The Astrophysical Journal, 910:17 (16pp), 2021 March 20 https://doi.org/10.3847/1538-4357/abe1b7

Orbital Torus Imaging: Using Element Abundances to Map Orbits and Mass in the Milky Way

Adrian M. Price-Whelan[1], David W. Hogg[1,2,3], Kathryn V. Johnston[1,4], Melissa K. Ness[4], Hans-Walter Rix[2], Rachael L. Beaton[5,6,14], Joel R. Brownstein[7], D. A. García-Hernández[8,9], Sten Hasselquist[7,15], Christian R. Hayes[10], Richard R. Lane[11], Matthew Shetrone[12], Jennifer Sobeck[13], and Gail Zasowski[7]
[1] Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY 10010, USA; aprice-whelan@flatironinstitute.org
[2] Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany
[3] Center for Cosmology and Particle Physics, Department of Physics, New York University, 726 Broadway, New York, NY 10003, USA
[4] Department of Astronomy, Columbia University, New York, NY 10027, USA
[5] Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08544, USA
[6] The Observatories of the Carnegie Institution for Science, 813 Santa Barbara Street, Pasadena, CA 91101, USA
[7] Department of Physics and Astronomy, University of Utah, 115 S. 1400 E., Salt Lake City, UT 84112, USA
[8] Instituto de Astrofísica de Canarias (IAC), E-38205 La Laguna, Tenerife, Spain
[9] Universidad de La Laguna (ULL), Departamento de Astrofísica, E-38206 La Laguna, Tenerife, Spain
[10] Department of Astronomy, University of Washington, Box 351580, Seattle, WA 98195, USA
[11] Instituto de Astronomía y Ciencias Planetarias de Atacama, Universidad de Atacama, Copayapu 485, Copiapó, Chile
[12] University of California Observatories, UC Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA
[13] Department of Astronomy, University of Washington, Seattle, WA 98195, USA
Received 2020 November 21; revised 2021 January 29; accepted 2021 January 29; published 2021 March 23

Abstract

Many approaches to galaxy dynamics assume that the gravitational potential is simple and the distribution function is time invariant. Under these assumptions there are traditional tools for inferring potential parameters given observations of stellar kinematics (e.g., Jeans models). However, spectroscopic surveys measure many stellar properties beyond kinematics. Here we present a new approach for dynamical inference, Orbital Torus Imaging, which makes use of kinematic measurements and element abundances (or other invariant labels). We exploit the fact that, in steady state, stellar labels vary systematically with orbit characteristics (actions), yet must be invariant with respect to orbital phases (conjugate angles). The orbital foliation of phase space must therefore coincide with surfaces along which all moments of all stellar label distributions are constant. Both classical-statistics and Bayesian methods can be built on this; these methods will be more robust and require fewer assumptions than traditional tools because they require no knowledge of the (spatial) survey selection function and do not involve second moments of velocity distributions. We perform a classical-statistics demonstration with red giant branch stars from the APOGEE surveys: we model the vertical orbit structure in the Milky Way disk to constrain the local disk mass, scale height, and the disk–halo mass ratio (at fixed local circular velocity). We find that the disk mass can be constrained (naïvely) at the few-percent level with Orbital Torus Imaging using only eight element-abundance ratios, demonstrating the promise of combining stellar labels with dynamical invariants.

Unified Astronomy Thesaurus concepts: Galaxy dynamics (591); Milky Way dynamics (1051); Stellar kinematics (1608); Astrostatistics (1882); Chemical abundances (224); Spectroscopy (1558)

1. Introduction

An important goal of modern physics and astronomy is to understand the detailed properties of dark matter on astrophysical scales as a way of informing constraints on its fundamental nature (see, e.g., recent reviews by Bullock & Boylan-Kolchin 2017; Buckley & Peter 2018). Significant effort has therefore gone into developing and applying tools for constraining the mass distributions (i.e., dark matter distributions) of Local Group galaxies using only kinematic observations of tracers (i.e., stars; Jeans 1919; Binney & Tremaine 2008 and references therein). This challenge—using observations of tracer objects to constrain the underlying force law—is conceptually similar to a much older problem faced by physicists of the 17th century, who worked out that the gravitational force law in the solar system is proportional to the inverse square of the distance from the Sun (Newton 1687). That inference was based on observations that showed that orbits in the solar system are closed ellipses, with the Sun at one focus (Kepler 1609)—that is, they made use of observations of a few tracers (the planets) at many orbital phases.

In our current efforts to map dark matter, we are instead in a regime where we observe *many* tracers (stars) but have little or no information about orbital phase: we observe only a single snapshot (in time) of the stellar orbits. The closest we come to seeing orbits directly in the Milky Way is in the study of stellar streams (e.g., Helmi & White 1999; Johnston et al. 1999; Eyre & Binney 2011; Sanders & Binney 2013; Price-Whelan et al. 2014; Bonaca & Hogg 2018), where ensembles of stars almost directly encode (differential) phase information about the orbit of their progenitor systems. But stellar streams are relatively sparse in the Milky Way, and orbits in the Galaxy have more degrees of freedom than orbits in the solar system, meaning

[14] Carnegie-Princeton Fellow.
[15] NSF Astronomy and Astrophysics Postdoctoral Fellow.

that we need many more orbits to span the phase space and obtain precise constraints on the mass distribution. One could argue then that our current task is much harder than that of Newton or Kepler: we seek to constrain the global, spatially extended distribution of dark matter around a galaxy—a time-evolving mass distribution with nontrivial shape and radial distribution—with only a snapshot of the tracer kinematics (and often only a subset of phase-space dimensions).

Given only a snapshot of the dynamics, the tools we use to constrain the Galactic force field (i.e., the dark matter distribution) are therefore typically statistical in nature and do not depend on our knowledge of the orbits of individual stars. These methods generally rely on making strong assumptions about the distribution function (DF) or mass model. For example, in the case of Jeans modeling (e.g., Jeans 1922; Oort 1932; Bahcall 1984; Romanowsky et al. 2003; Evans et al. 2009; Watkins et al. 2010; Walker & Peñarrubia 2011; Zhai et al. 2018; Buch et al. 2019), investigators make use of the Jeans equations, which relate spatial derivatives of velocity second moments to derivatives of the gravitational potential. The equations are correct for any collisionless tracer population of stars, but implementing Jeans models in practice requires the assumption that the system is in steady state or equilibrium and requires that an explicit, parametrized model be specified for the underlying gravitational potential (see, e.g., Read 2014, for a review of such methods as applied to the problem of determining the local dark matter density). More general approaches here instead attempt to model the DF explicitly (e.g., Bovy et al. 2010, 2016; McMillan & Binney 2013; Magorrian 2014, 2019; Binney et al. 2014), but these methods become computationally prohibitive for large data sets or flexible model forms. When applied in real-world contexts to stars in the Milky Way or dwarf spheroidal galaxies, standard methods for inferring a mass distribution from tracer kinematics therefore typically assume an equilibrium (or steady-state) DF, (coordinate) separability of the DF, symmetries, and simple parameterizations of the mass model (e.g., integrable models) and DF, among others.

A contemporary "challenge" to applying these methods to modern data (especially within the Milky Way) is that stellar kinematic data are now very precise, and the phase space is well sampled. For example, the second data release (DR2) from the Gaia mission (Gaia Collaboration et al. 2016, 2018a) has provided full phase-space kinematics (for a subset of stellar tracers) over a region several kiloparsecs in size around the Sun. These data have revealed evidence of dynamical disequilibrium throughout the Galactic disk and halo (Antoja et al. 2018; Gaia Collaboration et al. 2018b; Koppelman et al. 2018; Myeong et al. 2018; Eilers et al. 2020). Standard methods for estimating dark matter properties using stellar kinematics therefore rely on a list of assumptions that we now *know* are strongly violated in the real universe and Galaxy. The fact that we still use and apply these methods is a reflection of the fact that relaxing these assumptions make dynamical inferences[16] *far* more challenging and computationally expensive. For example, there are no straightforward methods for measuring the mass distribution when the assumption of dynamical equilibrium is relaxed.

One reason to remain hopeful in our goal of precisely constraining the dark matter is that, unlike Kepler and Newton, we have excellent working models of stars and access to much more information per star than kinematics alone. For example, spectroscopic surveys have measured the stellar parameters and element abundances of hundreds of thousands, and soon to be millions, of stars (e.g., Deng et al. 2012; Martell et al. 2017; Ahumada et al. 2020) over large regions of the Galactic disk and halo. It is therefore promising to think that utilizing stellar "labels" (element abundances, stellar ages, or other effectively invariant stellar properties) within dynamical inferences will provide additional information that may help interpret or model the Galaxy (see, e.g., Sanders & Binney 2015; Binney & Sanders 2016; Das & Binney 2016; Iorio & Belokurov 2021, for recent methods that begin to move in this direction, within the context of equilibrium models).

In this paper, we demonstrate that stellar surface abundances can be used to illuminate the orbit structure in the Milky Way, and are therefore extremely valuable for galaxy dynamics. In our current methodology, like all other dynamical inferences, we work only under strong assumptions and we are therefore still in the business of revealing the orbits indirectly. However, this class of approaches, here referred to as Orbital Torus Imaging, is qualitatively distinct from all previous methods for measuring the mass model. There are reasonable regimes in which it will be more precise than any other method, conditioned on the assumptions.

2. Methodological Generalities

In a well-mixed equilibrium population, stars are in a kinematic steady state. As time goes on, stars move along their orbits: these orbits can be represented by a set of dynamical invariants—actions—and a location along an orbit can be represented by a set of phase—conjugate angle—variables. These action–angle coordinates, $(\boldsymbol{J}, \boldsymbol{\theta})$, are canonical coordinates (i.e., a transformation from, say, Cartesian position and velocity) in which many dynamical computations are simpler or more efficient (see, e.g., Binney & Tremaine 2008). The actions $\boldsymbol{J}$ are integrals of motion like any other (e.g., energy or angular momentum), but are special in that they form a set of momentum coordinates whose conjugate position coordinates are angle variables. Action variables are defined as integrals of other canonical momentum coordinates $\boldsymbol{p}$ over a closed loop in the conjugate position coordinates $\boldsymbol{q}$,

$$J_i = \oint p_i \, dq_i. \tag{1}$$

As we see below, if the motion in a given coordinate pair (e.g., q_i, p_i) is separable from the other coordinates, the action integral simply computes the area enclosed by the orbit in the two-dimensional phase space (q_i, p_i).

