. Structurally this is identical to our current modeling of subhalos within the lens. Since the abundance of foreground substructure is expected to be at most comparable to the substructure within the lensing galaxy (depending on the source redshift), we expect that these additional factors in the joint likelihood ratio and joint score will not slow down the overall simulation significantly, and will not increase the variance of the inference techniques too much while
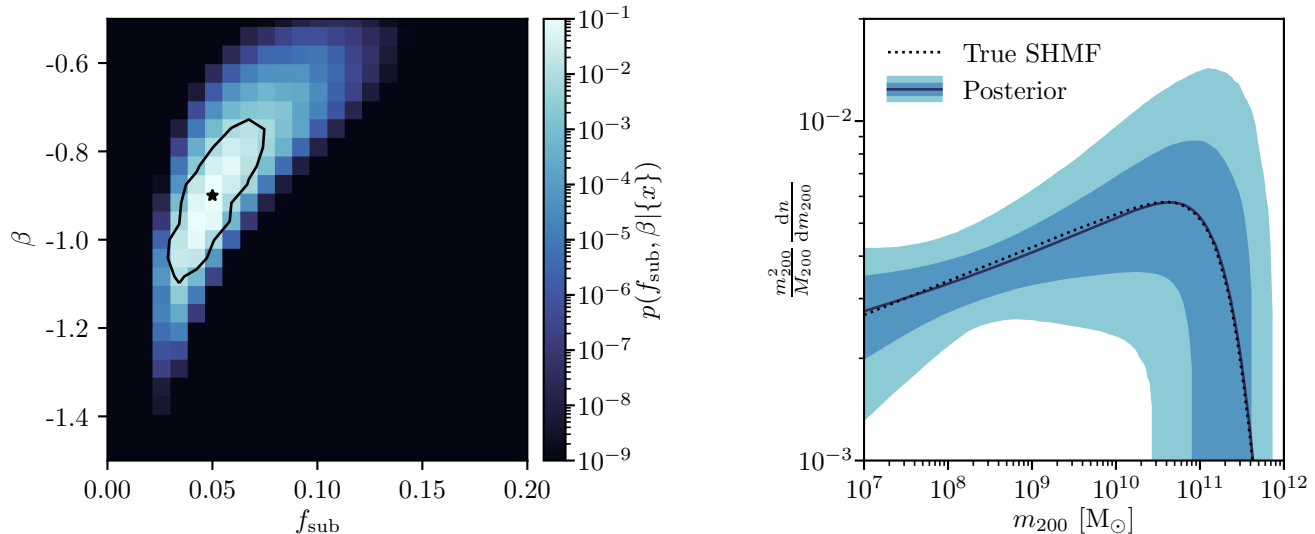
Figure 5. Left: Expected posterior and 95% credible region on subhalo mass function parameters for 100 observed lenses. The mock observations are generated for $f_{\rm sub} = 0.05$ and $\beta = -0.9$. We assume a uniform prior on $f_{\rm sub}$ and a Gaussian prior with mean -0.9 and standard deviation 0.1 for β. ⟨▶⟩ Right: Corresponding inferred subhalo mass function, marginalized over the host halo properties. We show the point-wise mean (solid line) as well as point-wise 68/95% credible intervals (cyan/blue bands) of the subhalo mass function per host halo mass. The true simulated subhalo mass function for $f_{\rm sub} = 0.05$ and $\beta = -0.9$ is shown as the dotted line. An animation showing the evolution of the substructure parameter and mass function posteriors as the number of lenses analyzed is increased can be found at ○. ⟨▶⟩

having the potential to improve the overall sensitivity of the analysis to substructure abundance in the Universe.

Modification to the subhalo mass function parameterization that we have considered may be desirable for constraining specific particle physics scenarios. For example, warm dark matter introduces a lower cutoff scale in the subhalo mass function (Bode et al. 2001) which can be parameterized and mapped onto the dark matter mass (Schneider et al. 2012; Lovell et al. 2014; Li et al. 2016; Birrer et al. 2017a). This would also require a straightforward modification of the joint likelihood ratio and joint score calculation depending on the specific parameterization.

It is expected that a sample of strong lenses will include image-to-image variations on the exposure, sky background, and detector effects like the point spread function depending on the specific scanning strategy of the observatory. The sky background can be marginalized over as usual. Rather than treating the exposure and PSF model as nuisance parameters, passing them as additional a priori known inputs to the network in addition to normalizing the network input to unit exposure is likely to improve performance. Multiple color bands can easily be modeled and included as inputs to the neural network as different color channels, something that is commonly done when using the ResNet architecture we consider. This can substantially improve discrimination between light from the source, host, and sky background which tend to have a degree of separation in color space.

While including these extensions in our simulation and inference code is feasible, the detailed modeling is beyond the scope of the current paper. We thus leave the implementation of these features and application to real lensing data to future work.

6. CONCLUSIONS

Strong lensing offers a unique way to probe the properties and distribution of dark matter on sub-galactic scales through the subtle imprint of substructure on lensed arcs. The high dimensionality of the underlying latent space characterizing substructure poses a significant challenge, however. In this paper, we have introduced a novel simulation-based inference technique based on the ideas introduced in Cranmer et al. (2015); Brehmer et al. (2018a,b,c); Stoye et al. (2018) for inferring high-level population properties characterizing the distribution of substructure in an ensemble of galaxy-galaxy strong lenses and demonstrated its feasibility through proof-of-principle examples.

