Gravitational Potential from small-scale clustering in action space: Application to Gaia DR2

Tianyi Yang1,2,3*, Supranta S. Boruah1,3,3,4†, Niyesh Afshordi1,2,3‡

1Waterloo Centre for Astrophysics, University of Waterloo, Waterloo, ON, N2L 3G1, Canada
2Department of Physics and Astronomy, University of Waterloo, Waterloo, ON, N2L 3G1, Canada
3Perimeter Institute of Theoretical Physics, 31 Caroline St. N., Waterloo, ON, N2L 2Y5, Canada
4Department of Applied Mathematics, University of Waterloo, Waterloo, ON, N2L 3G1, Canada

8 August 2019

Abstract

Most measurements of mass in Astronomy that use kinematics of stars or gas rely on assumptions of equilibrium that are often hard to verify. Instead, we develop a novel idea, first proposed by Sanderson et al. (2015), that uses the clustering in action space, as a probe of underlying gravitational potential. The correct potential should maximize small-scale clustering in the action space. We provide a first-principle derivation of likelihood using the two-point correlation function in action space, and test it against simulations of stellar streams. We then apply this method to the 2nd data release of Gaia, and use it to measure the radial force fraction $f_R$ and logarithmic slope $\alpha$ of dark matter halo profile. We investigate stars within 9-11 kpc and 11.5-15 kpc from Galactic centre, and find $(f_R, \alpha) = (0.297 \pm 0.007, 1.60 \pm 0.07)$ and $(0.297 \pm 0.009, 1.49^{+0.07}_{-0.07})$ respectively. We also confirm that the set of parameters that maximize the likelihood function correspond to the most clustering in the action space. The best-fit circular velocity curve for Milky Way potential is comparable but $\sim 4-17\%$ lower than previous studies that use other methods.

Our work provides a clear demonstration of the full statistical power that lies in the full phase space information, relieving the need for ad hoc assumptions such as virial equilibrium, circular motion, or steam-finding algorithms.

Key words: Dark Matter Halo – Milky Way – Tidal Streams

1 INTRODUCTION

Understanding the nature of dark matter is one of the most significant challenges in the 21st century for both physicists and astrophysicists. While the cold dark matter (CDM) paradigm has been the most popular model of dark matter, observational tensions on the galactic scale (Bullock & Boylan-Kolchin 2017; Del Popolo & Le Delliou 2018) and non-detection of dark matter particles in the ground-based experiments have led scientists to examine alternative possibilities. Part of the reason why it is difficult to probe the nature of dark matter is that (as far as we can tell) it only interacts with ordinary matter via gravity; only affecting astrophysical observations on very large scales ($\gtrsim$ few kpc). On these scales, extracting the full 6-dimensional phase space information (that is necessary to infer dark matter mass unambiguously), has been a difficult task.

Nonetheless, European Space Agency (ESA)’s Gaia mission has recently started probing the kinematics of the Milky Way stars with unprecedented precision (Gaia Collaboration et al. 2016). Gaia is a space-based observatory launched by European Space Agency (ESA) in December 2013, which aims at constructing the largest catalogue of 3d positions and velocities of the Milky Way stars, using Astrometric techniques. The first data set (Gaia DR1), released in September 2016, did not have the measurement of radial velocities. However, the Data Release 2 (Gaia DR2), which was released in 2018, with more complete magnitude measurement and longer span compared to Gaia DR1, contains the proper motion, parallax as well as the radial velocity information of more than 7 million stars (Gaia Collaboration et al. 2018). The proliferation of data on the kinematics of Milky Way’s stars has therefore opened a new avenue to probe the structure of the Milky Way potential and the nature of dark matter.

In this work, we propose a new method to constrain the potential of the Milky Way and apply it to two six dimensional subsamples of stars Gaia DR2. Our method is based on maximizing the statistical clustering of the stars in the space of actions. The current theories and observational evidence suggest that the growth of structure in our universe is hierarchical, where smaller structures merge to form bigger ones. During the formation of galaxies, however, the smaller structures are tidally disrupted and due to various relaxation mechanisms at play, the memory of their common origin
in configuration space is erased. This makes identifying stars with common origin nearly impossible. Nevertheless, the information regarding their common origin may still be present in the phase space of action variables. Since action variables are conserved, due to their common origin, the action variables of the various structures would be clustered on small scales (in action space). This principle has previously been proposed to infer the potential by maximizing Kullback-Liebler divergence (KLD) or “relative entropy” in the action space (Kullback & Leibler 1951; Sanderson et al. 2015, 2017). In these two studies, the viability of this idea was tested in simulated stellar distribution, where they successfully recover a spherical logarithmic (Sanderson et al. 2015) and a spherical NFW profile (Sanderson et al. 2017) using this method. The same method has also been applied to constrain the parameters of the globular cluster. By minimizing the KL entropy in the phase-space, Buckley et al. (2019) successfully constrain the mass and the King radius of the simulated M4 globular cluster, which provides another proof that the original parameters of a system can be recovered if the true phase-space information are found. Maximizing the statistical clustering in the action space does not require identification of the membership of any star, which is one of the major merits for this method.

This principle can be used to infer the potential of the Milky way. If the action variables are estimated using the incorrect potential, the resulting quantity will not be conserved with the dynamical evolution. Therefore, the clustering of the stars, in action space, on small scales will be destroyed if we use the wrong potential. Conversely, using the correct potential will maximize the small scale clustering in this space. We provide a first-principle derivation that the likelihood for the potential can be expressed as an integral (or KLD) over the 2-point correlation function in the action space, and test it using simulations of mock streams. Then, as an example, we fit a power-law dark matter profile (assuming a fixed form for bulge and disk component) to Gaia DR2, and compare our results to those that use other methods.

The paper is structured as follows: In Section 2, we introduce the required theoretical background for our method, including our parameterized models of the Milky Way potential and the computation of action variables. In Section 3, we briefly discuss the data sets we used and the selection cuts imposed on the raw data. Details related to the two-point correlation function and the likelihood test we used in the action space are presented in Section 4. To check the viability of our method, we first apply it to simulations with only stream stars, where the streams are simulated using the Python package galpy (Version 1.3.0. See Bovy 2015, for more details). Then we proceed to apply our method to the real observations taken from Gaia DR2 with the selection cuts listed in Section 3 from two different radial bins. The results of our analysis is presented in Section 5. In Section 6, we discuss the shortcomings and future possibilities of our method before concluding in Section 7.

2 THEORY

2.1 Modelling the Milky Way Potential

We use a parameterized model for the Milky Way potential which can be regarded as a combination of the bulge, the disk and the dark matter halo. The specific model we are using is a slight modification to the MWPotential2014 potential in galpy, which is assumed to be a good approximation to the Milky Way potential.

The bulge is modelled with a power-law density with an exponential cut-off:

\[
\rho_b(r) \propto r^{-1.8} \exp \left[ -\frac{r}{1.9 \text{kpc}} \right].
\]

(1)

The contribution from the central bulge is negligible at radius greater than 9 kpc, but we still include this component in the model for completeness.

The disk is modelled as a Miyamoto-Nagai Potential profile, but with fixed parameters (Bovy 2015):

\[
\Phi_d(r, z) \propto -\frac{1}{\sqrt{R^2 + [3 \text{ kpc} + \sqrt{z^2 + (0.28 \text{kpc})^2}]}}.
\]

(2)

where \(R\) and \(z\) are the radial and vertical galactocentric cylindrical coordinates, respectively (\(z = 0\) is the plane of the galaxy).