For gravitational potentials with more than one degree of freedom, it is not guaranteed that actions exist at all locations in phase space: generically, regions of phase space can contain chaotic orbits and resonances. However, when the resonant structure is weak (as is the case for many simple models of Galactic potentials), most of the phase space is typically stable and therefore admits transformation to action-space. In action–angle coordinates, a steady-state DF f is simply a function of the actions, $f(\boldsymbol{J})$. In such situations, every location in angle-space is equally likely, or equally probably populated in any snapshot of the kinematics of tracers orbiting in some mass distribution.

[16] By "dynamical inferences," we really mean statistical inferences of a mass model (or parameters of a mass model) given noisy observations of stellar kinematics (and other stellar labels).

How, then, do stellar abundances relate to dynamics? The "chemical tagging" insight (Freeman & Bland-Hawthorn 2002) notes that most stars also effectively preserve their surface element abundances (and other stellar labels) as they orbit over many Galactic orbital timescales. That is, in an integrable galaxy—a galaxy whose stellar orbits are associated with three invariant actions—the element abundances and the dynamical actions have something in common: they are invariant with time. In these models, only the conjugate angle coordinates are time-dependent. This means that, for a well-mixed population, the detailed element abundances can only be a function of actions, and never a function of conjugate angles. That is a remarkably informative constraint on the configuration of the Milky Way in the space of positions (dimensionality 3), velocities (3), and detailed abundances (10–30, depending on the spectroscopic survey).

Consider a collection of stars (localized, say, in phase space) for which we have measured abundances for D elements. This collection of stars will in general have a diversity of element abundances: their abundances are drawn from some distribution in D-dimensional element-abundance space. In general, the abundance distribution will depend on position in phase space: in the Milky Way, there are observed radial and vertical abundance gradients in the Galactic disk (e.g., Hayden et al. 2015; Queiroz et al. 2020). However, any changes we observe in these element-abundance distributions with respect to phase-space coordinates (i.e., the six-gradient with respect to the three position and velocity components) must not project onto the directions of increasing (or decreasing) conjugate angles in phase space. All gradients with respect to phase-space coordinates of the element-abundance distribution must be orthogonal to the directions of increase (or decrease) of the conjugate angles, and lie in the subspace of the directions of increase (or decrease) of the dynamical actions. The trajectories of stars in the phase space (the dynamical tori) must therefore lie along or describe level surfaces in the (ensemble mean) element-abundance distribution.

In this paper, we demonstrate the utility of these gradients and relations between element abundances and dynamical invariants in the context of dynamical inferences in the Milky Way. We consider stars in our Galaxy because here we can measure six-dimensional kinematics and element-abundances for individual stars, enabling relatively simple demonstrations of the concepts here. However, we note that a generative model built on these concepts would, in principle, be applicable in more general scenarios, such as for stars in Local Group satellite galaxies where only a subset of phase-space coordinates are measured.

3. Data

In our toy demonstrations below, our main data source is a cross-match between spectroscopic data from the APOGEE surveys (Majewski et al. 2017) and astrometric data from the Gaia mission (Gaia Collaboration et al. 2016, 2018a).

APOGEE is a spectroscopic sub-survey and component of the Sloan Digital Sky Survey IV (SDSS-IV; Eisenstein et al. 2011; Blanton et al. 2017) whose main goal is to map the chemical and dynamical properties of stars across the Milky Way disk. The survey uses two nearly identical, high-resolution ($R \sim 22,\ 500$; Wilson et al. 2019), infrared (H-band) spectrographs—one in the Northern hemisphere at Apache Point Observatory (APO) using the SDSS 2.5 m telescope (Gunn et al. 2006), and one in the Southern hemisphere at Las Campanas Observatory (LCO) using the 2.5 m du Pont telescope (Bowen & Vaughan 1973). The primary survey targets are selected with simple color and magnitude cuts (Zasowski et al. 2013, 2017; R. Beaton et al. 2021, in preparation; F. Santana et al. 2021, in preparation), but the sparse angular sky coverage and limited number of fibers per field lead to a "pencil-beam"-like sampling of the Milky Way stellar density. APOGEE spectra are reduced (Nidever et al. 2015) and then analyzed (i.e., to measure stellar parameters and abundances) using the APOGEE Stellar Parameters and Chemical Abundance Pipeline (ASPCAP; García Pérez et al. 2016; Holtzman et al. 2018; Jönsson et al. 2020); here we use abundance measurements from the standard APOGEE pipeline.

Here we make use of a recent internal data product (which includes all data taken through 2020 March) from the APOGEE surveys (post-DR16) that includes ≈60% more stars than the publicly available DR16 catalogs (Ahumada et al. 2020; Jönsson et al. 2020), but was reduced and processed using the same pipeline used to produce the DR16 release (i.e., the pipeline described in Jönsson et al. 2020). This APOGEE catalog contains calibrated element-abundance measurements for 18 elements, but these have a variety of physical origins and a range of reliabilities and measurement precisions. For demonstrations below, we therefore focus on a subset of eight well-measured (log) abundance ratios selected to have varied astrophysical origins: [Fe/H], [C/Fe], [N/Fe], [O/Fe], [Mg/Fe], [Si/Fe], [Mn/Fe], [Ni/Fe]. In some cases, we focus on just a single element abundance, [O/Fe], which is one of the most precisely and accurately determined element abundance measured with the DR16 pipeline (Jönsson et al. 2020).

Gaia is primarily an astrometric mission and survey (Gaia Collaboration et al. 2016) that obtains sky position, proper motion, and parallax measurements for >1 billion stars, limited only by their apparent magnitudes (Gaia $G \lesssim 20.7$). Here we use parallax and proper motion measurements released in Gaia DR2 (Gaia Collaboration et al. 2018a; Lindegren et al. 2018).

We cross-match the APOGEE sample to Gaia DR2 using the APOGEE-provided 2MASS (Skrutskie et al. 2006) identifiers, and the Gaia-provided cross-match between Gaia DR2 and the final 2MASS point source catalog (Marrese et al. 2019). We then apply a number of quality cuts and other selections to limit the catalog to chemically thin-disk, red giant branch (RGB) stars with well-measured stellar parameters ($\log g$, $T_{\rm eff}$) and abundances (the abundance ratios listed above), high signal-to-noise-ratio parallax measurements, but excluding stars targeted in stellar clusters and dwarf galaxies, as enumerated below. In detail, our selections include:

1. ASPCAP quality flag (`ASPCAPFLAG`) must not contain `STAR_BAD`or `STAR_WARN`,
2. $3500\ {\rm K} < T_{\rm eff} < 6500\ {\rm K}$,
3. $1.5 < \log g < 3.4$,
4. combined spectroscopic signal-to-noise ratio ${\rm S/N} > 20$,
5. APOGEE targeting bit flags must not indicate that the source was part of a special program to observe stellar clusters, dwarf galaxies, M31 stars, stellar streams, or moving groups; this *excludes* stars with the following bits enabled:
   (a) `APOGEE_TARGET1`: (9, 18, 24, 26)
   (b) `APOGEE_TARGET2`: (10, 18)
   (c) `APOGEE2_TARGET1`: (9, 18, 20, 21, 22, 23, 24, 26)
   (d) `APOGEE2_TARGET2`: (10)
   (e) `APOGEE2_TARGET3`: (5, 14, 15)

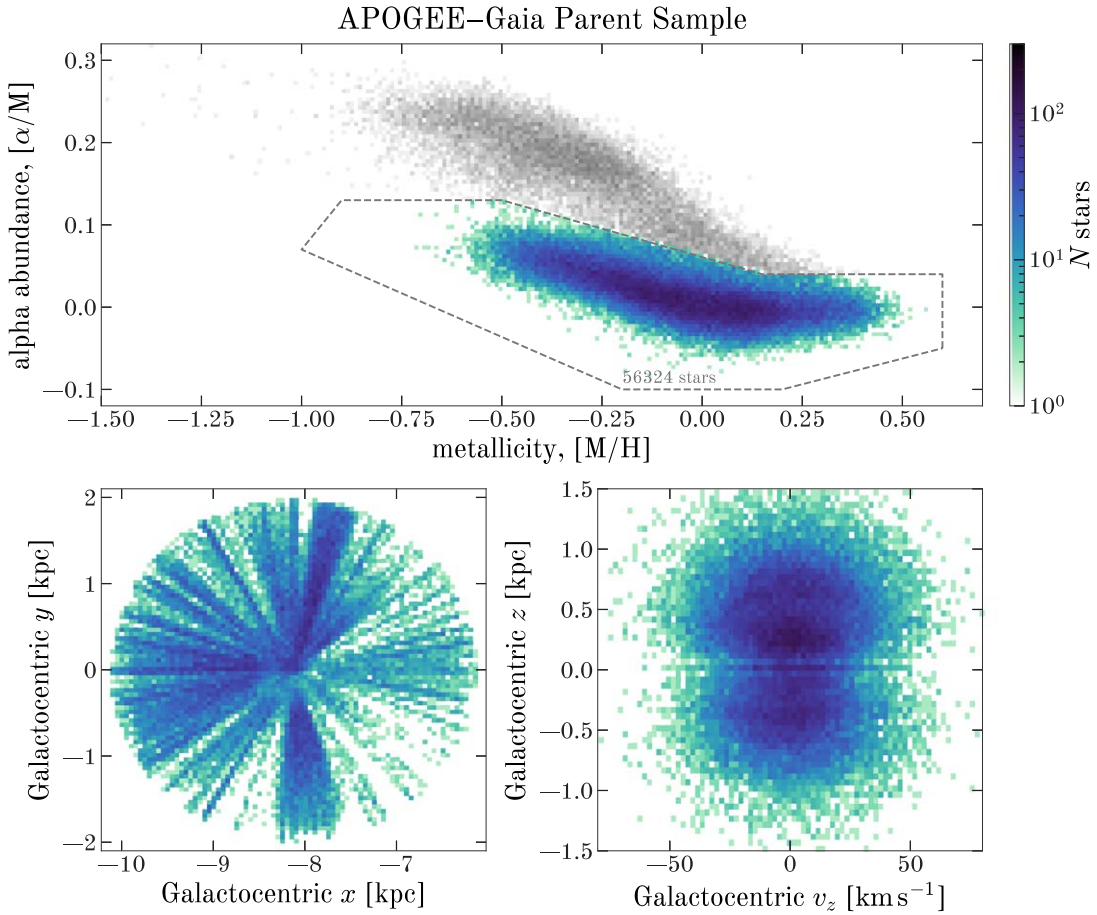

Figure 1. Data sample used in this project. Each panel shows a 2D histogram of the stars. Top panel: bulk abundance ratios measured by APOGEE and a selection boundary (dashed line) used to exclude the "high-α" stars that are generally older and kinematically hotter. Lower left panel: positions of the stars in the parent sample projected onto the Galactic plane, showing the spherical spatial cut and highly nonuniform (APOGEE) spatial selection. Heliocentric distances to the stars are obtained by naïvely inverting their Gaia parallax measurements. Lower right panel: distributions of stars in the parent sample in vertical z, v_z phase space. This panel shows that the sample is less populated at low z (mainly because of the APOGEE selection function), which in turn shows that the sample is less populated at certain values of vertical angle, θ_z. The methods presented in this paper do not require that all angles be equally populated in the sample.