Our results on simulated data demonstrate that this method, based on calibrated likelihood ratio estimators with a machine learning back end, offers a promising way to analyze extended-arc strong lensing images with the goal of inferring properties of dark matter substructure. Our proposed method offers several combined advantages over established techniques. In probing the collective effect of a large number of low-mass, sub-threshold subhalos it can offer sensitivity to the faint end of the

subhalo mass function where deviations from the concordance ΛCDM paradigm and the effects of new physics are most likely to be expressed. It can naturally be applied to perform fast, principled, and concurrent analyses of a large sample of strong lenses that share a common set of hyperparameters describing the underlying substructure population properties. By efficiently marginalizing out the individual subhalo properties and directly inferring the population-level parameters of interest we are able to sidestep the more expensive two-step procedure of characterizing individual subhalos before using them to infer higher-level population parameters. Population-level likelihood scans for individual images are thus a suitable alternative to probabilistic catalogs over subhalos, avoiding both prior dependence as well as the logistical complexity of communicating a complicated trans-dimensional posterior. Furthermore, rigorous selection of lensing images out of a large sample is not necessary within our framework since images with a smaller effective arc area or low overall fidelity simply do not contribute significantly to the simultaneous substructure analysis, and non-detections are just as valuable as detections. Finally, our analysis is performed at the level of image data without incurring loss of information associated with dimensionality reduction.

Although we have focused on a simple proof-of-principle example in this paper, extensions to more realistic scenarios—including more complicated descriptions of the host, source, and substructure populations—are easily admitted within our framework. The flexibility of the proposed method allows for applications beyond substructure population inference as well. For example, a large lens sample can be used to perform cosmological parameter estimation while accounting for substructure effects and in particular to independently constrain the Hubble constant (Chen et al. 2019; Wong et al. 2019) through its dependence on the angular diameter distance scales in lensing systems. In the spirit of Alsing & Wandelt (2018), these techniques can also be used to learn powerful summary statistics (Brehmer et al. 2018c).

We are currently at the dawn of a new era in observational cosmology, when ongoing and upcoming surveys—e. g., DES, LSST, *Euclid*, and WFIRST—are expected to discover and deliver images of thousands of strong lensing systems. These will harbor the subtle imprint of dark matter substructure, whose characterization could hold the key to unveiling the particle nature of dark matter. In this paper, we have introduced a powerful machine learning-based method that can be used to uncover the properties of small-scale structure within these lenses and in the Universe at large. The techniques presented have the potential to maximize the informa-

tion that can be extracted from a complex lens sample and zero in on signatures of new physics.

The code used to obtain the results in this paper is available at https://github.com/smsharma/mining-for-substructure-lens ⊙.

APPENDIX

A. MINIMUM OF THE LOSS FUNCTIONAL

A central step in our inference technique is numerically minimizing the functional $L[g(x,\vartheta)]$ given in Equation (21) to obtain an estimator for the likelihood ratio function. Here we will use calculus of variation to explicitly show that the solution given in Equation (22) in fact minimizes this loss, closely following Brehmer et al. (2018b); Stoye et al. (2018).

First consider the case of $\alpha = 0$, i.e. the functional

$$
\begin{aligned}
L[g(x,\vartheta)] &= \int \mathrm{d}\vartheta \int \mathrm{d}\vartheta' \int \mathrm{d}x \int \mathrm{d}z\, \pi(\vartheta)\pi(\vartheta')p(x,z|\vartheta)\Big(-s\log g - (1-s)\log(1-g) - s'\log g' - (1-s')\log(1-g')\Big) \\
&= \int \mathrm{d}\vartheta \int \mathrm{d}x\, \underbrace{\left[\int \mathrm{d}z\, \pi(\vartheta)\Big(p(x,z|\vartheta) + p_{\mathrm{ref}}(x,z)\Big)\Big(-s\log g - (1-s)\log(1-g)\Big)\right]}_{\equiv F(x,\vartheta)},
\end{aligned}
\tag{A1}
$$

where we use the shorthand notation $s \equiv s(x,z|\vartheta) \equiv 1/(1+r(x,z|\vartheta))$, $s' \equiv s(x,z|\vartheta') \equiv 1/(1+r(x,z|\vartheta'))$, $g \equiv g(x,\vartheta)$, $g' \equiv g(x,\vartheta')$. The function $g^*(x|\vartheta)$ that minimizes this functional has to satisfy

$$
0 \overset{!}{=} \left.\frac{\delta F}{\delta g}\right|_{g^*} = \int \mathrm{d}z\, \pi(\vartheta)\Big(p(x,z|\vartheta) + p_{\mathrm{ref}}(x,z)\Big)\Big(-\frac{s}{g^*} + \frac{1-s}{1-g^*}\Big)
\tag{A2}
$$

As long as $\pi(\vartheta) > 0$, this is equivalent to

$$
(1-g^*)\int \mathrm{d}z\Big(p(x,z|\vartheta) + p_{\mathrm{ref}}(x,z)\Big)s = g^*\int \mathrm{d}z\Big(p(x,z|\vartheta) + p_{\mathrm{ref}}(x,z)\Big)(1-s)
\tag{A3}
$$

and finally

$$
\begin{aligned}
g^*(x|\vartheta) &= \frac{\int \mathrm{d}z\Big(p(x,z|\vartheta) + p_{\mathrm{ref}}(x,z)\Big)s(x,z|\vartheta)}{\int \mathrm{d}z\Big(p(x,z|\vartheta) + p_{\mathrm{ref}}(x,z)\Big)} \\
&= \frac{\int \mathrm{d}z\Big(p(x,z|\vartheta) + p_{\mathrm{ref}}(x,z)\Big)\frac{1}{1+p(x,z|\vartheta)/p_{\mathrm{ref}}(x,z)}}{\int \mathrm{d}z\Big(p(x,z|\vartheta) + p_{\mathrm{ref}}(x,z)\Big)} \\
&= \frac{p_{\mathrm{ref}}(x)}{p(x|\vartheta) + p_{\mathrm{ref}}(x)} = \frac{1}{1+r(x|\vartheta)},
\end{aligned}
\tag{A4}
$$

in agreement with Equation (22). Note that this result is independent of the choice of $\pi(\vartheta)$, as long as this proposal distribution has support at all relevant parameter points.