Finally, we model the dark-matter halo profile as a spherical power-law profile

\[
\rho_{dm}(r) \propto r^{-\alpha},
\]

(3)

which is also a built-in potential expression categorized as PowerSphericalPotential in galpy package. Note in MWPotential2014, the halo potential is characterized as NFW profile (NFW: Navarro et al. 1996). Instead, in this work, we use a power law potential for simplicity. Also, as we shall see later, the data set for constraining the potential does not span over a large range, so a localized power law potential should be a good approximation for NFW profile.

Finally, we set the normalizations of bulge, disk, and dark matter components, \(\rho_b, \Phi_d, \rho_{dm}\), so that at \(R = 8 \text{ kpc}\) and \(z = 0\), the fraction of radial force due to dark matter is \(f_b\), while the ratio of force due to stellar bulge to disk is fixed to be 1:12. Therefore, in the end we are left with two free parameters, \(f_b\), and the power-law index of the density profile, \(\alpha\), which we aim to constrain using our method.

2.2 Action Variables

Regular (i.e. non-chaotic) orbits in the galactic potential should admit 3 integrals of motion (e.g., Mo et al. 2010). However, finding these integrals of motions in terms of the phase space coordinates could be a difficult task. Nevertheless, it may be possible to find canonical transformations so that, finding integrals of motions in these coordinates are easy. One particularly convenient system of canonical variables is the so-called action-angle variables [denoted by \((\theta, J)\)], where the canonical momenta \(J\), or actions, are also the integrals of motion. The angle variables \(\theta\) are periodic in orbital torus, where an increase of \(2\pi\) in the angle would be associated with the same point in phase space. The action conjugate to this angle is then defined as:

\[
J_i = \frac{1}{2\pi} \oint_{\gamma_i} \mathbf{p} \cdot d\mathbf{q}.
\]

(4)

where \(\gamma_i\) is the orbit section where the \(i\)-th angle, \(\theta_i\), increases from 0 to \(2\pi\).

Finding a canonical transformation to transform to the action-angle variables provide a convenient convention to define integrals of motion for an integrable potential.
2.2.1 Calculating the action variables

The action-angle variables provide a convenient way to find the integrals of motion. However, finding closed analytic forms for the actions is only possible for a few special potentials. Therefore, in galactic dynamics, we often have to rely on approximate methods which involve integration of the orbits, e.g., adiabatic approximation (Binney 2010) or the torus construction method (Binney & McMillan 2016). Here, we use the ‘Stäckel approximation’ (Binney 2012), which is implemented in galpy.

Stäckel potentials are a special class of potentials where the Hamiltonian can be written in a separable form, using a canonical transformation. The Stäckel potentials are expressed in the spherical coordinates, \((u, v)\) which are related to cylindrical coordinates through

\[
R = \Delta \sinh u \sin v; \quad z = \Delta \cosh u \cos v. \quad (5)
\]

In these coordinates, the Stäckel potential takes the form

\[
\Phi(u, v) = \frac{U(u) - V(v)}{\sinh^2 u + \sin^2 v}. \quad (6)
\]

The radial and the azimuthal actions can be expressed in closed analytic forms

\[
J_r = \frac{1}{\pi} \int_{u_{\text{min}}}^{u_{\text{max}}} p_r(u) du, \quad (7)
\]

\[
J_z = \frac{2}{\pi} \int_{v_{\text{min}}}^{v_{\text{max}}} p_z(v) dv, \quad (8)
\]

where,

\[
p_r^2 = \frac{2\Delta^2}{E} E \sinh^2 u - I_3 - U(u) - \frac{L_z^2}{2\Delta^2 \sinh^2 u}, \quad (9)
\]

\[
p_z^2 = \frac{2\Delta^2}{E} \sin^2 v + I_3 + V(v) - \frac{L_z^2}{2\Delta^2 \sin^2 v}. \quad (10)
\]

In the above relations, \(E\) is the energy of the orbit and \(I_3\) is a third integral of motion (apart from the energy and the azimuthal angular momentum) which can be expressed analytically in terms of the conjugate variables and the momenta.

The Stäckel approximation involves approximating any nominal potential of the Milky Way as a Stäckel potential, and use this potential to find approximate action variables. To do this, we need to find the effective confocal length, \(\Delta\) in Equation 5.

A prescription for this was given by Sanders (2012). To obtain the value of \(\Delta\), we make use of the fact that \(\Phi(\sinh^2 u + \sin^2 v)\) is a separable function of \(u\) and \(v\). Therefore, any mixed derivative of this quantity must vanish. Using the model potential, \(\Phi_{\text{model}}\) we obtain

\[
\frac{\partial^2}{\partial u \partial v} (\sinh^2 u + \sin^2 v)\Phi_{\text{model}} \approx 0. \quad (11)
\]

This equation can then be solved for \(\Delta\) in terms of the derivatives of the potential, \(\partial \Phi / \partial R, \partial \Phi / \partial \Omega, \partial^2 \Phi / \partial R^2, \partial^2 \Phi / \partial \Omega^2\) and \(\partial^3 \Phi / \partial R \partial \Omega\).

The relation between \(\Delta\) and the derivatives were obtained in Sanders (2012):

\[
\Delta^2 = z^2 - R^2 + \left(3 R \frac{\partial \Phi}{\partial \Omega} - 3z \frac{\partial \Phi}{\partial R} + 2 \left( \frac{\partial^2 \Phi}{\partial R^2} - \frac{\partial^2 \Phi}{\partial R \partial \Omega} \right) \right)^2. \quad (12)
\]

This algorithm is implemented using the python package galpy.

2.3 Clustering in the action space

Structure in the \(\Lambda\)CDM model of cosmology is formed hierarchically where smaller structures merge to form bigger ones. However, the memory of their initial conditions are quickly erased due to the different relaxation mechanisms such as phase mixing and violent relaxation. However, if we have any integral of motion, it will be conserved during the orbital evolution of these stars. In particular, the action variables will remain conserved as long as the host potential evolves adiabatically. When the smaller structures are tidally disrupted in the Milky Way potential, their spread the action space is much smaller than the rest of the stars in the Milky Way. Therefore, we expect the small scale structure of the action space to contain the hierarchical tidal disruption/assembly history of the Milky Way (Afshordi et al. 2009).

Here, we would like to exploit this principle to infer the correct potential of the Milky Way. We first use the parametric form of the Milky Way potential presented in Section 2.1 to estimate the action variables using the ‘Stäckel approximation’ (presented in Section 2.2). If we use the incorrect potential, the estimated action variables are not true integrals of motion, and thus the small scale clustering due to hierarchical structure formation will be destroyed. On the other hand, using the correct potential will preserve the clustering on smaller scales. Therefore, maximizing the clustering with respect to the parameters of the potential should yield a reasonable estimate for the gravitational potential of the Milky Way (Sanderson et al. 2015, 2017).