6. stars are part of the "low-α" (or "chemical thin disk") population (see polygonal selection in the top panel of Figure 1),
7. Gaia parallax $\varpi > 0.5$ mas,
8. Gaia parallax signal-to-noise ratio $\varpi/\sigma_\varpi > 8$.

The APOGEE catalog contains a small number of duplicates (duplicated source identifier `APOGEE_ID`); to avoid duplication in our sample in these cases, we keep only the entry with the highest signal-to-noise ratio. We select the low-α sequence primarily because the high- and low-α sequences have different scale heights (e.g., Bovy et al. 2016), and our toy model is not currently flexible enough to account for this (however, it is possible to include this complexity in our framework, as will be explored in future extensions of this model). The final parent sample contains 56,324 RGB stars with high-quality APOGEE and Gaia data.

For each star in the parent sample, we compute naïve distance estimates by inverting the parallax, $d = 1/\varpi$. While this is generally not a safe way of computing distance from parallaxes (see, e.g., Bailer-Jones 2015), our sample stars are (by construction) relatively nearby and have high signal-to-noise-ratio parallax measurements. Using a parallax signal-to-noise-ratio selection has its own consequences, especially in that it makes the sample selection function complex and nonintuitive. However, our methodology should be relatively insensitive to selection effects, and since the demonstrations that make use of these distance measurements below are meant to be illustrative examples, we ignore these details in what follows. This sample is visualized in bulk element-abundance ratios and Galactocentric positions in Figure 1.

4. Milky Way Mass Model and Action-angle Computations

Our methodology and demonstrations below rely on the computation of actions and angles for stars, which depend on the mass distribution of the Milky Way and the solar position and motion with respect to a Galactocentric reference frame.

For the solar position, we use the recent, precise measurement of the Sun–Galactic center distance from the GRAVITY collaboration, $r_\odot = 8.122$ kpc (Gravity Collaboration et al. 2018), and initially adopt $z_\odot = 20.8$ pc as the solar height above the Galactic midplane (Bennett & Bovy 2019). In Galactocentric Cartesian coordinates, we use a right-handed coordinate system such that the Sun is at $\boldsymbol{x}_\odot = (-8.1219,\ 0,\ 0.0208)$ kpc and the solar velocity is $\boldsymbol{v}_\odot = (12.9,\ 245.6,\ 7.78)\,\mathrm{km\,s^{-1}}$ (Reid & Brunthaler 2004; Drimmel & Poggio 2018; Gravity Collaboration et al. 2018).

We represent the density distribution (or gravitational potential) of the Milky Way using an idealized, four-component

mass model consisting of a spherical Hernquist bulge (Hernquist 1990), spherical Hernquist nucleus, an axisymmetric Miyamoto–Nagai disk (Miyamoto & Nagai 1975), and a spherical Navarro–Frenk–White dark matter halo (Navarro et al. 1996). Most of the parameters of these components are fixed to their default values from the `MilkyWayPotential` class implemented in the *gala* Python package (v1.2; Price-Whelan 2017): briefly, the bulge parameters (mass and scale radius) and disk parameters (total mass, scale height, and scale radius) are initially set to match the `MWPotential2014` implemented in *galpy* (Bovy 2015), and the dark matter halo parameters (virial mass and scale radius) are initially set by fitting the enclosed mass profile of the mass model to a compilation of recent enclosed mass measurements.[17] However, the mass of the disk is then adjusted to match a circular velocity of $v_{\rm circ}(R_\odot) = 229\,{\rm km\,s^{-1}}$ at the solar radius (Eilers et al. 2019). Our fiducial mass model therefore adopts the default `MilkyWayPotential` parameters except for the disk mass, which is set to $M^\star_{\rm disk} = 6.526 \times 10^{10}\,M_\odot$. Later in this paper, we vary the disk mass $M_{\rm disk}/M^\star_{\rm disk}$ and disk scale height at the solar position $h_{z,\odot}$, but we adjust the dark matter halo mass to keep the circular velocity at the solar radius, $v_{\rm circ}(R_\odot)$, constant.

In a given potential model, we compute actions, $\boldsymbol{J} = (J_R, J_\phi, J_z)$, and angles, $\boldsymbol{\theta} = (\theta_R, \theta_\phi, \theta_z)$, for a star using the "Stäckel Fudge" (Binney 2012; Sanders 2012) as implemented in *galpy* (Bovy 2015). We solve for the focal length, Δ, of the locally approximating Stäckel potential for the orbit of each star by numerically computing the orbit of each star for four orbital periods.[18]

We assess the accuracy of using the Stäckel Fudge for orbits with large vertical excursions from the Galactic disk by comparing actions computed with the Stäckel Fudge to actions computed with a more accurate action solver. For a set of trial orbits (meant to span the range of vertical actions we see in our sample of APOGEE stars) over a range of disk masses (from $M_{\rm disk}/M^\star_{\rm disk} = 0.5$–$1.5$), we compute actions both with the Stäckel Fudge and with the "O2GF" method defined in Sanders & Binney (2014, 2016). The O2GF method works by numerically integrating an orbit for a given star and solving for the generating function to transform from actions computed in a toy potential model to the actions in the potential model of interest (as defined in Sanders & Binney (2014), and implemented in *gala*). This method has some tuning parameters related to the total orbital integration time and time step, and the number of Fourier components to include in the Fourier expansion used to represent the generating function, $N_{\rm max}$. We use an Isochrone potential as our toy potential model, set the total integration time for each star to 128 (radial) orbital periods, $T = 128\,P_r$, and set the time step to $\Delta t = P_r/256$. For all stars, we set $N_{\rm max} = 8$ (following Sanders & Binney 2016). We find that in all cases, the values of the actions agree to within $<10\%$, $<0.1\%$, and $<0.5\%$ for J_R, J_ϕ, and J_z, respectively, with the largest disagreements only affecting a small fraction of the stars in our sample ($<8\%$) with the largest excursions ($\gtrsim 3$ scale heights) from the Galactic plane.

For computational efficiency, we therefore use the Stäckel Fudge as our primary method for computing actions. To additionally speed up computation, we typically also parallelize the computation of the actions (over stars) using the Python package *schwimmbad* (Price-Whelan & Foreman-Mackey 2017).

[17] As described in https://gala.adrian.pw/en/latest/potential/define-milky-way-model.html.

[18] For computational efficiency, we actually compute the locally fitting Stäckel potential focal length parameter Δ (Sanders 2012) using *gala* (Price-Whelan 2017).

5. Motivation from Observed Element Abundance Gradients

A significant motivation for this work came from plots of elemental abundance ratios of stars as a function of vertical height z and vertical velocity v_z in Galactocentric Cartesian coordinates. As an example, Figure 2 shows the mean abundance ratios of stars in bins of their vertical phase-space coordinates using data from the APOGEE and Gaia surveys (see Section 3). The stars shown in this figure are selected following the quality cuts defined above (Section 3) and are selected to lie in the low-alpha sequence (Figure 1). In these plots, the eye is drawn to *abundance gradients*: the stars at small heights and small vertical velocities have different abundance ratios, on average, than stars at large heights and large vertical velocities. But these positional and velocity gradients are related: stars at large absolute vertical velocities v_z will, as they orbit, reach large absolute vertical positions z (i.e., far from the Galactic plane), and stars at large absolute vertical positions will, in the future, reach large absolute vertical velocities. In other words, the stars will orbit in the Galaxy, which projects onto this z–v_z plane as (to zeroth order) roughly elliptical trajectories. For example, Figure 3 shows two Galactic orbits computed in a 3D model for the Milky Way (see Section 4) in different projections of phase-space coordinates: in the space of z–v_z (center panel), a Galactic orbit will form a close-to-elliptical band whose enclosed area scales with the vertical action, J_z, whose thickness depends on the eccentricity of the orbit, and a given position on its "ellipse" can nearly be mapped to a vertical angle, θ_z.

To very high precision, stars do not change their abundances as they orbit. One consequence of this is a new method for inferring the orbit structure of the Milky Way: if two small neighborhoods in phase space lie on the same orbit—that is, they correspond to the same dynamical actions but with different conjugate angles—they must contain stars with the same distribution of element abundances. This prediction depends on many detailed assumptions; for example, that the Galaxy is (approximately) phase mixed, and that the potential is (approximately) time invariant and integrable. Of course, the *usefulness* of this prediction for inference depends on the existence of element-abundance gradients: if there are no element-abundance-ratio gradients, there will be no information to exploit.

For the sake of illustration and simplicity, we visualize and demonstrate these concepts using the vertical kinematics of stars in our parent sample. Figure 4 again shows the mean [O/Fe] abundance ratios (in all panels), but now with two overlaid orbits (green overlaid bands) computed in three different Milky Way models (with varied disk mass, as indicated; see Section 4). The two orbits were chosen for illustrative purposes (one with low J_z, one with higher J_z), and are defined such that their three actions (J_R, J_ϕ, J_z) have the same values in all mass models. In the fiducial mass model ($M_{\rm disk}/M^\star_{\rm disk} = 1.0$), the two overlaid orbits nearly follow mean abundance contours. In the model in which the disk is made less massive (and the halo more massive to keep the circular velocity constant; $M_{\rm disk}/M^\star_{\rm disk} = 0.4$), the orbits change shape: there is more positional extent to an orbit relative