Similarly it can be shown that the gradient term in the loss functional weighted by α is minimized when the gradient of the log likelihood ratio estimated by the neural network is equal to the true score,

$$
\nabla_\vartheta \log \hat{r}(x|\vartheta) \equiv \nabla_\vartheta \log \frac{1-g^*(x,\vartheta)}{g^*(x,\vartheta)} = \nabla_\vartheta \log r(x|\vartheta).
\tag{A5}
$$

We refer the reader to Brehmer et al. (2018b) for the derivation. While not strictly necessary for the inference technique, including this term in the loss function substantially improves the sample efficiency of the algorithm, similar to how gradient information makes any fit converge faster.

B. SIMPLIFIED SCENARIOS

In order to validate our setup and to disentangle the impact of different latent variables on the inference results we consider three additional versions of our simulation. In the simplest one, which we call "fix", all source and host
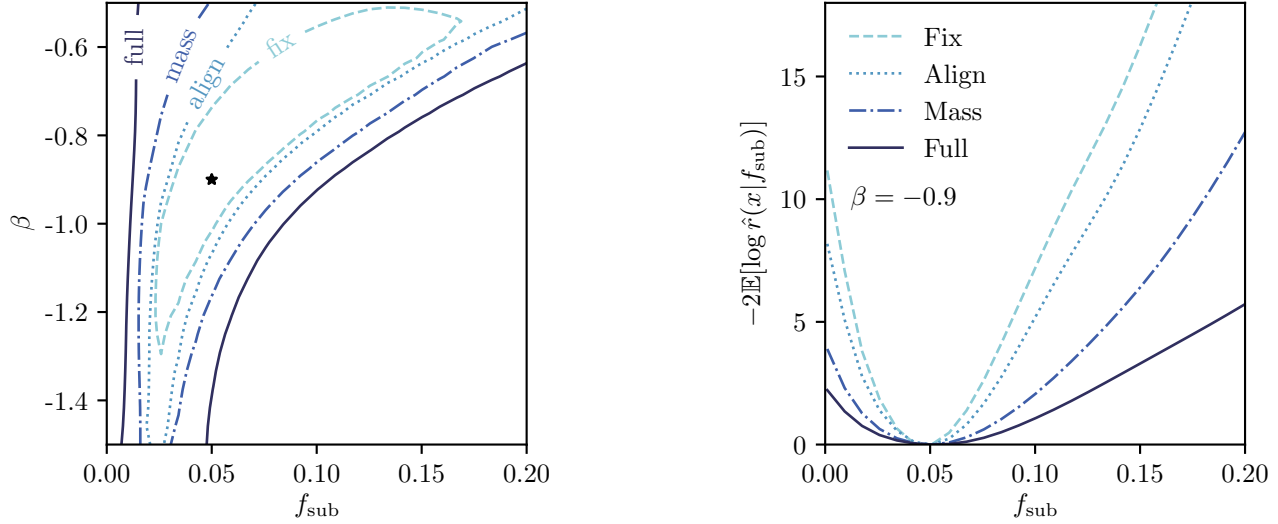
Figure 6. Left: The expected 95% CL exclusion limits for 5 observed lenses for four different levels of complexity of the simulator. Right: The expected likelihood ratio along a one-dimensional slice through the parameter space at $\beta = -0.9$ for the same four simulator scenarios. In both panels we compare the "full" simulator discussed in Section 2, a scenario in which the host mass is varied but the offset relative to the source is fixed at zero ("mass"), a case in which the source offset is varied but the host halo mass is fixed ("align"), and a toy scenario in which both the offset and the mass of the host halo are fixed ("fix"). The data was generated for $\beta = -0.9$ and $f_{\rm sub} = 0.05$. 🔗

properties are fixed to particular value, including the host halo mass and the offset between source and lens, which is set to zero. In the "align" scenario we relax the restriction on the source offset variables $\Delta\theta_x$ and $\Delta\theta_y$ and draw them from a Gaussian as described in Section 2. In the "mass" version, on the other hand, the offset is fixed at zero, but the host halo mass is drawn from a distribution as described above. We train separate neural networks on lens images generated in these three scenarios and calculated likelihood maps as described in Section 3, although to save computation time we do not perform a calibration procedure.

The expected confidence limits for 5 observed lens images in the three simplified scenarios and our "full" setup are compared in Figure 6. As expected, the more latent variables we keep fixed, the more the inference technique becomes more sensitive to the parameters of interest. In particular fixing the source-host alignment substantially increases the strength of the expected limits.