3 DATA SET

The kinematics of stars in this paper are obtained from the Data Release 2 (DR2) of the Gaia mission, which was released in April 2018. All data can be accessed though Gaia Archive\footnote{Gaia Archive webiste: http://gea.esac.esa.int/archive/}. With the aid of the Gaia radial velocity spectrometer (Cropper et al. 2018), we can directly obtain the full six-dimensional phase space information of approximately 7 million stars. For the sample we used for our analysis, we make a few simple quality cuts in order to avoid the stars with large errors on the parallax or proper motion measurement. We shall impose the cut that the relative error on the parallax, and the proper motion are less than 20%:

\[
\frac{\Delta \rho}{\rho} < 0.2, \quad \frac{\Delta \mu_{\rho}}{\mu_{\rho}} < 0.2, \quad \frac{\Delta \mu_{\alpha}}{\mu_{\alpha}} < 0.2, \quad \frac{\Delta V_{\text{radial}}}{V_{\text{radial}}} < 0.2. \quad (13)
\]

where, \(\rho, \mu_{\rho}, \mu_{\alpha}, V_{\text{radial}}\) are the parallax, proper motion in the right ascension direction, the proper motion in the declination direction, and radial velocity respectively. Here, \(\Delta\) denotes the measurement error in each of these quantities. After applying the selection cuts mentioned above to the raw data, we are left with around 5.6 million stars in the data sample for further analysis\footnote{Note that for the actual data analysis, there are additional cuts for halo vs. all stars, and radial distribution, which will be discussed in Section 5.2.}.

Even though the 20% cut (while common as in Astraatmadja & Bailler-Jones (2016); Schönrich & Aumer (2017); Schönrich et al. (2019)) is ad hoc, we further verify that this choice has little effect on our results, as stars with large uncertainties in their phase space coordinates are unlikely to form close pairs in action space.

This data allow us to constrain the two-parameter power law potential in Equation 3 by using the likelihood test, which we discuss in the next section. As calculations are conducted by using the built-in functions in galpy, the inputs for most of the functions we used are in cylindrical coordinates. For the convenience of computation, all calculations are done in galactocentric coordinate system, and
the coordinate transformation are as well handled by the built-in functions in galpy library (Bovy 2015).

4 METHOD

We are interested in the small scale clustering of the stars in the action space. There are many different measures of clustering which are useful for different purposes. Here, we use two-point correlation function as our measure of clustering, which should be one of the most straightforward ones. However, to define the correlation function, we need to have a measure of the distance. While this choice is not unique, we shall use the following measure to find the distance of two stars in action space

\[ D = \sqrt{(\Delta J_R/\sigma_{J_R})^2 + (\Delta J_\phi/\sigma_{J_\phi})^2 + (\Delta J_z/\sigma_{J_z})^2}, \]

where \( \Delta J_i \) denotes the difference in the action coordinates of the two stars, while \( \sigma_{J_i} \)'s are standard deviations of \( J_i \)'s over all stars. The reason why the distance is not directly calculated using the difference between action variables but instead with the addition of the variance is discussed in Section 6.

The calculation of \( \sigma_{J_i} \) can be affected by the outliers in the raw data or numerical artifacts in galpy action approximation. Therefore, another constraint is added to effectively exclude outliers out of the sample with \( \frac{|J_i - \bar{J}_i|}{\sigma_{J_i}} > 3 \). We then use the remaining action variables that satisfy the above criterion to re-calculate the standard deviation.

For points distributed randomly with a uniform distribution in a three dimensional action space, the probability of finding pairs at a separation between \( D \) and \( D + dD \) is given by

\[ P(D)_{\text{uniform}} = D^2 dD. \]

However, the actual probability distribution \( P(ln D) \) will be different from \( P(ln D)_{\text{uniform}} \approx D^3 \) due to clustering in the action space. This clustering can be quantified using the 2-point correlation function \( \xi(ln D) \):

\[ 1 + \xi(ln D) \equiv \frac{P(ln D)}{P(ln D)_{\text{uniform}}} = \frac{D_{\text{max}}^3}{3D^3} \int_{-D_{\text{max}}}^{D_{\text{max}}} P(ln D') d ln D'. \]

where we used the fact that both \( P \) and \( P_{\text{uniform}} \) should integrate to unity over the range \( ln D \in (-\infty, ln D_{\text{max}}) \).

As it turns out, with certain assumptions, the statistical likelihood of any action-space distribution can be expressed in terms of \( \xi(ln D) \). The key idea here is to assume some small scale clustering in the action space is the Poisson sampling of a near-uniform background plus a random gaussian field. The correlation function of this random gaussian field encodes all the clustering information at small scale in the action space. This model is agnostic about the distribution function, \( f(J) \) and instead relates the likelihood to the correlation function in the action-space \( \xi(ln D) \), after marginalizing over all possible \( f(J) \)'s. More explicitly, we find that the log-likelihood for a potential is given by

\[ \ln L(f_0, \alpha | \text{data}) = \sum_{\text{pairings}} \left( \ln \left( 1 + \xi(ln D_{\text{pairing}}) \right) \right) = N_{\text{pairs}} \int_{-\infty}^{D_{\text{max}}} P(ln D) \ln \left( 1 + \xi(ln D) \right) d ln D. \]

of stars in the sample. A detailed derivation of this expression is presented in Appendix A.

We further note that relative entropy (Kullback & Leibler 1951) of the distribution \( P(ln D) \), with respect the uniform distribution, is defined as

\[ S_{\text{relative}} = -\int P(ln D) \ln \left( \frac{P(ln D)}{P_{\text{uniform}}} \right) d\ln D = -\frac{\ln L(f_0, \alpha | \text{data})}{N_{\text{pairs}}}, \]

i.e. the maximization of the likelihood function corresponds to minimizing the entropy relative to the uniform pair distribution. In other words, the best-fit values for the dark matter halo density produce the most non-uniform distribution of pairs in the action space. We should note that while this is similar to the criterion proposed by Sanderson et al. (2015), their relative entropy is based on phase space density in the action space \( f(J) \), while our derivation in Appendix A shows that likelihood depends on the relative entropy of the pair distance probability distribution \( P(ln D) \).

Another important point is the choice of \( D_{\text{max}} \), which is the scale of homogeneity in the action space background. We shall also discuss the dependence on this scale below.

5 RESULTS

5.1 Simulations

To validate our method, we simulate the orbits of a few stars in a known parameterized potential of the same form and then applied the above mentioned analysis to check if we can recover the true parameters of the potential. The simulation includes three groups of tidal stream stars with different initial conditions of progenitors. This is achieved by using the built-in modelling method in galpy package (Bovy 2014). One can specify the gravitational potential that stars evolve in, the method for action variables calculation, the initial conditions of progenitors, the velocity distribution of progenitors...
and the time when the disruption began. Initial conditions of the progenitors’ orbit for three streams are tabulated in Table 1. The header of the table is organized in the order of R, φ, z, v_R, v_T, v_z, velocity dispersion (σ_v) and the disruption time (t_{disrupt}). Each of the three streams consist of 3000 stream stars. With the simulated stellar trajectories, the action variables of stars can be calculated based on the Stäckel approximation as explained in Section 2. We now wish to test whether our proposed likelihood function (17) leads to constraints that are consistent with parameters that are used in our simulated host potential.

There are two free parameters in the expression of the dark matter halo density profile (Equation 3), f_h which fixes the normalization, and the logarithmic slope α. We choose the mass fraction of the halo f_h = 0.35 based on Table 1 in Bovy (2015), and we produce two sets of stream simulations with different choices of α = 1.70 and 2.00 respectively. The progenitor stars are evolved in these two host gravitational potentials respectively. We then compute the action variables on a grid in the (f_h, α) space, and compute the corresponding likelihood function using Equation (17). The likelihood functions evaluated with 9000 simulated stream stars for both potentials are shown in Figure 2 (assuming ln D_{max} = −1). For each case, we find clear constraints on both parameters as expected, in reasonable agreement with input parameters of the simulations, subject to caveats that we discuss next.