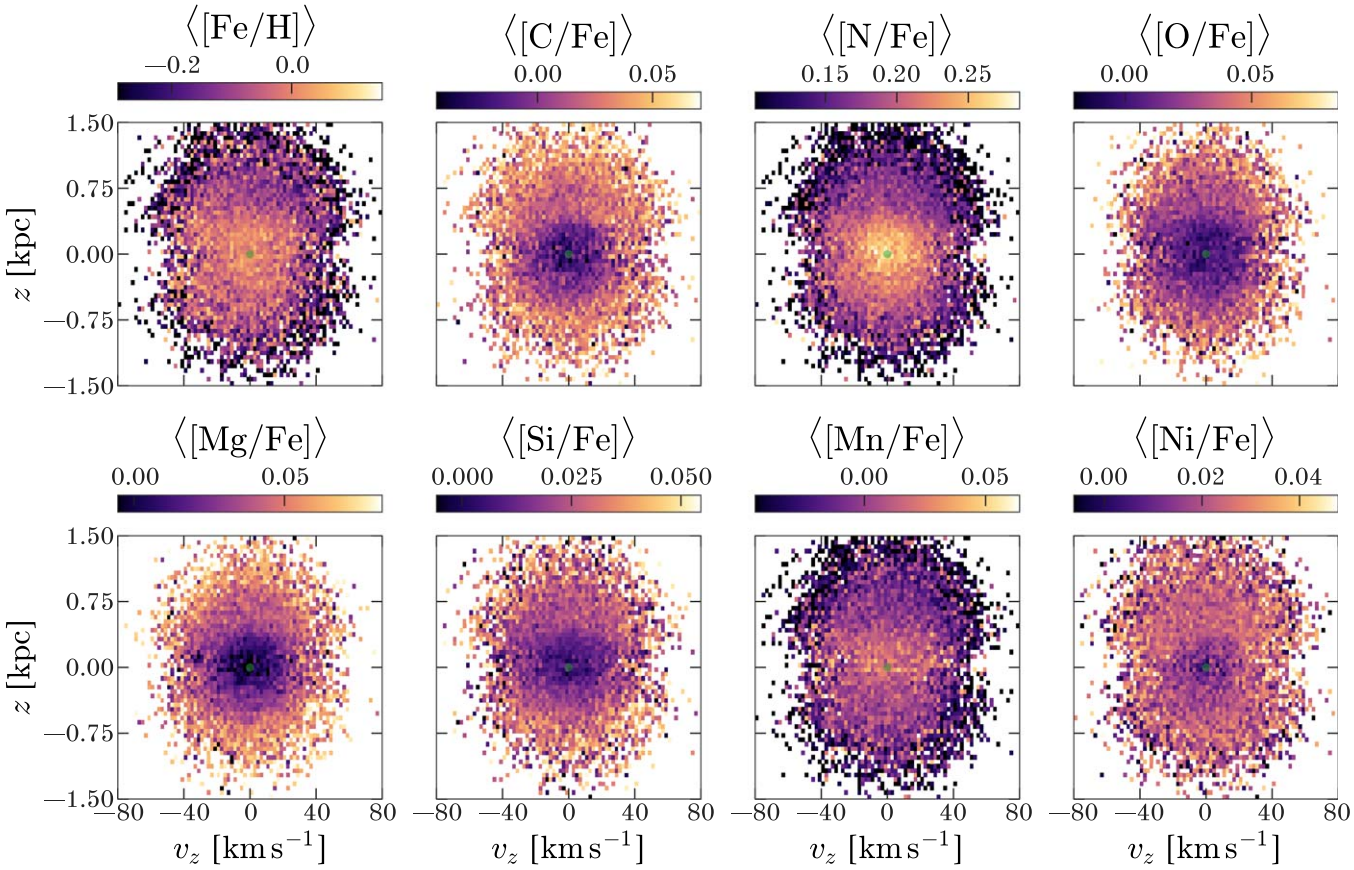

Figure 2. Means of various element-abundance ratios as a function of Galactic vertical height z and vertical velocity v_z. Averages are taken in z, v_z boxels. The stars at lower $|z|$ and lower $|v_z|$ (i.e., the stars with lower overall vertical action J_z) show higher overall metallicity on average, but lower α-to-iron. These plots are somewhat affected by APOGEE selection effects, in that different z, v_z boxels are projections through different extents in Galactocentric radius (see Figure 1). This explains some of the visible asymmetries.

to its velocity extent. If stars were traveling on these lower-disk-mass orbits, they would have to obtain higher abundances when they are passing through the disk midplane, and lower abundances when they are at their greatest absolute vertical heights, which is absurd: stars do not change their abundances as they orbit. In our method below, we utilize this fact to infer the disk mass.

6. Method, Assumptions, and Toy Applications

There are two families of approaches to Orbital Torus Imaging. In the first—which we call *classical* (in the sense of classical statistics)—we make explicit the fact that the element-abundance distribution should not depend on the conjugate angles. The best-fit or inferred mass-model parameters are those that lead to no residual dependence of abundances (or mean abundance ratio, or any moment of the abundance-ratio distribution) on angles. In the second—which we call *generative*—we would construct a predictive model for the high-dimensional abundance distribution as a function of actions alone (and not angles). The best-fit or inferred mass-model parameters are those that maximize the combined probability (density), evaluated at the observed abundances. The classical approaches are frequentist and the generative approaches produce likelihoods and can be used in Bayesian inferences.

Both classical and generative implementations of Orbital Torus Imaging depend on a specific set of assumptions, enumerated below. We emphasize that these assumptions are not necessarily correct, but that our method is *conditionally* correct, conditioning on these core assumptions:

Integrable: each stellar orbit is regular and has three dynamically invariant actions and three conjugate angles. This assumption could in principle be relaxed in future implementations.

Phase mixed: at any position in action-space, the stars in our sample are inherently uniformly distributed in angle variables; all angles are equally likely. This assumption will be violated in the data (as we discuss below; see Section 7), but this is the fundamental assumption of the vast majority of inferences of the Milky Way mass distribution.

Properly selected: the selection function depends on position (or velocity) in the Milky Way, but not on element abundances at a given position. In detail, the APOGEE selection function *will* depend on abundances through gradients between stellar parameters and element-abundance ratios, and implicit selections on stellar parameters (through, e.g., color and magnitude cuts). We attempt to mitigate these issues here by selecting a limited (in $\log g$) subset of red giant stars that are dominated by red clump stars (see Section 3), and we discuss this further in Section 7.

Measurable gradients: there exist gradients in the abundances with respect to the kinematics (i.e., actions) and those gradients are measurable at the precision of the individual stellar abundance measurements in our spectroscopic data set.

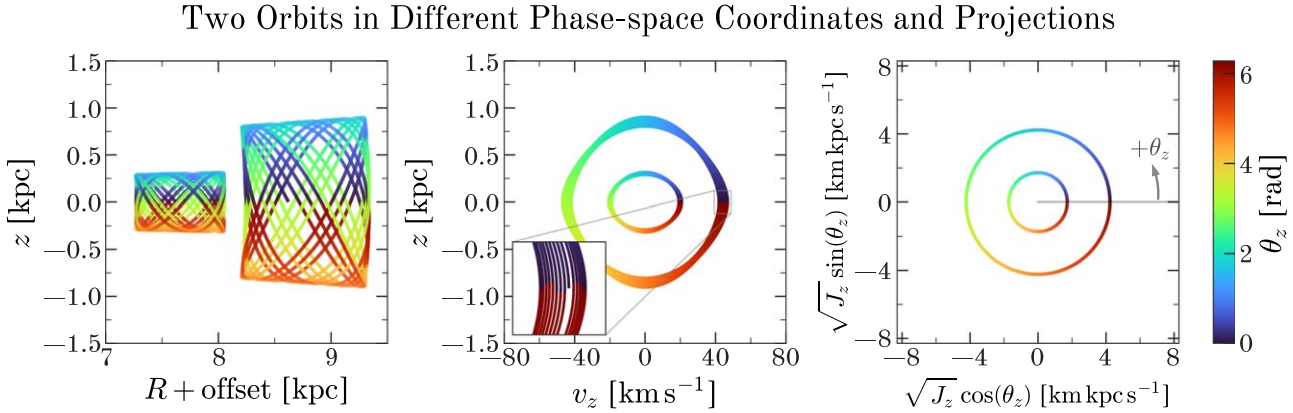

Figure 3. Left panel: two orbits projected onto the plane of Galactic vertical height z and Galactocentric cylindrical radius R. The orbits fill the surfaces of 3-tori in 6D phase space. Middle panel: same two orbits, but projected onto the plane of vertical height z and vertical velocity v_z. In this projection, it becomes clearer that the orbital lines are colored by the angle θ_z that is conjugate to vertical action J_z. The inset shows that conjugate angle wraps nontrivially, because the action (by construction) wraps at constant angular velocity, whereas the vertical period is a (weak) function of the other orbital phases. Note that although this projection is close to "face on" for these two orbits, the fact that they fill the surfaces of 3-tori means that they project to finite-width bands in z, v_z space. Right panel: same two orbits, but now plotted in vertical-action, vertical-angle space. In this space, the two orbits trace perfect circles.

Invariant abundances: stellar surface element abundances are time invariant. In detail, this assumption is violated by stellar and planetary evolution, as surface abundances can change, for example, during red giant branch evolution (e.g., Iben 1965; Martig et al. 2016) for solar-mass stars. However, we expect the magnitude of this to be unimportant and should not dominate our systematic errors as long as the timescales over which the abundances change are sufficiently different from the orbital timescales (hundreds of millions of years).

In this paper, we consider only classical-statistics approaches for illustrative purposes. However, we expect that generative approaches will produce at least slightly more precise inferences, because they will be protected by the arguments and proofs of Bayesian inference. In addition to the core assumptions listed above, we also make additional assumptions specific to our implementation and demonstrations. In particular, here we assume that the 6D phase-space positions are sufficiently accurate and precise such that we condition over these quantities directly (and ignore their reported uncertainties). For the parent sample used here, the median distance uncertainty (from inverting the Gaia parallaxes) is $\sigma_d \approx 50$ pc and the median velocity uncertainty is $\approx 2\,\mathrm{km\,s^{-1}}$. We also here assume that stars in different parts of phase space receive the same quality of abundance measurements, an assumption that we know is weakly violated because of known temperature (Jönsson et al. 2020) and $\log g$ (A. C. Eilers et al. 2021, in preparation) dependences on the APOGEE abundance measurements, and stars with different temperatures or surface gravities or luminosities can be seen at different distances and z heights. However, our parent sample is selected to contain a relatively small range of surface gravities along the red giant branch and predominantly consists of red clump stars; we therefore expect these issues to be negligible.[19] Finally, here we additionally assume that the Milky Way potential model is time-independent. This is not a strict or core assumption of Orbital Torus Imaging: we only require that the actions exist, and that the orbital phases of stars are mixed at any location in action-space. For example, this could still be satisfied in weakly time-dependent potentials in which the actions remain adiabatically invariant.