REFERENCES

Aaboud, M., Aad, G., Abbott, B., et al. 2019, JHEP, 2019, 142, arXiv: 1903.01400, doi: 10.1007/JHEP05(2019)142

Agrawal, P., Cyr-Racine, F.-Y., Randall, L., & Scholtz, J. 2017, JCAP, 2017, 021, arXiv: 1702.05482, doi: 10.1088/1475-7516/2017/08/021

Agrawal, P., & Randall, L. 2017, JCAP, 2017, 019, arXiv: 1706.04195, doi: 10.1088/1475-7516/2017/12/019

Akerib, D. S., Alsum, S., Araújo, H. M., et al. 2017, PhRvL, 118, 021303, arXiv: 1608.07648, doi: 10.1103/PhysRevLett.118.021303

Albert, A., Anderson, B., Bechtol, K., et al. 2017, ApJ, 834, 110, arXiv: 1611.03184, doi: 10.3847/1538-4357/834/2/110

Alsing, J., Charnock, T., Feeney, S., & Wandelt, B. 2019, MNRAS, 488, 4440, arXiv: 1903.00007, doi: 10.1093/mnras/stz1960

Alsing, J., & Wandelt, B. 2018, MNRAS, 476, L60, arXiv: 1712.00012, doi: 10.1093/mnrasl/sly029

Aprile, E., Aalbers, J., Agostini, F., et al. 2018, PhRvL, 121, 111302, arXiv: 1805.12562, doi: 10.1103/PhysRevLett.121.111302

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33, arXiv: 1307.6212, doi: 10.1051/0004-6361/201322068

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123, arXiv: 1801.02634, doi: 10.3847/1538-3881/aabc4f

Baldi, P., Cranmer, K., Faucett, T., Sadowski, P., & Whiteson, D. 2016, European Physical Journal C, 76, 235, arXiv: 1601.07913, doi: 10.1140/epjc/s10052-016-4099-4

Baltz, E. A., Marshall, P., & Oguri, M. 2009, JCAP, 2009, 015, arXiv: 0705.0682, doi: 10.1088/1475-7516/2009/01/015

Bartelmann, M., & Schneider, P. 2001, PhR, 340, 291, arXiv: astro-ph/9912508, doi: 10.1016/S0370-1573(00)00082-X

Bechtol, K., Drlica-Wagner, A., Balbinot, E., et al. 2015, ApJ, 807, 50, arXiv: 1503.02584, doi: 10.1088/0004-637X/807/1/50

Birrer, S., & Amara, A. 2018, Physics of the Dark Universe, 22, 189, arXiv: 1803.09746, doi: 10.1016/j.dark.2018.11.002

Birrer, S., Amara, A., & Refregier, A. 2015, ApJ, 813, 102, arXiv: 1504.07629, doi: 10.1088/0004-637X/813/2/102

—. 2017a, JCAP, 2017, 037, arXiv: 1702.00009, doi: 10.1088/1475-7516/2017/05/037

Birrer, S., Welschen, C., Amara, A., & Refregier, A. 2017b, JCAP, 2017, 049, arXiv: 1610.01599, doi: 10.1088/1475-7516/2017/04/049

Bode, P., Ostriker, J. P., & Turok, N. 2001, ApJ, 556, 93, arXiv: astro-ph/0010389, doi: 10.1086/321541

Bonaca, A., & Hogg, D. W. 2018, ApJ, 867, 101, arXiv: 1804.06854, doi: 10.3847/1538-4357/aae4da

Bonaca, A., Hogg, D. W., Price-Whelan, A. M., & Conroy, C. 2019, ApJ, 880, 38, arXiv: 1811.03631, doi: 10.3847/1538-4357/ab2873

Bond, J. R., & Szalay, A. S. 1983, ApJ, 274, 443, doi: 10.1086/161460

Boyanovsky, D., de Vega, H. J., & Sanchez, N. G. 2008, PhRvD, 78, 063546, arXiv: 0807.0622, doi: 10.1103/PhysRevD.78.063546

Boyanovsky, D., & Wu, J. 2011, PhRvD, 83, 043524, arXiv: 1008.0992, doi: 10.1103/PhysRevD.83.043524

Brehmer, J., Cranmer, K., Louppe, G., & Pavez, J. 2018a, PhRvL, 121, 111801, arXiv: 1805.00013, doi: 10.1103/PhysRevLett.121.111801

—. 2018b, PhRvD, 98, 052004, arXiv: 1805.00020, doi: 10.1103/PhysRevD.98.052004

Brehmer, J., Kling, F., Espejo, I., & Cranmer, K. 2019, arXiv: 1907.10621

Brehmer, J., Louppe, G., Pavez, J., & Cranmer, K. 2018c, arXiv: 1805.12244

Brennan, S., Benson, A. J., Cyr-Racine, F.-Y., et al. 2019, MNRAS, 488, 5085, doi: 10.1093/mnras/stz1607

Brewer, B. J., Huijser, D., & Lewis, G. F. 2016, MNRAS, 455, 1819, arXiv: 1508.00662, doi: 10.1093/mnras/stv2370

Brooks, A. M. 2018, arXiv: 1812.00044

Buckley, M. R., & DiFranzo, A. 2018, PhRvL, 120, 051102, arXiv: 1707.03829, doi: 10.1103/PhysRevLett.120.051102

Buckley, M. R., & Peter, A. H. G. 2018, Physics Reports, 761, 1, arXiv: 1712.06615, doi: 10.1016/j.physrep.2018.07.003

Buckley, M. R., Zavala, J., Cyr-Racine, F.-Y., Sigurdson, K., & Vogelsberger, M. 2014, PhRvD, 90, 043524, arXiv: 1405.2075, doi: 10.1103/PhysRevD.90.043524