To determine the location and the uncertainties of the measurements at each D_{max}, we fit the log-likelihood distribution with a quadratic function around its maximum. The assumption made by this procedure is that the likelihood only has a single peak that can be approximated by a gaussian distribution. In order to check the validity of this assumption, we plot the posterior of parameter at the location of the maximum likelihood (Equation 17) based on clustering in action space can yield reasonable constraints on simulated potentials, subject to small systematic errors of 1% (4%) on normalization and logarithmic slope. There also does not seem to be any significant dependence on the maximum separation of included pairs in action space D_{max}. A more exciting step is to apply our method to real Gaia DR2 data to see how well it can constrain the Milky Way potential, which we shall do next.

5.2 Real data

After confirming the reliability of the method, we proceed with our analysis using real data from Gaia DR2. The criteria for data selection were already discussed in Section 3. Let us now introduce some additional selection cuts. Recall that in the derivation of likelihood function (Appendix A), we assume the stellar distribution in the action space is a uniform background plus fluctuations. This assumption as more appropriate for halo stars in our galaxy, as disk stars have J_z = 0. Additionally, as we are trying to constrain dark matter profile, which mostly occupies the Milky Way halo, halo stars should be better candidates compared to disk stars. Due to these considerations, we only select stars that have vertical distance to the galactic plane > 1 kpc³.

Figure 3 shows the galactocentric distance and the tangential velocity distribution (in cylindrical coordinate) for all the data with relative measurement error smaller than 20% and |z| > 1 kpc. There are around 337,022 stars in total. As expected, the peak of radial distribution is around solar radius and the peak of v_T distribution is around the value of circular velocity at solar radius. Here, we assume we take R_⊙ = 8 kpc, V_{φ,⊙} = 220 km/s, but shall discuss this choice further in Sec. 6. Due to the limitation of the computational time, another galactocentric radius cut is also applied to the data: We choose two different radial ranges 9 kpc < R < 11 kpc (hereafter real-data-9-11) and 11.5 kpc < R < 15 kpc (hereafter real-data-115-15). After applying all of these cuts, there are approximately 57,000 and 15,000 stars in each sample, respectively.

Now, taking the NFW profile as reference, the expected value of the α in the power law density profile should be within 1 to 3: For r >> r_s, the density is proportional to r^{-3}, while for r ≲ r_s, it goes to r^{-1}. Furthermore, Bovy & Rix (2013) used the assumption of Jean’s equilibrium for G-dwarfs from SEGUE survey to constrain α < 1.53 (at 95% confidence) between R=4 kpc and 9 kpc. Therefore, to allow for a conservative prior, we consider the range:

\[ 0.5 < \alpha < 2.5, \]
\[ 0.25 < f_h < 0.55, \]  

for our dataset within 9 kpc to 15 kpc.

The likelihood plots showing the constraints on the mass fraction and the index for both radial samples are shown in the left panels of Figure 4. Although calculated within the same ln D_{max} = −1, the log-likelihood values for “real-data-9-11” is larger than those of “real-data-115-15” as there are more stellar pairs included in the

³ We will discuss the effect of this cut, as well as the measurements error cuts on the final results later in Section 6.
Table 1. Initial conditions of progenitor for the generation of stream stars (where galactocentric radius and velocity are normalized by solar radius value, and \(t_{\text{disrupt}}\) is rescaled by the time unit defined by bovy_conversion.time_in_Gyr(220, 8.).)

| Stream Number | \(R\) | \(\phi\) | \(z\) | \(V_R\) | \(V_T\) | \(V_Z\) | \(\sigma_v\) (km/s) | \(t_{\text{disrupt}}\) |
|---------------|------|------|-----|-------|-------|-------|-----------------|----------------|
| Stream 1      | 1.56 | 0.12 | 0.89| 0.35  | −1.15 | −0.48 | 0.3             | 50             |
| Stream 2      | 1.00 | −0.05| 0.001| −0.60 | 0.51  | 0.0086| 0.3             | 50             |
| Stream 3      | 1.20 | −0.05| −1  | −0.30 | 0.51  | 0.16  | 0.3             | 50             |

Figure 2. Top panel: likelihood test and error bar plot for case \([f_h = 0.35, \alpha = 1.70]\). Bottom panel: likelihood test and error bar plot for case \([f_h = 0.35, \alpha = 2.0]\). The maximum likelihood gives constraints on the parameters \(f_h = 0.35, \alpha = 1.63\) for the first case and \(f_h = 0.35, \alpha = 1.95\) for the second case. The initial set of parameter is indicated as black plus sign on the likelihood plot for either case. Error bars are determined based on the paraboloid fitting, where the black points with black error bars are determined with whole 9000 stars simulated from three different streams (initial conditions are listed in Table 1), while the red points with red error bars are a randomly-chosen sub-sample of stars. The ultimate results for both parameters are not significantly changed with the variation of \(D_{\text{max}}\) and the size of the sample. Based on this figure, an approximate 4% systematic discrepancy might be expected between the likelihood evaluation and the actual value of \(\alpha\), and a 1% systematic discrepancy might exist in the \(f_h\) evaluation. The \(D_{\text{max}}\) values are taken as \(\ln D_{\text{max}} = -1.0\) for both cases.

more nearby sample. Furthermore, the errorbar plots in Figure 4 show how the best fits (black points) and median\(\pm 1\)-\(\sigma\) constraints on parameters (green errorbars) vary with different choices of \(D_{\text{max}}\). We see that for \(\ln D_{\text{max}} \leq -1\), the \(f_h\) and \(\alpha\) constraints are stable and robust to the choice of free parameter \(D_{\text{max}}\).

We notice that for the more distant sample “real-data-115-15”, the convergence for both parameters is not reached until \(\ln D_{\text{max}} \approx -1\). Taking this into account, we treat -1.0 as our final choice of \(D_{\text{max}}\), and to be consistent, use the same value of \(D_{\text{max}}\) for both Gaia samples.

Having seen statistical constraints on both parameters from the likelihood plots, we would like to evaluate the true uncertainties of the measurements. However, the determination of uncertainties is more subtle compared with the simulations. Unlike simulations, where we found the posterior distribution of parameters had a sharp gaussian peak, we notice that the likelihood 2D plots have multiple peaks for real Gaia data. This can be seen more clearly in the 1D posterior distributions in Figure 5, where (depending on the choice of \(D_{\text{max}}\)) there can be multiple peaks. As a result, it is no longer appropriate to simply assume the likelihood distribution is approximated by a gaussian. In particular, the jump in the best fit \(f_h\) around \(\ln D_{\text{max}} \approx -0.7\) in “real-data-115-15” sample is due to the change in relative heights of the two main peaks in posteriors shown in the top panel of Figure 5. Therefore, we calculated the median of the parameters using the full posterior distribution within our prior range (Equation 20), as it is a more robust statistical estimator than average whenever multiple peaks or outliers are presented in the distribution. The 68% confidence interval around the median, which can be also computed from the posterior distribution, is treated as the 1\(\sigma\) error on the parameter.