6.1. Initial Demonstration of Fitting Procedure

As an initial demonstration, we focus here on the vertical kinematics of stars (but we do not assume separability) and we ask whether there is a choice of mass model that leads to dynamical actions and conjugate angles such that the abundance-ratio distributions do not depend on the conjugate angle θ_z. To assess the dependence of a particular element-abundance ratio $[X/Y]$ on the vertical angle, we define a "mean abundance-ratio deviation," $\Delta^{[X/Y]}$, for each n star, which is the difference between the (logarithmic) abundance ratio of a given star and the mean of the (logarithmic) abundance ratios of its kinematic neighbors (in action-space),

$$\Delta_n^{[X/Y]} = [X/Y]_n - \langle [X/Y] \rangle_J . \quad (2)$$

That is, we use a given mass model to compute the three actions $\boldsymbol{J}$ for each star in the parent sample and choose the closest K neighbors in $\boldsymbol{J}$-space (with the isotropic Euclidean metric distance) for each star to compute the action-local mean abundance. In reality, we perform a weighted mean in order to account for the fact that there could be strong number-density gradients in action-space that could bias our estimates of the mean abundance values (we discuss this in more detail in the Appendix). For a specific star, its neighbors are a function of the mass-model parameters because all actions will change values in different mass models. We choose to use $K = 64$ based on experimentation: smaller K values more accurately resolve the abundance gradients but have more shot noise, but larger K values smooth out the abundance gradients. However, we find that our results are not very sensitive to this choice (within a multiplication or division by a factor of a few).

Our goal then is to minimize (over mass model parameters) the dependence of the mean abundance-ratio deviation distribution on the vertical angle θ_z. To quantify this dependence, we fit

[19] We tried repeating all subsequent analyses in this paper using a sample of high-confidence red clump stars (using the selection defined in Bovy et al. 2014) and found no significant differences in our results.

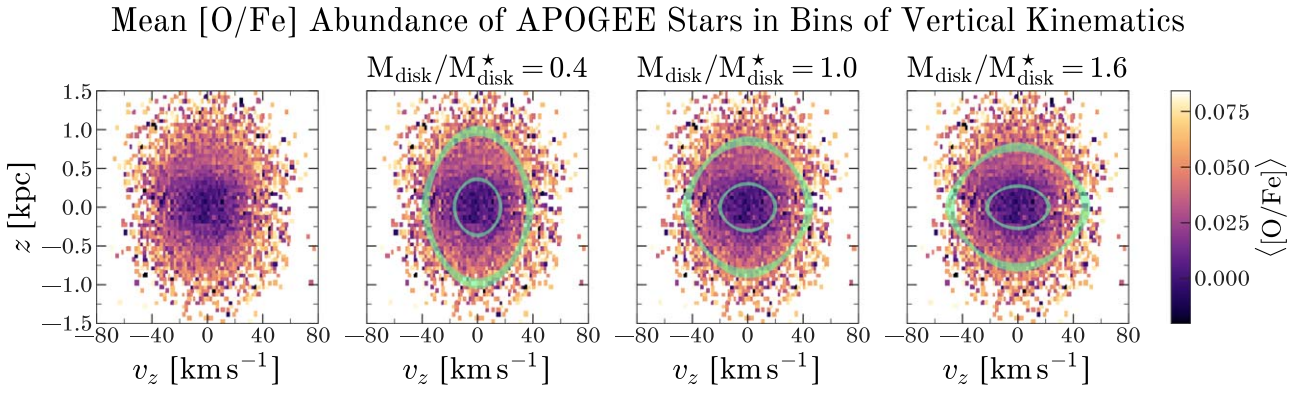

Figure 4. Left panel: repeat of the [O/Fe] panel of Figure 2. Other panels: same as the left panel, but with the two orbits from Figure 3 over-plotted for three different Milky Way potentials. These three potentials have the fiducial Milky Way disk mass (see the text for details), or a disk less massive by a factor of 0.4 or more massive by a factor of 1.6, as noted in each panel title. All potentials are constrained to have a circular velocity at the solar circle of 229 km s^{-1}. *Which of the three panels most suggests that the orbits are coincident with isopleths of the mean abundance?* This question is asked for illustrative purposes only: these plots distort the data by projection in phase space, and the methods we use for the inferences we perform (see Section 6) do not rely on or use any projections.

a smooth model of the form

$$\Delta(\theta_z) = c_0 + a_1 \cos\theta_z + b_1 \sin\theta_z + a_2 \cos 2\theta_z + b_2 \sin 2\theta_z, \tag{3}$$

where the five parameters (c_0, a_1, a_2, b_1, b_2) are the coefficients of a Fourier series expressed out to $m = 2$. The functional form captures our expectations for how the mean abundance-ratio deviations will depend on vertical angle when our model is wrong: if the orbit contours at a given action have the wrong shape, we expect there to be an $m = 2$ variation to the abundance-ratio deviations (e.g., Figure 4). If our assumed solar position relative to the midplane or solar velocity relative to the local standard of rest are wrong, we expect there to be sin and cos (i.e., $m = 1$) variations. In the smooth model for the mean abundance deviations (Equation (3)), the parameter a_1 is sensitive to the vertical component (v_z component) of the local standard of rest or the solar motion. Parameter b_1 is sensitive to the vertical (z) location of the Sun relative to the disk midplane. Parameter a_2 is sensitive to the local mass density concentrated in the disk, or the disk-mass parameter $M_{\rm disk}/M^{\star}_{\rm disk}$. Parameter b_2 should not exist and is included as a test of model assumptions; in detail, it is sensitive to tilts in the coordinate system, and nonphase-mixed structures. As we note later, the "best" setting of the mass model parameters should minimize a combination of these amplitudes.

We fit this model (Equation (3)) to all of the $N = 56,324$ individual abundance-ratio deviations and their uncertainties $(\Delta_n, \sigma_{\Delta_n})_N$ (without binning) using least-squares fitting. In detail, for a given mass model, we start by computing the actions and angles for all N stars in our sample. We then use the three actions for each star to compute the mean abundance deviations Δ_n and the associated uncertainty σ_{Δ_n} for all stars (Equation (2)). We construct the design matrix $\mathbf{M}$ using the vertical angle values $\theta_{z,n}$,

$$\mathbf{M} = \begin{pmatrix} 1 & \cos\theta_{z,1} & \sin\theta_{z,1} & \cos 2\,\theta_{z,1} & \sin 2\,\theta_{z,1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \cos\theta_{z,N} & \sin\theta_{z,N} & \cos 2\,\theta_{z,N} & \sin 2\,\theta_{z,N} \end{pmatrix} \tag{4}$$

the "data" vector $\boldsymbol{y}$ using the mean abundance deviations,

$$\boldsymbol{y} = \begin{pmatrix} \Delta_1 & \cdots & \Delta_N \end{pmatrix}^{\mathsf{T}} \tag{5}$$

and the covariance matrix $\mathbf{C}$ using the mean abundance-deviation uncertainties,

$$\mathbf{C} = \begin{pmatrix} \sigma^2_{\Delta_1} & & \\ & \ddots & \\ & & \sigma^2_{\Delta_N} \end{pmatrix}. \tag{6}$$

We compute the best-fitting parameter vector $\hat{\boldsymbol{g}} = (c_0, a_1, a_2, b_1, b_2)$ by solving the linear least-squares problem,

$$\hat{\boldsymbol{g}} = (\mathbf{M}^{\mathsf{T}}\mathbf{C}^{-1}\mathbf{M})^{-1}\,\mathbf{M}^{\mathsf{T}}\,\mathbf{C}^{-1}\,\boldsymbol{y} \tag{7}$$

Algorithm 1. Procedure for computing the classical-statistics objective function

Input: Parameter vector ($M_{\rm disk}$, $h_{z,\odot}$, $z_\odot$, $v_{z,\odot}$) and bootstrap sample of parent stellar sample
1. Compute the halo mass at fixed $v_{\rm circ}(R_\odot) = 229$ km s^{-1}
2. Compute Galactocentric positions and velocities for each star
3. Compute Stäckel potential focal length parameter for each star
4. Compute actions and angles for all stars (Stäckel fudge)
5. Compute action-local mean abundance deviation for each star (Equation (2))
6. Use linear least-squares to compute the optimal Fourier parameters (Equation (3))
7. Compute the objective function (Equation (8)).

Figure 5 shows smooth fits of the above model computed from bootstrapped resamplings of the data (blue curves) and, for visualization only, we show binned means of the measured abundance-ratio deviations (black histogram). We emphasize that no binning is performed at any time in performing these fits. The results shown in Figure 5 are for abundance deviations in [O/Fe], and just for three particular settings of the disk mass parameters, $M_{\rm disk}/M^{\star}_{\rm disk}$(with all other mass model parameters fixed). The data are bootstrapped prior to the construction of the abundance deviations, because the abundance-deviation estimates depend on the data set in play, and also the mass-model parameters. Note that the sign of the $m = 2$ term flips

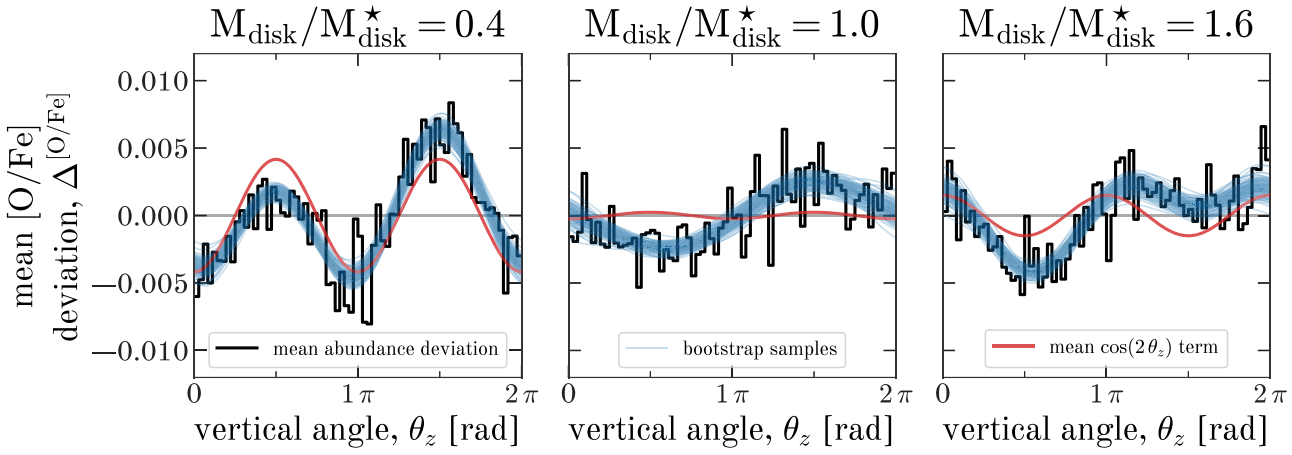

Figure 5. Mean [O/Fe] abundance deviation, $\Delta^{[O/Fe]}$, as a function of vertical angle for three different values of the mass of the disk (at fixed circular velocity at the solar circle). The abundance deviation for each star in the sample is the difference between the abundance measured in each star and a mean of the $K = 64$ nearest neighbors to that star in three-dimensional action space (for that mass model). In each panel, the mean abundance deviation is shown three ways. The black histogram shows the mean of the abundance deviation in small bins in vertical angle (binned for visualization purposes only: the binned data are never used in the analysis). The blue lines show continuous fits to the unbinned data, which are continuous linear combinations of sines and cosines (see Section 6); there are 128 blue lines, one for each of 128 independent bootstrap trials. The red line shows the mean of the $\cos 2\theta_z$ terms across the 128 bootstrap trials: the amplitude of this parameter is an indication of the goodness of fit of a particular choice of mass model parameters. A "perfect" fit would produce a red curve that is flat or has minimal amplitude. Comparing the three panels, the red curves show an amplitude of opposite sign in the highest-disk-mass panel relative to the other panels, suggesting that the best setting of the disk mass is in between the fiducial disk mass model and the higher-disk-mass model (as we find; see Section 6).