Buschmann, M., Kopp, J., Safdi, B. R., & Wu, C.-L. 2018, PhRvL, 120, 211101, arXiv: 1711.03554, doi: 10.1103/PhysRevLett.120.211101

Carlberg, R. G. 2012, ApJ, 748, 20, arXiv: 1109.6022, doi: 10.1088/0004-637X/748/1/20

Carlberg, R. G., & Grillmair, C. J. 2013, ApJ, 768, 171, arXiv: 1303.4342, doi: 10.1088/0004-637X/768/2/171

Chang, L. J., Lisanti, M., & Mishra-Sharma, S. 2018, PhRvD, 98, 123004, arXiv: 1804.04132, doi: 10.1103/PhysRevD.98.123004

Chatterjee, S., & Koopmans, L. V. E. 2018, MNRAS, 474, 1762, arXiv: 1710.03075, doi: 10.1093/mnras/stx2674

Chen, G. C. F., Fassnacht, C. D., Suyu, S. H., et al. 2019, arXiv: 1907.02533

Chen, R. T. Q., Rubanova, Y., Bettencourt, J., & Duvenaud, D. 2018, arXiv: 1806.07366

Ciotti, L., & Bertin, G. 1999, A&A, 352, 447

Colín, P., Avila-Reese, V., Valenzuela, O., & Firmani, C. 2002, ApJ, 581, 777, arXiv: astro-ph/0205322, doi: 10.1086/344259

Collett, T. E. 2015, ApJ, 811, 20, arXiv: 1507.02657, doi: 10.1088/0004-637X/811/1/20

Cowan, G., Cranmer, K., Gross, E., & Vitells, O. 2011, European Physical Journal C, 71, 1554, arXiv: 1007.1727, doi: 10.1140/epjc/s10052-011-1554-0

Cranmer, K., Pavez, J., & Louppe, G. 2015, arXiv: 1506.02169

Cui, X., Abdukerim, A., Chen, W., et al. 2017, PhRvL, 119, 181302, doi: 10.1103/PhysRevLett.119.181302

Cyr-Racine, F.-Y., Keeton, C. R., & Moustakas, L. A. 2019, PhRvD, 100, 023013, arXiv: 1806.07897, doi: 10.1103/PhysRevD.100.023013

Cyr-Racine, F.-Y., Moustakas, L. A., Keeton, C. R., Sigurdson, K., & Gilman, D. A. 2016, PhRvD, 94, 043505, arXiv: 1506.01724, doi: 10.1103/PhysRevD.94.043505

Dalal, N., & Kochanek, C. S. 2002, ApJ, 572, 25, arXiv: astro-ph/0111456, doi: 10.1086/340303

Dalcanton, J. J., & Hogan, C. J. 2001, ApJ, 561, 35, arXiv: astro-ph/0004381, doi: 10.1086/323207

Dark Energy Survey Collaboration, Abbott, T., Abdalla,
F. B., et al. 2016, MNRAS, 460, 1270, arXiv: 1601.00329,
doi: 10.1093/mnras/stw641

Davé, R., Spergel, D. N., Steinhardt, P. J., & Wandelt,
B. D. 2001, ApJ, 547, 574, arXiv: astro-ph/0006218,
doi: 10.1086/318417

Davis, M. 2012–, palettable: Color palettes for Python.
https://jiffyclub.github.io/palettable/

Daylan, T., Cyr-Racine, F.-Y., Diaz Rivero, A., Dvorkin,
C., & Finkbeiner, D. P. 2018, ApJ, 854, 141,
arXiv: 1706.06111, doi: 10.3847/1538-4357/aaaa1e

Despali, G., & Vegetti, S. 2017, MNRAS, 469, 1997,
arXiv: 1608.06938, doi: 10.1093/mnras/stx966

Despali, G., Vegetti, S., White, S. D. M., Giocoli, C., & van
den Bosch, F. C. 2018, MNRAS, 475, 5424,
arXiv: 1710.05029, doi: 10.1093/mnras/sty159

Diaz Rivero, A., Cyr-Racine, F.-Y., & Dvorkin, C. 2018,
PhRvD, 97, 023001, arXiv: 1707.04590,
doi: 10.1103/PhysRevD.97.023001

Díaz Rivero, A., Dvorkin, C., Cyr-Racine, F.-Y., Zavala, J.,
& Vogelsberger, M. 2018, PhRvD, 98, 103517,
arXiv: 1809.00004, doi: 10.1103/PhysRevD.98.103517

Dinev, T., & Gutmann, M. U. 2018, arXiv: 1810.09899

Dinh, L., Krueger, D., & Bengio, Y. 2014, arXiv: 1410.8516

Dinh, L., Sohl-Dickstein, J., & Bengio, S. 2016,
arXiv: 1605.08803

Drlica-Wagner, A., Bechtol, K., Rykoff, E. S., et al. 2015,
ApJ, 813, 109, arXiv: 1508.03622,
doi: 10.1088/0004-637X/813/2/109

Drlica-Wagner, A., Mao, Y.-Y., Adhikari, S., et al. 2019,
arXiv: 1902.01055

Efstathiou, G. 1992, MNRAS, 256, 43P,
doi: 10.1093/mnras/256.1.43P

Elbert, O. D., Bullock, J. S., Kaplinghat, M., et al. 2018,
ApJ, 853, 109, arXiv: 1609.08626,
doi: 10.3847/1538-4357/aa9710

Errani, R., Peñarrubia, J., Laporte, C. F. P., & Gómez,
F. A. 2017, MNRAS, 465, L59, arXiv: 1608.01849,
doi: 10.1093/mnrasl/slw211