As can be seen in the errorbar plots in Figure 4, the best-fit parameters (black points, corresponding to the peaks of likelihood) are all consistent with the median values within 1\(\sigma\) deviation (blue points with green errorbars), and the errorbars become larger for smaller \(D_{\text{max}}\), where fewer stellar pairs are included. We notice
that, when choosing \( \ln(D_{\text{max}}) \) as -0.88, the constraints we get under both situations are summarized in Table 2.

Comparing the constraints on \( f_h \) and \( \alpha \) at same \( D_{\text{max}} \), the error bars determined with stars ranging from 9-11 kpc are relatively smaller than those with 11.5-15 kpc as there are more pairs included in the case of “real-data-9-11” at same \( D_{\text{max}} \).

Furthermore, recall that from simulated measurements in Section 5.1, we do expect an additional 4% percent systematic discrepancy for index measurements and an 1% percent offset in mass fraction measurement. We do include these estimates in Table 2 as systematic errors, which can be combined with our stochastic errors to obtain the total expected uncertainties.

Let us now perform the same consistency checks we did in Section 5.1 for simulated data, and see how stars from real data are distributed in the action space. Figures 6 and 7 shows how (2D projections of) the stellar distribution in the action space, as well as its two-point correlation function change as we vary \( f_h \) or \( \alpha \) in the Milky Way halo potential. As expected, the correlation function
The posterior distribution of \( f_h \) (upper panel) and \( \alpha \) (lower panel) at three different values of \( D_{\text{max}} \) calculated using stars with radial coverage from 11.5-15 kpc. Unlike simulation, the appearance of multiple peaks is obvious in the probability distribution and the peaks could vary with \( D_{\text{max}} \) as well. This indicates that the paraboloid fitting can not be used for uncertainties determination.

**Table 2.** Constraints on normalization and logarithmic slope of dark matter profile \((f_h, \alpha)\) for both radial samples with selection cuts. Here, we summarize all sources of measurement errors: the stochastic errors estimated from the posterior distribution of parameters, the systematic errors from simulation, and the total error given by the root of stochastic errors squared plus systematic errors squared.

| Error Type | \([9.0 \text{ kpc} < R < 11.0 \text{ kpc}]\) | \([11.5 \text{ kpc} < R < 15.0 \text{ kpc}]\) |
|------------|---------------------------------|---------------------------------|
| \(f_h\)    | \(0.297\) \(\pm\) 0.006        | \(0.306\) \(\pm\) 0.003        |
| \(\alpha\) | \(1.656\) \(\pm\) 0.040        | \(1.606\) \(\pm\) 0.066        |

**Figure 5.** The posterior distribution of \( f_h \) (upper panel) and \( \alpha \) (lower panel) at three different values of \( D_{\text{max}} \) calculated using stars with radial coverage from 11.5-15 kpc. Unlike simulation, the appearance of multiple peaks is obvious in the probability distribution and the peaks could vary with \( D_{\text{max}} \) as well. This indicates that the paraboloid fitting can not be used for uncertainties determination.

In this situation, stars are extended along \( J_\phi \) axis with no distinct \( J_R, J_z \) contributions, which indicates the property of circular (or disk) motion for most stars. This is not surprising as can be seen from the tangential velocity distribution in Figure 3; the disk stars are still the dominant component in our real data samples even though we performed a \(|z| > 1 \text{ kpc} \) cut. Just as in the simulations, stars within the potential that maximizes the small-scale clustering statistics (Figure 7) present the most compact distribution in the action space. The fact that stellar distribution reduces to circular motion for the best-fit potential is in practice consistent with the traditional assumption of circular motion for disk stars, in order to estimate the mass of Milky Way galaxy. However, our method does not explicitly make this assumption, and thus can account for deviations from circular motion, effectively combining (thin+thick) disk+halo stars.
Figure 6. Stellar distribution in the $J_R$ and $J_\phi$ 2D projected plane varying with different choices of potential, where $J_R$ and $J_\phi$ are defined as $J_R/\sigma_{J_R}$ and $J_\phi/\sigma_{J_\phi}$, i.e. radial and angular action variables normalized by their standard deviations over all stars in the sample. First two rows show the stellar distribution for the first case of read data with fixed $f_h(\alpha)$ in the first (second) row. Last two rows show the result for the second case of real data. Interestingly, while approaching the potential that maximizes the likelihood, stars are tend to be more disk-like and display the properties of circular motion.

6 DISCUSSION

In the previous analysis, we used some measurement error cuts and a vertical distance cut to real data. However, selection cuts to the raw data could cause unexpected biases in the measured parameters. To investigate this issue, we randomly choose 90,000 stars from 9-11 kpc sample\(^4\) and take all data from 11.5-15 kpc (88,370 stars in total), without imposing any of the previous error or distance cuts\(^5\). The errorbar plots are shown in Figure 8. For comparison, we also overplot the results obtained before using the sample with selection cuts. As expected, at same value of $D_{\text{max}}$, the uncertainties on parameters are significantly reduced when using the data samples without selection cuts. To be consistent, for both radial ranges, we still take $\ln(D_{\text{max}}) \approx -1$ and check the corresponding constraints on $f_h$ and $\alpha$. The results are tabulated in Table 3.

Compared with the constraints at the same $D_{\text{max}}$ obtained previously but with error selection cuts with the results from full data

\(^4\) There are 544,829 stars in total, but we only choose a subset of the catalogue due to the limitation of computational time. This procedure does not affect the final results. Including more stars only narrows down the uncertainties in parameters. However, as discussed later, this could not do help to explain the problem we met.

\(^5\) However, we did apply some minimal cuts to the raw data in order to get rid of the unreliable observations, including $|z| < 10$ kpc and the absolute values of all three components of velocity in cylindrical coordinates are smaller than 500 km/s

MNRAS 000, 1–77 (2019)
Figure 7. Correlation function \( \frac{P(\ln D)}{D^3} \) as a function of the distance in the action space in natural logarithm scale. The purpose of this figure is to check how the two-point correlation function varies with different choices of potential and whether it is extremized around the set of parameter that maximize the likelihood function. Top panel: the behaviour of two-point correlation function for the first case of real data with fixed \( f_h(\alpha) \) on the left (right). Bottom panel: the behaviour of two-point correlation function for the other case with fixed \( f_h(\alpha) \) on the left (right). Different colors indicate the values of \( \ln(\frac{P(\ln D)}{D^3}) \) at different choices of potential.

| Error Type            | (no |z| cut)              | [9.0 kpc <R<11.0 kpc] | [11.5 kpc <R<15.0 kpc] |
|-----------------------|-----------------------|------------------------|------------------------|
|                       | \( f_h \) \( \alpha \) | \( f_h \) \( \alpha \) | \( f_h \) \( \alpha \) |
| Stochastic            | \( \pm 0.005 \) \( \pm 0.005 \) | \( \pm 0.003 \) \( \pm 0.003 \) | \( \pm 0.003 \) \( \pm 0.003 \) |
| Systematic            | \( \pm 0.006 \) \( \pm 0.006 \) | \( \pm 0.006 \) \( \pm 0.006 \) | \( \pm 0.006 \) \( \pm 0.006 \) |
| Total                 | \( \pm 0.006 \) \( \pm 0.007 \) | \( \pm 0.006 \) \( \pm 0.007 \) | \( \pm 0.006 \) \( \pm 0.007 \) |

Table 3. Constraints on normalization and logarithmic slope of dark matter profile \((f_h, \alpha)\) for both radial samples without selection cuts. Here, we summarize all source of measurement errors: the stochastic errors estimated from the posterior distribution of parameters, the systematic errors from simulation, and the total error given by the root of stochastic errors squared plus systematic errors squared.

set, we find a consistency in \( f_h \) constraint for both radial ranges: the estimates of mass fraction are within 1\( \sigma \) standard deviation for both samples. However, for \( \alpha \), we notice a 6% systematic discrepancy within 9-11 kpc, and an 11% discrepancy for the 11.5-15 kpc. For both radial ranges, the index estimates for the uncut sample are lower than those of the cut sample.