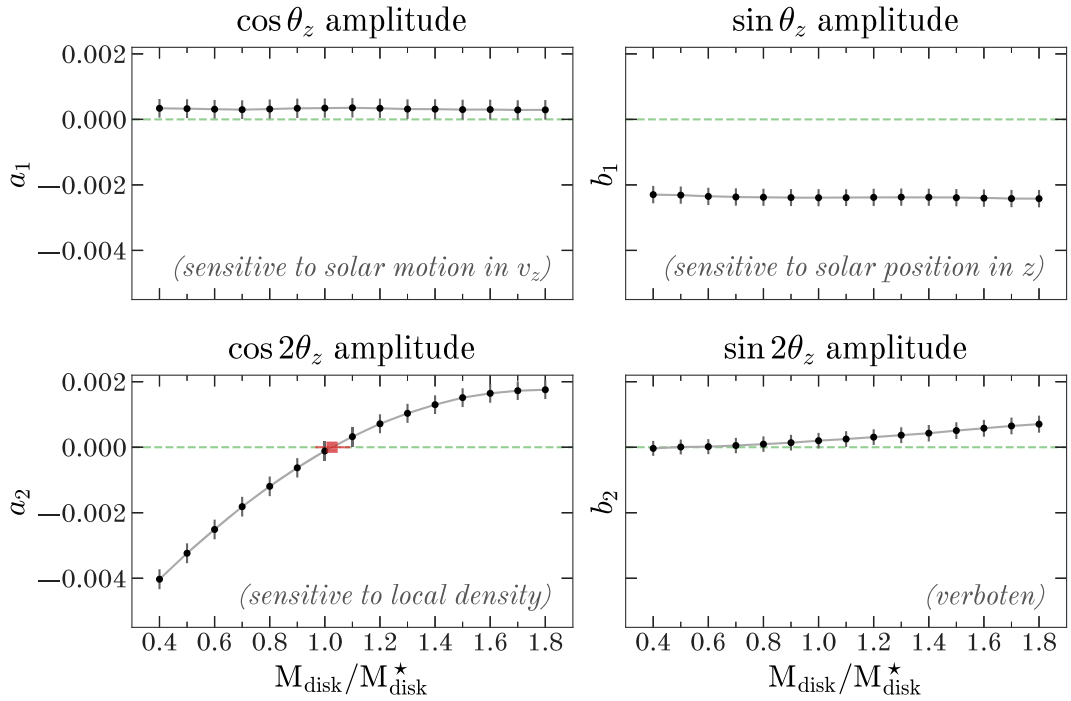

Figure 6. Parameters of the smooth sine-and-cosine fits to the dependence of [O/Fe] abundance deviation on vertical angle θ_z (the blue lines in Figure 5) as a function of the disk-mass parameter $M_{disk}/M^{\star}_{disk}$. The black markers show the values of the disk-mass parameter at which we performed the sine-and-cosine fits; the vertical error bars show the bootstrap uncertainties for each coefficient. The amplitude of the $\cos 2\theta_z$ term is the amplitude that is sensitive to the local density of the disk; it crosses zero when the model has the best-fit disk mass. The best-fit disk mass and its measurement uncertainty are shown as the square (red) marker and a horizontal error bar. The other terms shown have different dependencies on Milky Way parameters: the $\cos\theta_z$ would vary strongly if we varied the solar motion (the vertical component of the local standard of rest). The $\sin\theta_z$ term would vary strongly if we varied the location of the midplane of the disk. The $\sin 2\theta_z$ term cannot be nonzero; the fact that we find a nonzero value for this amplitude may suggest a weak violation of our model assumptions (see Section 6). These figures show that we can precisely measure the disk mass.

between the left and right panels of Figure 5, indicating that the best-fit mass model must be at an intermediate value of $M_{disk}/M^{\star}_{disk}$, close to but slightly larger than the fiducial value.

We repeat this fitting procedure for a grid of disk-mass parameter values $M_{disk}/M^{\star}_{disk} = 0.4$–1.8: for each mass model (i.e., each setting of $M_{disk}/M^{\star}_{disk}$), we estimate the Fourier coefficient parameters and uncertainties on the coefficients using 128 bootstrap trials. Figure 6 shows the inferred coefficients for the abundance ratio [O/Fe] as a function of disk mass $M_{disk}/M^{\star}_{disk}$. Conceptually, to turn this into a constraint on the disk mass, we then look for the value of $M_{disk}/M^{\star}_{disk}$ that minimizes the (absolute) value of a_2. We

obtain an estimate for the disk mass parameter $M_{\rm disk}/M^{\star}_{\rm disk}$ and an associated uncertainty by linearly interpolating the measurements and bootstrap error bars onto the $a_2 = 0$ intercept. Our best-fit disk mass and its uncertainty is shown as the square (red) marker in Figure 6 to emphasize that, though this is a simple model and we only vary one parameter in this demonstration, the measurement is encouragingly precise with a formal error bar of ≈7% for just a single element abundance. We note that here we also apparently measure a finite value for the $\sin 2\theta_z$ term, which should not exist in the universe defined by our assumptions (and is therefore labeled "verboten"). This implies that our assumptions seem to be lightly violated (the inferred amplitude is only marginally significant), and we discuss this further in the context of our assumptions below (Section 7).

6.2. Construction of a Mass Model Objective Function

The demonstrations above show how we can compute best-fitting coefficients for our model of the variations of the mean abundance deviations as a function of vertical angle θ_z. To use these coefficients in a more general fitting procedure (i.e., to derive constraints on the mass model parameters) we construct an objective function that we can then optimize over the mass model parameters. However, the upper panels of Figure 6 show that, at all values of the disk mass parameter $M_{\rm disk}/M^{\star}_{\rm disk}$, there is a finite $m = 1$ amplitude for both the cos and sin terms. As noted before, these coefficients are sensitive to the solar motion and solar position (in z), respectively. We therefore consider four parameters in our objective function: the solar position relative to the midplane $z_{\odot}$, the solar z velocity relative to the local standard of rest $v_{z,\odot}$, the disk mass parameter $M_{\rm disk}/M^{\star}_{\rm disk}$, and the disk scale height $h_{z,\odot}$(the Miyamoto–Nagai scale height parameter, not an exponential scale height; Miyamoto & Nagai 1975). For each setting of these parameters $M_{\rm disk}/M^{\star}_{\rm disk}$, $h_{z,\odot}$, $z_{\odot}$, $v_{z,\odot}$, we compute the Fourier coefficients (as described in the previous section) and minimize the objective function

$$f(a_1, b_1, a_2; M_{\rm disk}/M^{\star}_{\rm disk}, h_{z,\odot}, z_{\odot}, v_{z,\odot}) = a_1^2 + b_1^2 + a_2^2. \quad (8)$$

Note that here we ignore the constant term c_0 and the amplitude of the $\sin 2\theta_z$ term b_2, but we have verified that including these coefficients in the objective function does not significantly (within a few percent) change the inferred parameters for any of the elements used here. This choice of objective function (Equation (8)) is somewhat arbitrary, but is sufficient for following the "classical statistics" approach we take in this paper. A Bayesian or likelihood-based formulation of the ideas described here could instead construct a more physical model for $\Delta(\theta_z)$, and directly fit for, or sample over (with priors), the mass model and Milky Way parameters; we consider this to be out of scope for this work. Using the objective function above, we again perform 128 bootstrap trials per element, and perform the bootstrap resampling outside of the entire procedure (so that each optimization is performed independently with a bootstrap sample). We use Nelder–Mead optimization (Gao & Han 2012) to minimize our objective function, as implemented in the *scipy* package (Virtanen et al. 2020).

Figure 7 shows a summary of our constraints on the disk mass parameter and disk scale height for all of the elements shown in Figure 2 (colorful ellipses), and the joint constraint from all of these elements combined (dark, black ellipse). Here we show 1σ and 2σ error ellipses (darker and lighter ellipses for each color), with means and covariance matrices estimated from the optimization results for the bootstrap resamplings of the data for each abundance ratio.

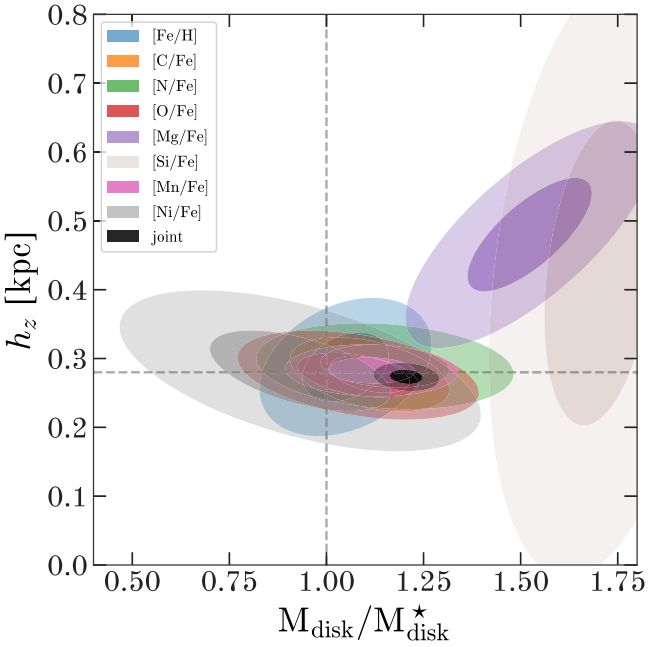

Figure 7. Summary of our joint constraints on the disk mass parameter $M_{\rm disk}/M^{\star}_{\rm disk}$ and the scale height $h_{z,\odot}$ for each element individually (colored ellipses) and the joint constraint (black). Each color shows 1σ and 2σ error ellipses for each abundance ratio. The means and covariance matrices of each error ellipse are computed from optimizing the objective function (Equation (8)) for 128 bootstrap resamplings of the parent sample. Each individual element-abundance ratio already provides fairly precise ($\lesssim$10%) constraints on these parameters, but the joint constraint for these eight elements has a precision of ≈2.5% for both of these parameters. The dashed vertical and horizontal lines indicate our fiducial values (Section 4).