Fadely, R., & Keeton, C. R. 2012, MNRAS, 419, 936,
arXiv: 1109.0548, doi: 10.1111/j.1365-2966.2011.19729.x

Fan, J., Katz, A., Randall, L., & Reece, M. 2013, Physics of
the Dark Universe, 2, 139, arXiv: 1303.1521,
doi: 10.1016/j.dark.2013.07.001

Fan, Y., Nott, D. J., & Sisson, S. A. 2012, arXiv: 1212.1479

Fitts, A., Boylan-Kolchin, M., Elbert, O. D., et al. 2017,
MNRAS, 471, 3547, arXiv: 1611.02281,
doi: 10.1093/mnras/stx1757

Fitts, A., Boylan-Kolchin, M., Bozek, B., et al. 2018,
arXiv: 1811.11791

Garrison-Kimmel, S., Wetzel, A., Bullock, J. S., et al. 2017,
MNRAS, 471, 1709, arXiv: 1701.03792,
doi: 10.1093/mnras/stx1710

Germain, M., Gregor, K., Murray, I., & Larochelle, H.
2015, arXiv: 1502.03509

Gilman, D., Birrer, S., Nierenberg, A., et al. 2019a,
arXiv: 1908.06983

Gilman, D., Birrer, S., Treu, T., Keeton, C. R., &
Nierenberg, A. 2018, MNRAS, 481, 819,
arXiv: 1712.04945, doi: 10.1093/mnras/sty2261

Gilman, D., Birrer, S., Treu, T., Nierenberg, A., & Benson,
A. 2019b, MNRAS, 487, 5721, arXiv: 1901.11031,
doi: 10.1093/mnras/stz1593

Grathwohl, W., Chen, R. T. Q., Bettencourt, J., Sutskever,
I., & Duvenaud, D. 2018, arXiv: 1810.01367

Gutmann, M. U., Dutta, R., Kaski, S., & Corander, J.
2017, Statistics and Computing, 1

Han, J., Cole, S., Frenk, C. S., & Jing, Y. 2016, MNRAS,
457, 1208, arXiv: 1509.02175, doi: 10.1093/mnras/stv2900

He, K., Zhang, X., Ren, S., & Sun, J. 2016, in Proceedings
of the IEEE conference on computer vision and pattern
recognition, 770–778

Hermans, J., Begy, V., & Louppe, G. 2019,
arXiv: 1903.04057

Hezaveh, Y., Dalal, N., Holder, G., et al. 2016a, JCAP,
2016, 048, arXiv: 1403.2720,
doi: 10.1088/1475-7516/2016/11/048

Hezaveh, Y. D., Perreault Levasseur, L., & Marshall, P. J.
2017, Nature, 548, 555, arXiv: 1708.08842,
doi: 10.1038/nature23463

Hezaveh, Y. D., Dalal, N., Marrone, D. P., et al. 2016b,
ApJ, 823, 37, arXiv: 1601.01388,
doi: 10.3847/0004-637X/823/1/37

Hiroshima, N., Ando, S., & Ishiyama, T. 2018, PhRvD, 97,
123002, arXiv: 1803.07691,
doi: 10.1103/PhysRevD.97.123002

Hsueh, J.-W., Enzi, W., Vegetti, S., et al. 2019,
arXiv: 1905.04182

Huang, C.-W., Krueger, D., Lacoste, A., & Courville, A.
2018, arXiv: 1804.00779

Hunter, J. D. 2007, Computing In Science & Engineering,
9, 90

Jimenez Rezende, D., & Mohamed, S. 2015,
arXiv: 1505.05770

Johnston, K. V., Zhao, H., Spergel, D. N., & Hernquist, L.
1999, ApJL, 512, L109, arXiv: astro-ph/9807243,
doi: 10.1086/311876

Jones, E., Oliphant, T., Peterson, P., et al. 2001–, SciPy:
Open source scientific tools for Python.
http://www.scipy.org/

Kahlhoefer, F., Kaplinghat, M., Slatyer, T. R., & Wu, C.-L. 2019, arXiv: 1904.10539

Kamada, A., Kaplinghat, M., Pace, A. B., & Yu, H.-B. 2017, PhRvL, 119, 111102, arXiv: 1611.02716, doi: 10.1103/PhysRevLett.119.111102

Kaplinghat, M., Keeley, R. E., Linden, T., & Yu, H.-B. 2014, PhRvL, 113, 021302, arXiv: 1311.6524, doi: 10.1103/PhysRevLett.113.021302

Kaplinghat, M., Tulin, S., & Yu, H.-B. 2016, PhRvL, 116, 041302, arXiv: 1508.03339, doi: 10.1103/PhysRevLett.116.041302

Keeton, C. R. 2001, arXiv: astro-ph/0102341

Keeton, C. R., Kochanek, C. S., & Seljak, U. 1997, ApJ, 482, 604, arXiv: astro-ph/9610163, doi: 10.1086/304172

Kingma, D. P., & Dhariwal, P. 2018, arXiv: 1807.03039

Kluyver, T., Ragan-Kelley, B., Pérez, F., et al. 2016, in ELPUB

Koposov, S., Belokurov, V., Evans, N. W., et al. 2008, ApJ, 686, 279, arXiv: 0706.2687, doi: 10.1086/589911

Koposov, S. E., Belokurov, V., Torrealba, G., & Evans, N. W. 2015, ApJ, 805, 130, arXiv: 1503.02079, doi: 10.1088/0004-637X/805/2/130