This systematic shift is primarily due to the selection cut to the vertical distance, \( z \). Although some measurement error cuts are also imposed on the raw data, \( z \) distance cut seems to be the most severe: 80% of raw data survives the measurement error cuts, while only around 6% remain after the \( |z| \geq 1 \) kpc cut is imposed. One possible reason for this systematic difference could be the inaccuracy of the simple analytic model for the disk potential used in Equation 2. It remains to be seen whether a more realistic model (e.g. using other datasets), or including the disk parameters in the likelihood marginalization, could lead to more consistent (and realistic) estimates.

In order to account for additional potential systematic errors due selection cuts, we use the probability function defined in Appendix C, which yields our final constraints on the mass fraction of dark matter and the index in the localized density profile in Table 4.

After obtaining the constraints on both parameters in the dark matter halo density profile, we can translate them to less model-dependent constraints by computing the rotation curve (circular Keplerian velocity) of the Milky Way, as a function of distance from the centre. This result can then be compared to other studies that use different parameterizations and methods. To obtain a more robust estimation to the circular rotation curve, we evaluate the average and standard deviation of \( v_{\text{circ}}(R) = \sqrt{\frac{\partial \Phi}{\partial \ln R}} \) given the likelihoods found from our different Gaia samples (Equation 17) over our prior range of \( f_h \) and \( \alpha \) (Equation 20).

As the expression of disk potential is fixed (where we also fix
Figure 8. Errorbar plot using stars from 9 to 11 kpc (top) and 11.5 to 15 kpc (bottom) without selection cuts. Black points shows the constraints to the parameters by directly finding the maximum from the likelihood plot, and this gives $f_h = 0.298$ and $\alpha = 1.555$ for the nearby sample and $f_h = 0.297$ and $\alpha = 1.445$ for the other. The blue circle point and green line represent the median of parameter interpreted from the posterior distribution at each $D_{\text{max}}$. For comparison, the median values and 1$\sigma$ errorbars of two parameters determined from the sample with selection cuts are also over-plotted on the same figure, which are shown as blue star points and red lines.

$z = 0$ in Equation 2), its contribution to the total rotation curve (as well as that of the bulge) can be simply added to the halo part in quadrature. Figure 9 displays the circular velocity curve obtained from this work (purple and black solid line) and its corresponding uncertainty shaded area. Although the index parameter, $\alpha$ obtained from the sample with and without selection cuts are not consistent with each other at 1$\sigma$-level, the estimated circular velocity curves overlap$^6$. For comparison, the results obtained from Bovy (2015), McMillan (2017), Vasiliev (2019), and Eilers et al. (2019) are also shown in the same figure. McMillan (2017) used kinematic data from maser observations with (expected) near-circular motion to fit a Milky Way model with an NFW spherical halo, a stellar and gas disk plus a central bulge. Using a nearly identical model, Vasiliev (2019) assumed Jeans equilibrium of Milky Way globular clusters in Gaia data to constrain the gravitational potential. Eilers et al.

$^6$ Recall we used a slightly different method when finding the constraints to the potential due to the multi-modal of posterior distribution and the limitation of prior. However, when ln($D_{\text{max}}$) is chosen to be around -1, we checked that the best-fit estimation of parameters from weighted average or median fitting are consistent with each other within 1$\sigma$ deviation.
Figure 9. The rotation curves of halo part only (left) and the total rotation curves (right) calculated from different potential models. Result for this work is indicated as purple solid line (for the sample with selection cuts) and black solid line (or the sample without selection cuts). The shaded area indicates the 1-σ uncertainty calculated based on the weighted likelihood (see text for more details). For comparison, results estimated from other works are over-plotted on the same figure. Our results are comparable to those of Bovy (2015) (green dashed line) but are systematically lower than three other studies by 9-17% [McMillan (2017): blue dashed line, Vasiliev (2019); black dashed line and Eilers et al. (2019): red points with errorbars]

|        | [9.0 kpc < R < 11.0 kpc] | [11.5 kpc < R < 15.0 kpc] |
|--------|--------------------------|---------------------------|
| \( h_0 \) | 0.297±0.006              | 0.306±0.006               |
| \( \alpha \) | 1.656±0.040               | 1.606±0.080               |
| \( (|z|>1 \text{kpc, stochastic error only}) \) | 1.656±0.040               | 1.606±0.080               |

Table 4. Constraints on dark matter halo parameters for both samples with/without selection cuts. The top four rows compare results for different cuts (including only, mostly independent, stochastic errors), while the final two rows is an attempt to combine these results, including the systematic errors introduced due to selection, as detailed in Appendix C.

Yang et al. (2019) also used Jeans equilibrium for Gaia luminous red-giant stars to determine the circular velocity of the Milky way over radial range 5 kpc < R < 25 kpc. Although we approximate the localized halo density profile as a simple power law, our result is still relatively close to (but still 4% lower than) the best-fit NFW dark matter potential found in Bovy (2015) (NMWPotential2014). However, the circular velocity (radial force) is about 9-17% (18-30%) smaller than the other three studies. However, compared with these studies, our method might be more robust as it does not rely on assumptions of circular motion or Jeans equilibrium, and can be equally applied to halo or disk stars.

An important consideration for comparison to other measurements of circular velocity is our choices of \( V_{\odot} \) and \( R_{\odot} \). While we made this choice for consistency with analysis of Bovy (2015), the measurements of \( V_{\odot} \) and \( R_{\odot} \) have been progressively improving. For example, Gravity Collaboration et al. (2018) finds the best-fit values of \( R_{\odot} = 8.122 \pm 0.031 \text{kpc} \) and \( V_{\odot} = 246 \pm 1 \text{km/s} \) (using the Sag. A* proper motion found in Reid & Brunthaler 2004). Using these values does increase our total circular velocity by \( \sim 5\% \). While this brings our measurement closer to others, it is not enough to relieve the discrepancy.

Let us now comment on our choice of distance (or metric) in the action space, Equation (14). The reason why we normalize action variables by their standard deviation to compute distance is partly due to the assumption we made in the likelihood derivation in Appendix A. Our derivation starts from a uniformly distributed background plus gaussian fluctuations which model clustering in the action space. Therefore, the structures we consider should be on smaller scale than the background distribution in the action space. Since the extent of the background could be different in different directions in the action space, the normalization has the effective role of making the distribution homogeneous and isotropic, at least for \( D \ll 1 \), i.e. close pairs.