Figure 8 is similar to Figure 7, but shows our joint constraints on the other projections of our parameter space: this figure shows that a few of the element-abundance ratios prefer a significantly different solar position relative to the midplane: while most elements are consistent with past measurements of the solar height of $z_{\odot} \sim 20$ pc (Bennett & Bovy 2019 and Bland-Hawthorn & Gerhard 2016 and references therein), both [Mg/Fe] and [Si/Fe] suggest that the Sun is on the opposite side of the midplane. While we do not have a simple explanation for this discrepancy, our constraints from different element-abundance ratios will effectively weight stars in different parts of action space in ways that may amplify issues with our assumptions. In particular, there are known asymmetries in the vertical density and kinematics of stars in the local disk that affect stars with different vertical actions with a different phase and amplitude. In general, structures that are coherent in orbital phase (e.g., the known substructure in the local velocity distribution, e.g., Hunt et al. 2018, and the vertical "phase-space spiral," e.g., Antoja et al. 2018). If the abundance-ratio gradients (with respect to vertical action) emphasize stars with different vertical actions, we will in general find disagreements between the results for different abundance ratios.

Our results are summarized in Table 1, where we list the joint constraints on our four parameters utilizing all eight abundance ratios (center column), or excluding [Mg/Fe] and [Si/Fe] (right column), which are clear outliers in their preferred solar position

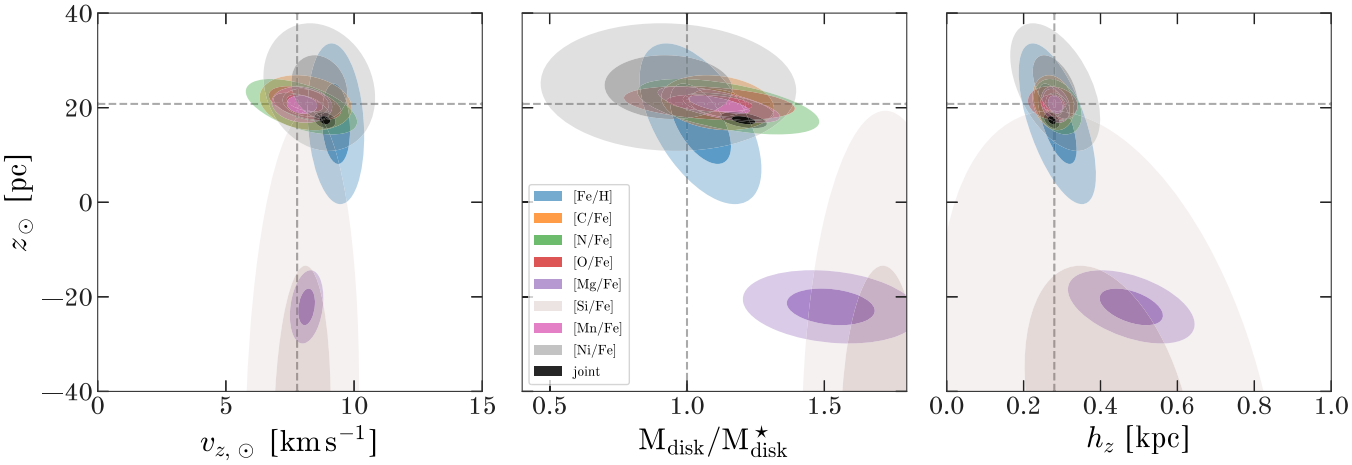

Figure 8. Same as Figure 7, but showing other projections of our parameter space. Note that in the solar height above the midplane parameter $z_\odot$, we find some significant disagreements between [Mg/Fe] and [Si/Fe] and the rest of the abundance ratios we consider. The dashed vertical and horizontal lines indicate our fiducial values (Section 4).

Table 1
A Summary of Our Results from Combining the Constraints on the Disk Mass, Scale Height, Solar Position, and Solar Motion from Eight Independent Element Abundance Ratios (Center Column)

Parameter	All Eight Abundance Ratios	Excluding [Mg/Fe], [Si/Fe]
$M_{disk}/M^\star_{disk}$	1.21 ± 0.04	1.07 ± 0.05
$h_{z,\odot}$	0.27 ± 0.01 kpc	0.28 ± 0.01 kpc
$z_\odot$	16.9 ± 0.8 pc	20.6 ± 0.9 pc
$v_{z,\odot}$	8.8 ± 0.2 km s^{-1}	8.4 ± 0.3 km s^{-1}
M_{disk}	$7.89 \pm 0.26 \times 10^{10}\,M_\odot$	$6.98 \pm 0.32 \times 10^{10}\,M_\odot$
$M_{disk}/M_{halo}(<8.1$ kpc)	2.11 ± 0.21	1.51 ± 0.16

Note. We also show joint results for all abundance ratios excluding [Mg/Fe] and [Si/Fe], which are clear outliers in their preferred solar position parameter values.

values. We also include two derived quantities: the total disk mass M_{disk}, and the disk to halo mass ratio within the solar circle $M_{disk}/M_{halo}(<8.1$ kpc), which we find is slightly larger than both our fiducial model ≈1.3 (Price-Whelan 2017) and the implied value for the `MWPotential2014` (Bovy 2015) of ≈1.7.

7. Discussion

Here we define and demonstrate a promising new method, Orbital Torus Imaging, for using measurements of both stellar kinematics and stellar surface abundances to infer the underlying mass distribution of the Milky Way. This method has promise in that it utilizes the high-dimensional stellar labels that are measured by spectroscopic surveys to improve dynamical inferences. Here we briefly review the method and its connection to other dynamical inference methods, return to the assumptions that underpin Orbital Torus Imaging and discuss their applicability in the Milky Way, and discuss possible extensions or reformulations of Orbital Torus Imaging (e.g., in a Bayesian context).

7.1. Fundamentals of Orbital Torus Imaging

The existence of element-abundance gradients in the Milky Way combined with the assumption that stars do not (rapidly) change their surface abundances as they orbit leads to the core concept of Orbital Torus Imaging: stellar abundance ratios can depend on the three invariant actions, but they cannot depend on the conjugate angles (or phases). An implication of this is that gradients of stellar abundances with respect to any six-dimensional phase-space coordinates will be locally tangent to the 3-tori defined by the surfaces of constant actions at any position in phase space. The six-dimensional phase space is foliated by a complete set of orbital 3-tori, each of which is specified by the three actions. Therefore, globally, each of these orbital tori will be a level 3-surface (in the six-dimensional phase space) of all moments or statistics of the element-abundance distribution.

From an information theory perspective, this implies that any predictive model for the element abundances of stars that depends on both actions and angles will not predict the stellar abundances more precisely than a model that only depends on the actions. That is, in the six-dimensional phase space, only the action coordinates provide information about the element abundances. In this context, we expect that the precision with which a mass model can be constrained is better (i.e., the uncertainties on parameters are smaller) when the abundance gradients are stronger. We also expect that the precision will increase as the width or dispersion of the local (in phase space) abundance distribution gets smaller. This dispersion will have contributions from the intrinsic distribution of abundances, which is related to both star formation and dynamical mixing, and also from observational uncertainties. An important consideration here is that inferences using Orbital Torus Imaging will not always continue to improve as the individual abundance measurements improve: at some point, the finite scatter of the (kinematically local) abundance distribution will cause the improvements to saturate. However, the precision will increase as the number of stars increases, especially as the stars cover more of the range of possible conjugate angles.

In this paper, in our toy implementation, we focus primarily on fitting a model of the means of (logarithmic) abundance ratios of stars. But Orbital Torus Imaging is much more general: any statistical moments of any statistics of any invariant stellar labels would also be allowed. The method also does not assume that the element abundances are uniquely or even specifically predicted by the actions. This method only requires that there be gradients in some moments of the abundance distribution. Generically, the abundance-ratio

distributions at any point in action space are expected to have (and are observed to have) substantial dispersion.

7.2. Connection to Other Methods

Like Jeans modeling, Schwarzschild modeling, or DF modeling, Orbital Torus Imaging provides a method for measuring the underlying acceleration field traced by stars for which we only have access to instantaneous measurements of phase-space coordinates. However, unlike Jeans modeling, Orbital Torus Imaging *does not* require separability or assumptions about symmetries of the acceleration field, nor does it require accurate measurements (or spatial derivatives) of second moments of the velocity distribution. We choose in this paper to focus on the vertical kinematics of stars, but we do not assume that motion in z is decoupled from the radial kinematics (i.e., the orbits plotted in Figure 4 show finite-width projections onto the vertical phase space). We therefore expect Orbital Torus Imaging to be more stable and less restrictive than Jeans methods.

Another significant advantage of Orbital Torus Imaging over other standard dynamical inference methods is that it does not require knowledge of the survey selection function, provided that the selection is made in position space (or phase space) and does not act strongly on the element-abundance ratios that enter into the inference. This condition is sufficiently satisfied for the APOGEE data used here, and will generally be satisfied for other spectroscopic surveys of stars in the Galaxy (e.g., GALAH; Martell et al. 2017; Buder et al. 2018).

In some ways, Orbital Torus Imaging is related to the concept of "extended distribution functions" (EDFs; e.g., Sanders & Binney 2015; Das & Binney 2016), in which the dynamical DF in action space $p(\boldsymbol{J})$ (sometimes written as $f(\boldsymbol{J})$) is extended into a DF in actions $\boldsymbol{J}$, abundances $\boldsymbol{X}$, and (perhaps) other (near-) invariants to make a joint DF $p(\boldsymbol{J}, \boldsymbol{X})$. So far, EDFs have been primarily used as tools to study the intrinsic properties of the stellar tracer density distributions and to constrain models of diffusive dynamical phenomena (e.g., radial migration). In other words, EDFs have been used in a mode where the Galactic mass model is fixed and the tracer kinematics are used to constrain the intrinsic chemical and kinematic structure of the tracer populations, but they have not yet been used to infer the mass model for the Galaxy. While it is possible, in principle, to use EDFs to simultaneously constrain the mass model, most applications would require precise knowledge of the survey selection functions (for any data used) and would therefore be challenging to implement with existing survey data.