Kormann, R., Schneider, P., & Bartelmann, M. 1994, A&A, 284, 285

Li, R., Frenk, C. S., Cole, S., et al. 2016, MNRAS, 460, 363, arXiv: 1512.06507, doi: 10.1093/mnras/stw939

Lisanti, M., Mishra-Sharma, S., Rodd, N. L., & Safdi, B. R. 2018, PhRvL, 120, 101101, arXiv: 1708.09385, doi: 10.1103/PhysRevLett.120.101101

Louppe, G., & Cranmer, K. 2017, arXiv: 1707.07113

Lovell, M. R., Frenk, C. S., Eke, V. R., et al. 2014, MNRAS, 439, 300, arXiv: 1308.1399, doi: 10.1093/mnras/stt2431

LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv: 0912.0201

Lueckmann, J.-M., Bassetto, G., Karaletsos, T., & Macke, J. H. 2018, arXiv: 1805.09294

Lueckmann, J.-M., Goncalves, P. J., Bassetto, G., et al. 2017, arXiv: 1711.01861

Madau, P., Diemand, J., & Kuhlen, M. 2008, ApJ, 679, 1260, arXiv: 0802.2265, doi: 10.1086/587545

Moliné, Á., Sánchez-Conde, M. A., Palomares-Ruiz, S., & Prada, F. 2017, MNRAS, 466, 4974, arXiv: 1603.04057, doi: 10.1093/mnras/stx026

Morningstar, W. R., Hezaveh, Y. D., Perreault Levasseur, L., et al. 2018, arXiv: 1808.00011

Morningstar, W. R., Perreault Levasseur, L., Hezaveh, Y. D., et al. 2019, arXiv: 1901.01359

Nadler, E. O., Mao, Y.-Y., Green, G. M., & Wechsler, R. H. 2019, ApJ, 873, 34, arXiv: 1809.05542, doi: 10.3847/1538-4357/ab040e

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563, doi: 10.1086/177173

—. 1997, ApJ, 490, 493, doi: 10.1086/304888

Neyman, J., & Pearson, E. S. 1933, Philosophical Transactions of the Royal Society of London Series A, 231, 289, doi: 10.1098/rsta.1933.0009

Nierenberg, A. M., Treu, T., Wright, S. A., Fassnacht, C. D., & Auger, M. W. 2014, MNRAS, 442, 2434, arXiv: 1402.1496, doi: 10.1093/mnras/stu862

Nierenberg, A. M., Treu, T., Brammer, G., et al. 2017, MNRAS, 471, 2224, arXiv: 1701.05188, doi: 10.1093/mnras/stx1400

Nightingale, J. W., Dye, S., & Massey, R. J. 2018, MNRAS, 478, 4738, arXiv: 1708.07377, doi: 10.1093/mnras/sty1264

Nishikawa, H., Boddy, K. K., & Kaplinghat, M. 2019, arXiv: 1901.00499

Oguri, M., & Marshall, P. J. 2010, MNRAS, 405, 2579, arXiv: 1001.2037, doi: 10.1111/j.1365-2966.2010.16639.x

Paige, B., & Wood, F. 2016, arXiv: 1602.06701

Papamakarios, G., & Murray, I. 2016, in Advances in Neural Information Processing Systems, 1028–1036

Papamakarios, G., Pavlakou, T., & Murray, I. 2017, arXiv: 1705.07057

Papamakarios, G., Sterratt, D. C., & Murray, I. 2018, arXiv: 1805.07226

Paszke, A., Gross, S., Chintala, S., et al. 2017, in NIPS-W

Perez, F., & Granger, B. E. 2007, Computing in Science and Engineering, 9, 21, doi: 10.1109/MCSE.2007.53

Perreault Levasseur, L., Hezaveh, Y. D., & Wechsler, R. H. 2017, ApJL, 850, L7, arXiv: 1708.08843, doi: 10.3847/2041-8213/aa9704

Peter, A. H. G., Rocha, M., Bullock, J. S., & Kaplinghat, M. 2013, MNRAS, 430, 105, arXiv: 1208.3026, doi: 10.1093/mnras/sts535

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13, arXiv: 1502.01589, doi: 10.1051/0004-6361/201525830

Qian, N. 1999, Neural Netw., 12, 145, doi: 10.1016/S0893-6080(98)00116-6

Read, J. I., Iorio, G., Agertz, O., & Fraternali, F. 2017, MNRAS, 467, 2019, arXiv: 1607.03127, doi: 10.1093/mnras/stx147

Refregier, A., Amara, A., Kitching, T. D., et al. 2010, arXiv: 1001.0061

Robles, V. H., Kelley, T., Bullock, J. S., & Kaplinghat, M. 2019, arXiv: 1903.01469

Rubin, D. B. 1984, Ann. Statist., 12, 1151, doi: 10.1214/aos/1176346785

Sánchez-Conde, M. A., & Prada, F. 2014, MNRAS, 442, 2271, arXiv: 1312.1729, doi: 10.1093/mnras/stu1014

Schewtschenko, J. A., Wilkinson, R. J., Baugh, C. M., Bœhm, C., & Pascoli, S. 2015, MNRAS, 449, 3587, arXiv: 1412.4905, doi: 10.1093/mnras/stv431

Schneider, A., Smith, R. E., Macciò, A. V., & Moore, B. 2012, MNRAS, 424, 684, arXiv: 1112.0330, doi: 10.1111/j.1365-2966.2012.21252.x