As we noted in Sec. 1, there are other proposals to use the action-angle (or similar) variables to constrain the potential. Sanders & Binney (2013) use the correlations in the angle-frequency space for stars of a single stream to constrain gravitational potential. For a true potential, the angle and frequency differences of stars in a long narrow stream should lie along a straight line. An incorrect potential could cause a misalignment between the stream orbit and the underlying progenitor orbit. By minimizing this misalignment, which is potential-dependent, they manage to recover the expected constraints to a spherical logarithm potential using a simulated tidal stream. While this method uses more information (i.e. angle vari-

---

7 Even though it may appear that the discrepancy with Bovy (2015) is resolved by this shift, this is far from obvious, as they also adopt \( V_{\odot} \approx 220 \text{ km/s} \) and \( R_{\odot} = 8 \text{kpc} \).

8 Note that we ignore the covariance between different action variables, an assumption that can be easily relaxed in the future.

9 For an integrable system, frequencies can be simply thought as another coordinate system in the action space.
ables) than ours, and thus can be potentially more precise, it requires identifying only stream stars and relies on the assumption of a cold stream, which does limit its precision and accuracy. The methods introduced in Peñarrubia et al. (2012), Sanderson et al. (2015), Sanderson et al. (2017), and Buckley et al. (2019) are the closest compared with our methodology, which minimize relative entropy (or KLD) of a system in the space of action variables (or more generally, integrals of motion). However, these studies do not provide any proof that relative entropy of action-space density is the correct statistical representation of likelihood. Indeed, our derivation in Appendix A suggests that relative entropy of the distance distribution $P(\ln D)$ [rather than $f(J)$] is more directly related to the likelihood. On a more practical note, given that the density of stars (or pairs of stars) is discrete, the answer does depend on the coarse-graining procedure. However, since there are many more stellar stars than stars $(N(N - 1)/2 vs N)$, our likelihood computation is much more robust to coarse-graining. Furthermore, to our knowledge, none of these methods have yet been applied to real data.

Finally, to be fair, we should also highlight some of the caveats in our study. Several assumptions are made in the derivation of our likelihood test in Appendix A, most importantly that of a uniform distribution. Several assumptions are made in the derivation of our method to analyze two samples from Gaia DR2 over radial ranges of 9-11 kpc and 11.5-15 kpc, and studied the effect of selection cuts on the final results. Including all the known systematic errors, we find the parameters $(\mu, \sigma) = (0.297 \pm 0.007, 1.60 \pm 0.07)$ and $(0.297 \pm 0.009, 1.49^{+0.07}_{-0.07})$, for 9-11 kpc and 11.5-15 kpc respectively, for the median and 1-$\sigma$ uncertainty from the posterior distribution. For both simulations and real data, we can visually confirm that the potential that maximizes the likelihood function does indeed correspond to the largest two-point correlation function and most compact distribution in the action space, which again, demonstrates the reliability of our method.

To our knowledge, this is the first study that constrains the halo potential of the Milky Way using the action space clustering with real data. While more work is needed to fully understand the systematic error of this method (as discussed in Section 6), its sheer statistical power is formidable as it scales with the number of all the stars in the sample, and with proper calibrations can provide exquisite constraints on dark matter potential. Further improvements (or checks) may come from identification of streams beforehand or other criteria to separate disk and halo components (Bonaca et al. 2017; Helmi et al. 2017; Myeong et al. 2018; Necib et al. 2018). Additionally, in this study, we only varied the parameters in the dark matter halo density profile but kept the stellar disk potential fixed. More robust constraints, left for future work, requires varying the parameters in the disk potential as well, and possibly include other probes of stellar density.

ACKNOWLEDGEMENTS

We would like to thank Ana Bonaca, Jo Bovy, Charlie Conroy, and David Hogg for useful discussion.

This work was supported by the University of Waterloo, Natural Sciences and Engineering Research Council of Canada (NSERC), and the Perimeter Institute for Theoretical Physics. Research at the Perimeter Institute is supported by the Government of Canada through Industry Canada, and by the Province of Ontario through the Ministry of Research and Innovation.

This work has made use of data from the European Space Agency (ESA) mission Gaia (https://www.cosmos.esa.int/gaia), processed by the Gaia Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement.

REFERENCES

Afshordi N., Mohayaee R., Bertschinger E., 2009, Phys. Rev. D, 79, 083526
Astraatmadja T. L., Bailier-Jones C. A. L., 2016, ApJ, 832, 137
Binney J., 2010, MNRAS, 401, 2318
Binney J., 2012, MNRAS, 426, 1324
Binney J., McMillan P. J., 2016, MNRAS, 456, 1982
Bonaca A., Conroy C., Wetzel A., Hopkins P. F., Kereš D., 2017, ApJ, 845, 101
Bovy J., 2014, ApJ, 795, 95
Bovy J., 2015, ApJS, 216, 29
Bovy J., Rix H.-W., 2013, ApJ, 779, 115
Buckley M. R., Hogg D. W., Price-Whelan A. M., 2019, arXiv e-prints, p. arXiv:1907.00987
Bullock J. S., Boylan-Kolchin M., 2017, ARA&A, 55, 343
Cropper M., et al., 2018, arXiv e-prints, p. arXiv:1804.09369
Del Popolo A., Le Delliou M., 2017, Galaxies, 5, 17
Eilers A.-C., Hogg D. W., Rix H.-W., Ness M. K., 2019, ApJ, 871, 120
Gaia Collaboration et al., 2016, A&A, 595, A1
Gaia Collaboration et al., 2018, A&A, 616, A1
Gravity Collaboration et al., 2018, A&A, 615, L15
Helmi A., Veljanoski J., Breddels M. A., Tian H., Sales L. V., 2017, A&A, 598, A58
Kullback S., Leibler R. A., 1951, Ann. Math. Statist., 22, 79
McMillan P. J., 2017, MNRAS, 465, 76
Mo H., van den Bosch F. C., White S., 2010, Galaxy Formation and Evolution...
Yang et al.

Myeong G. C., Evans N. W., Belokurov V., Amorisco N. C., Koposov S. E., 2018, MNRAS, 475, 1537

Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563

Necib L., Lisanti M., Belokurov V., 2018, arXiv e-prints, p. arXiv:1807.02519

Peñarrubia J., Koposov S. E., Walker M. G., 2012, ApJ, 760, 2

Reid M. J., Brunthaler A., 2004, ApJ, 616, 872

Sanders J., 2012, MNRAS, 426, 128

Sanders J. L., Binney J., 2013, MNRAS, 433, 1826

Sanderson R. E., Helmi A., Hogg D. W., 2015, ApJ, 801, 98

Sanderson R. E., Hartke J., Helmi A., 2017, ApJ, 836, 234

Schönrich R., Aumer M., 2017, MNRAS, 472, 3979

Schönrich R., McMillan P., Eyer L., 2019, MNRAS, 487, 3568

Vasiliev E., 2019, MNRAS, 484, 2832
APPENDIX A: DERIVATION OF THE LIKELIHOOD FUNCTION

In this section, we discuss the derivation of the likelihood which we use to constrain the parameters of the Milky Way potential. Using Bayes’ Theorem, we have,

\[ P(\Theta|D) = \frac{P(D|\Theta)P(\Theta)}{P(D)} \]  

(A1)

where \( \Theta \) are the parameters we want to constrain, while \( D \) is the data we have available. In particular, in the context of constraining the Milky Way potential which is our goal here, \( \Theta \) represents the parameters of the potential, \((f_h, \alpha)\).

Let us divide the action space into \( M \) small bins such that the average number of stars per bin (over all of action space) is \( \bar{n} \ll 1 \). However, because of hierarchical structure formation, the expected number counts of stars in different bins will not be independent of each other. In order to capture this, we assume that the star count in bin \( a \) is a Poisson sampling of a mean, \( \bar{n} + \bar{n}_a \), where \( \bar{n}_a \)'s are correlated random gaussian variables. Therefore, the probability of measuring star counts \( \{v_a\} \) is given by:

\[ P(\{v_a\}, \{\bar{n}_a\}) = \exp\left(-\frac{1}{2} \sum_{a,b} \bar{n}_a \bar{n}_b \xi_{ab}\chi_{ab} \right) \prod_a \frac{v_a!}{\bar{n}_a!} \exp(-\bar{n} - \bar{n}_a). \]  

(A2)

From this definition, it follows that

\[ \langle \bar{n}_a \bar{n}_b \rangle = \xi_{ab}. \]  

(A3)

is the covariance matrix of the random gaussian variables \( \{\bar{n}_a\} \). Since we do not directly observe \( \bar{n}_a \)'s, we should marginalize over them. Therefore, in the limit of \( M \to \infty \) (i.e. when \( v_a = 0 \) or 1) the posterior (Equation A1) is given by:

\[ P(f_h, \alpha|\{v_a\}) \propto P_{\text{prior}}(f_h, \alpha) \int \frac{\exp\left(-\frac{1}{2} \sum_{a,b} \bar{n}_a \bar{n}_b \xi_{ab}\chi_{ab} \right) \prod_k \left[1 + \chi_{ab}\right]}{\sqrt{\det(\xi)(2\pi)^{M/2}}} \prod_a \bar{d}_ya. \]  

(A4)

where \( a_k \) is the action-space bin in which the \( k \)-th star lies, and we assume \( \sum_a \bar{n}_a = 0 \) over the entire action space. Using Wick’s theorem, the above Gaussian integral can be expressed in terms of a sum of the product of 2-point functions over all possible pairings of stars. This yields the likelihood (defined as the ratio of posterior to prior):

\[ L(f_h, \alpha|\{a_k\}) \equiv \frac{P(f_h, \alpha|\{a_k\})}{P_{\text{prior}}(f_h, \alpha)} \propto \left( \prod_k \left[1 + \chi_{ab}\right] \right)^{-1} \left( \sum_{\text{pairings}} \prod_{\text{pairs}} \left[1 + \xi_{\text{pair}}\right] \right)^{-1}. \]  

(A5)

Now, for a large number of pairs, we can expect that the exponent in this expression have smaller and smaller relative fluctuations around its mean for different possible pairings. This is often known as the mean-field approximation in statistical mechanics (where the sum represents the partition function and the exponent is proportional to the energy), and allows us to move the average inside the exponent:

\[ \ln L(f_h, \alpha|\{a_k\}) = \sum_{\text{pairs}} \ln \left(1 + \xi_{\text{pair}}\right) \]  

(A6)

This equation defines our log-likelihood formula adopted in Equation 17 in the main text (and subsequent statistical analyses), where we further assume that the 2-point function \( \xi \) only depends on the normalized distance \( D \) (Equation 14) in the action space.

APPENDIX B: SANITY CHECK FOR SIMULATIONS

To check whether the maximum likelihood test does correspond to the most clustering in the action space for either simulations, we plot the 2D projection of stellar distribution and two-point correlation function varying the choice of potential parameters \( f_h \) and \( \alpha \) used in the action computation. Here, we summarize the results. Figure B2 shows how a 2D projection of stellar distribution in the action space varies with different choices of potential for one of the simulations. In these figures, \( f_h(\alpha) \) is fixed, while \( \alpha (f_h) \) is varying across its correct value. As expected, for both simulations, the most compact distributions occur when parameters approach the correct values for the simulation (middle panel in both figures). This is also verified in the behavior of the two-point correlation function in Figure B3, where one of the parameters is fixed and the other one is varying. For both simulations, we see that the two-point correlation function is indeed maximized around the expected value, which proves the viability of our method.

APPENDIX C: COMBINING MEASUREMENTS WITH UNKNOWN SYSTEMICS

Here, we discuss how to combine measurements \( x_i \) (of a single quantity \( x \)) that have independent known stochastic gaussian errors \( \sigma_i \), as well as an unknown (but independent) systematic gaussian error \( \sigma_{\text{sys}} \). The joint likelihood is given by:

\[ L(x, \sigma_{\text{sys}}|\{x_i, \sigma_i\}) = \prod_i \frac{\exp\left(-\frac{(x-x_i)^2}{2(\sigma_i^2 + \sigma_{\text{sys}}^2)}\right)}{\sqrt{2\pi(\sigma_i^2 + \sigma_{\text{sys}}^2)}}. \]  

(C1)
Figure B1. The posterior distribution of $f_h$ (upper panel) and $\alpha$ (lower panel) for case $[f_h = 0.35, \alpha = 1.70]$ in simulation. The distributions are evaluated at three different values of $D_{\text{max}}$. A Gaussian fit has also been over-plotted on each panel for comparison. We see that the posterior distribution is well approximated by a Gaussian.

Now, assuming a flat prior on $\sigma_{\text{sys}}$, up to some maximum $\sigma_{\text{sys,max}}$, we can find the posterior on the parameter $x$:

$$P(x) \propto \int_0^{\sigma_{\text{sys,max}}} d\sigma_{\text{sys}} \prod_i \exp \left[ \frac{(x-x_i)^2}{2(\sigma_i^2 + \sigma_{\text{sys}}^2)} \right] \frac{1}{\sqrt{2\pi(\sigma_i^2 + \sigma_{\text{sys}}^2)}}$$  \hspace{1cm} (C2)

In practice, based on the difference in results we find with and without the cuts, we make conservative choices of $\sigma_{\text{sys,max}} = 0.02$ and 0.2 for $f_h$ and $\alpha$ determinations, respectively.
Figure B2. Stellar distribution in the $J_R$ and $J_\phi$ 2D projected plane, where $J_R$ and $J_\phi$ are defined as $J_R/\sigma_{J_R}$ and $J_\phi/\sigma_{J_\phi}$. This figure is aimed at presenting a general view of how the “compactness” (clustering behaviour) of stars in the action space varying with different choices of potential. First two rows show the action distribution for the simulation with $[f_h = 0.35, \alpha = 1.70]$, varying $f_h$ and $\alpha$ used in the calculation of action variable. Last two rows show the same thing the simulation with $[f_h = 0.35, \alpha = 2.00]$. As expected, stars appear to be most clustered if the correct parameters are used to compute the actions.
Figure B3. Correlation function \( \frac{P(\ln D)}{D^3} \) as a function of the distance in the action space in natural logarithm scale. The purpose of this figure is to check how the two-point correlation function varies with different choices of potential, and whether the correlation function is maximized at the correct set of parameter. Top panel: the behaviour of two-point correlation function for case \([f_h = 0.35, \alpha = 1.70]\) with fixed \(f_h, \alpha\) on the left (right). Different colors indicate the values of \(\ln \left( \frac{P(\ln D)}{D^3} \right)\) at different choices of potential. Bottom panel: the behaviour of two-point correlation function for case \([f_h = 0.35, \alpha = 2.00]\) with fixed \(f_h, \alpha\) on the left (right).