Schneider, P. 1997, MNRAS, 292, 673, arXiv: astro-ph/9706185, doi: 10.1093/mnras/292.3.673

Schneider, P., Ehlers, J., & Falco, E. E. 1992, Gravitational Lenses, doi: 10.1007/978-3-662-03758-4

Sérsic, J. L. 1963, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 6, 41

Simon, J., Birrer, S., Bechtol, K., et al. 2019, BAAS, 51, 153, arXiv: 1903.04742

Sirunyan, A. M., Tumasyan, A., Adam, W., et al. 2017, Physics Letters B, 769, 520, arXiv: 1611.03568, doi: 10.1016/j.physletb.2017.02.012

Spergel, D. N., & Steinhardt, P. J. 2000, PhRvL, 84, 3760, arXiv: astro-ph/9909386, doi: 10.1103/PhysRevLett.84.3760

Springel, V., Wang, J., Vogelsberger, M., et al. 2008, MNRAS, 391, 1685, arXiv: 0809.0898, doi: 10.1111/j.1365-2966.2008.14066.x

Stoye, M., Brehmer, J., Louppe, G., Pavez, J., & Cranmer, K. 2018, arXiv: 1808.00973

Suyu, S. H., Marshall, P. J., Hobson, M. P., & Bland ford, R. D. 2006, MNRAS, 371, 983, arXiv: astro-ph/0601493, doi: 10.1111/j.1365-2966.2006.10733.x

Tagore, A. S., & Jackson, N. 2016, MNRAS, 457, 3066, arXiv: 1505.00198, doi: 10.1093/mnras/stw057

Tagore, A. S., & Keeton, C. R. 2014, MNRAS, 445, 694, arXiv: 1408.6297, doi: 10.1093/mnras/stu1671

Thomas, O., Dutta, R., Corander, J., Kaski, S., & Gutmann, M. U. 2016, arXiv: 1611.10242

Tran, D., Ranganath, R., & Blei, D. 2017, in Advances in Neural Information Processing Systems 30, ed. I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett, 5523–5533

Treu, T. 2010, Annual Review of Astronomy and Astrophysics, 48, 87, arXiv: 1003.5567, doi: 10.1146/annurev-astro-081309-130924

Uria, B., Côté, M.-A., Gregor, K., Murray, I., & Larochelle, H. 2016, arXiv: 1605.02226

van den Oord, A., Kalchbrenner, N., & Kavukcuoglu, K. 2016a, arXiv: 1601.06759

van den Oord, A., Kalchbrenner, N., Vinyals, O., et al. 2016b, arXiv: 1606.05328

van den Oord, A., Dieleman, S., Zen, H., et al. 2016c, arXiv: 1609.03499

van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, Computing in Science and Engineering, 13, 22, arXiv: 1102.1523, doi: 10.1109/MCSE.2011.37

Van Tilburg, K., Taki, A.-M., & Weiner, N. 2018, JCAP, 2018, 041, arXiv: 1804.01991, doi: 10.1088/1475-7516/2018/07/041

Vegetti, S., Czoske, O., & Koopmans, L. V. E. 2010a, MNRAS, 407, 225, arXiv: 1002.4708, doi: 10.1111/j.1365-2966.2010.16952.x

Vegetti, S., Koopmans, L. V. E., Bolton, A., Treu, T., & Gavazzi, R. 2010b, MNRAS, 408, 1969, arXiv: 0910.0760, doi: 10.1111/j.1365-2966.2010.16865.x

Vegetti, S., Lagattuta, D. J., McKean, J. P., et al. 2012, Nature, 481, 341, arXiv: 1201.3643, doi: 10.1038/nature10669

Verma, A., Collett, T., Smith, G. P., Strong Lensing Science Collaboration, & the DESC Strong Lensing Science Working Group. 2019, arXiv: 1902.05141

Vogelsberger, M., Zavala, J., Cyr-Racine, F.-Y., et al. 2016, MNRAS, 460, 1399, arXiv: 1512.05349, doi: 10.1093/mnras/stw1076

Vogelsberger, M., Zavala, J., & Loeb, A. 2012, MNRAS, 423, 3740, arXiv: 1201.5892, doi: 10.1111/j.1365-2966.2012.21182.x

Vogelsberger, M., Zavala, J., Schutz, K., & Slatyer, T. R. 2019, MNRAS, 484, 5437, arXiv: 1805.03203, doi: 10.1093/mnras/stz340

Wald, A. 1943, Transactions of the American Mathematical Society, 54, 426

Warren, S. J., & Dye, S. 2003, ApJ, 590, 673, arXiv: astro-ph/0302587, doi: 10.1086/375132

Wechsler, R. H., & Tinker, J. L. 2018, ARA&A, 56, 435, arXiv: 1804.03097, doi: 10.1146/annurev-astro-081817-051756

Wilks, S. S. 1938, Annals Math. Statist., 9, 60, doi: 10.1214/aoms/1177732360

Wong, K. C., Suyu, S. H., Chen, G. C. F., et al. 2019, arXiv: 1907.04869

Yoshida, N., Springel, V., White, S. D. M., & Tormen, G. 2000, ApJL, 544, L87, arXiv: astro-ph/0006134, doi: 10.1086/317306

Zahid, H. J., Sohn, J., & Geller, M. J. 2018, ApJ, 859, 96, arXiv: 1804.04492, doi: 10.3847/1538-4357/aabe31

Zavala, J., Vogelsberger, M., & Walker, M. G. 2013, MNRAS, 431, L20, arXiv: 1211.6426, doi: 10.1093/mnrasl/sls053