Orbital Torus Imaging is also closely related to the concept of "mono-abundance populations" (MAPs; e.g., Bovy et al. 2012; Bovy & Rix 2013; Bovy et al. 2016; Mackereth & Bovy 2020), which have been used to study the structure of the galaxy divided into sub-populations that each have similar element abundances. In these applications, analyses are typically carried out independently for each MAP and therefore do not make use of any gradients in the element abundances with respect to the kinematics. Studies that make use of MAPs can therefore be seen as factorizing a conditional distribution $p(\boldsymbol{J} \mid \boldsymbol{X})$ out of the extended DF $p(\boldsymbol{J}, \boldsymbol{X})$ (although this is not precisely how they were formulated or described).

In contrast to EDFs and MAPs, Orbital Torus Imaging can be seen as factorizing out (and inferring) the conditional distribution $p(\boldsymbol{X} \mid \boldsymbol{J})$ from the joint DF $p(\boldsymbol{J}, \boldsymbol{X})$. The great advantage of this factorization is that it does not require knowledge of the survey selection function—it is *conditioned* on the actions $\boldsymbol{J}$, which are functions of the positions and velocities of stars—provided that the selection acts in position- or action-space alone (and not abundance-space).

The closest related dynamical methodology we are aware of is *Orbital Roulette* (Beloborodov & Levin 2004). Like Orbital Torus Imaging, *Orbital Roulette* assumes that objects are not observed at a special time and therefore that the distributions of phases for all tracers in some observed population should be uniform. This assumption is then turned into an estimator, which can be used to optimize for the parameters of a mass model by making the distribution of phases approach uniformity. Our assumption is weaker: we only assume that the distribution of abundances is invariant of phase.

7.3. Returning to Our Assumptions

Like any dynamical inference method that seeks to measure properties of a mass distribution from instantaneous measurements of tracer positions and velocities, we have relied on a set of strong assumptions to formulate Orbital Torus Imaging (Section 6). Already in our demonstrations above we see suggestions that these assumptions are violated by the data in hand: for example, some elements prefer a finite $\sin 2\theta_z$ amplitude (as shown in the lower-right panel of Figure 6), and the inferred solar position relative to the Galactic midplane is significantly different for [Mg/Fe] and [Si/Fe] as compared to the other elements (as shown in Figure 8).

For our particular toy application of Orbital Torus Imaging to the APOGEE sample defined in this paper, and for more general applications using stars in the Milky Way disk or halo, the "phase mixed" assumption is likely the most strongly violated of our list of assumptions. The Galactic disk is now known to show clear signatures of external perturbations, intrinsic time-dependent phenomena, and significant phase-space substructure (e.g., Antoja et al. 2018; Hunt et al. 2018; Schönrich & Dehnen 2018; Kamdar et al. 2019; Khanna et al. 2019; Monari et al. 2019; Coronado et al. 2020; Laporte et al. 2020; Poggio et al. 2020), all of which reflect the extent to which the stars in the disk are not phase mixed. In the case of vertical kinematics, the most striking illustration of this is the recently discovered phase-space spiral (Antoja et al. 2018). The spiral itself is only a $\approx$10% perturbation to the local DF (Laporte et al. 2019), providing some limit to the amount with which this will impact inferences with Orbital Torus Imaging. In principle, future implementations could make use of the spiral by explicitly modeling the perturbation and its dependence on stellar kinematics and position and stellar labels. That is out of scope for this paper, but suggests possible generalizations of Orbital Torus Imaging that account for stellar populations that are incompletely angle-mixed. Beyond the Galactic disk in the stellar halo, as with Jeans modeling, Orbital Torus Imaging will be biased by the existence of unmixed substructures (streams, shells, and merger remnants; e.g., Grillmair & Carlin 2016; Shipp et al. 2018).

Another key assumption that could lead to significant biases for certain stellar samples is our assumption that the selection function only depends on phase-space coordinates and not element abundances. A different way of stating this assumption is that we require element abundances to be measured and calibrated equivalently (in a statistical sense) at all orbital phases. In general, this assumption can be violated through a combination of subtle

effects: if a survey selection function implicitly depends on luminosity or temperature, and there are systematic trends in the measured abundances with surface gravity or temperature, the abundance distributions in any location of phase space will appear to be different from these effects alone. For the APOGEE sample used here, we have attempted to mitigate these issues by selecting a small range of surface gravities (and therefore temperatures) along the giant branch. However, there are known systematic trends in APOGEE between abundances and stellar parameters (Jönsson et al. 2020; Wheeler et al. 2020; A. C. Eilers et al. 2021, in preparation). In principle, a joint dynamics and calibration model could be made that simultaneously fits for the abundances as a function of conjugate angle and also housekeeping data (like stellar surface gravity or spectrograph line-spread function) that cause abundance systematics.

In our current implementation we have made use of transformations to action–angle coordinates, which places implicit constraints on the mass models we can consider. In particular, any mass models must be integrable and not dominated by resonances or chaotic regions, as we assume that the actions provide a continuous foliation of the phase space. At the precision with which we are operating, we do not think that this will be a dominant issue in our analyses. However, we are interested in the prospect of moving away from parameterizing the mass model directly, as with a sufficiently large sample of stars the abundance gradients should be enough to image the orbital tori directly.

As mentioned in Section 6, dynamically, Orbital Torus Imaging only strictly depends on the assumptions of phase-mixing and integrability discussed above, which could be satisfied even in weakly time-dependent potentials in which the actions remain adiabatically invariant. However, there is substantial evidence for nonadiabatic time-dependence in the dynamics of stars throughout the Galactic disk (e.g., Price-Whelan et al. 2015; Antoja et al. 2018; Xu et al. 2020). In practice, time-dependence in the Milky Way could therefore cause the actions to oscillate (e.g., Peñarrubia 2013; Buist & Helmi 2015), possibly with an action-dependent amplitude (Burger et al. 2020), which would lead to a change in the mean abundance distribution as a function of actions. Our implicit assumption that the potential is time-independent (through other core assumptions) could therefore lead to biases in the inferred parameters of the adopted time-independent models; This is a limitation of this work, and it will be important to quantify the magnitude of the expected biases. However, because our assumptions are slightly weaker, these biases may end up being less severe than for other methods that strictly require time-independence (e.g., Jeans modeling).

Much of the core methodology described here relies on the assumption that stellar surface element abundances are invariants. As mentioned above in our original list of assumptions, this is known to be violated in detail due to gravitational settling, stellar evolution, or planetary engulfment. However, as long as the timescales over which surface abundances can change are much shorter or much longer than the Galactic orbital timescales, the distributions of element abundances at any orbital phase should not be affected by these subtleties. We therefore expect this to be inconsequential.

Finally, we have assumed all along that there are measurable gradients in stellar abundances with respect to kinematics in the Galaxy or stellar population of interest. In the Milky Way, these are readily observed with existing spectroscopic surveys, and our measurements of individual element abundances will only become more numerous and precise in the coming years.

7.4. Extensions of Orbital Torus Imaging

The implementation of Orbital Torus Imaging used in this paper falls into the category of classical (frequentist) statistics. However, the concepts introduced here also lead naturally to probabilistic generative model (Bayesian) versions of Orbital Torus Imaging. Bayesian implementations of this method at first look very different: instead of trying to minimize the dependence of abundance-ratio statistics on the conjugate angles (as we do here), Bayesian formulations would instead involve building a flexible forward model for the element-abundance distribution as a function of the actions alone. A non-intuitive aspect of this is that the angles would never explicitly appear in the probabilistic model. The Bayesian implementations would therefore look conceptually like extended DFs (Sanders & Binney 2015), but while also varying the mass models, or forward models of the DF (e.g., Magorrian 2014), where element abundances or stellar labels are treated as additional invariants.

In the implementation presented here, we use an abundance deviation that is based on a nearest-neighbor interpolation in action space (see Section 6 and Appendix). That is a blunt tool; in principle, our results would improve if we instead explicitly modeled the abundance distribution moments as functions of the actions. One option for this modeling would be to use a generative, physical model for the star formation history of the Galaxy (or sample), including the products of nucleosynthesis and effects of chemical enrichment and stellar migration (similar to what is done in Sanders & Binney 2015). Another option would be to use a flexible machine-learning method, such as a Gaussian process.

We have parameterized the Milky Way mass model with a very simple, few-component mass model. Another option would be to instead parameterize the torus foliation of phase space directly. The representation of this foliation could be very general: any foliation of phase space with closed 3-tori is, in principle, allowed by dynamics. A direct reconstruction of the foliation could then be interpreted in terms of the force law, and hence the potential or mass model. That would be a more data-driven approach—with far fewer assumptions—than what has been executed here.

We used the means of a set of hand-chosen abundance-ratio deviations as our invariants for the inferences. There are many other choices we could have made. For example, we could have searched for maximally informative labels from among the abundance ratios or functional combinations thereof. We could have used moments other than the mean. And we could have used nonabundance labels, such as stellar ages, angular momenta, or even binary-companion properties. We could even have used the *other actions* J_R, J_ϕ as invariant labels for the vertical-angle θ_z fits. That is an interesting thought for future work, but the actions can only be used as labels here if the survey selection function is precisely known and accounted for. That condition negates one of the principal advantages of Orbital Torus Imaging over other methods.

Lastly, here we have focused on the vertical kinematics of stars through the vertical action J_z and angle θ_z, but in principle this methodology—and any Bayesian extensions—would also work in all three actions and angles. In practice, the classical statistics approach would not work with the azimuthal angle θ_ϕ

or action J_ϕ because our sample spans a small range of θ_ϕ (local to ≈2 kpc around the Sun). However, we could have equally used the radial action J_R and angle θ_R here for demonstrations.

8. Conclusions

We present a new method—Orbital Torus Imaging—for inferring the mass model (or acceleration field) traced by a phase-mixed stellar distribution. Orbital Torus Imaging is novel in that it provides a way of using measurements of element-abundance ratios and other stellar labels to improve the precision of dynamical inferences. This method is also more statistically robust than traditional methods (e.g., Jeans modeling) in that it does not require the measurement of second moments of the stellar velocity distribution or knowledge of the spatial (or phase space) survey selection function. This method does, however, depend on the existence of element-abundance gradients with respect to kinematics and our ability to measure these gradients: the stars on different orbits (i.e., with different actions) must have—on average or statistically—different compositions. Fortunately these gradients are ubiquitous and measurable in the Milky Way (and most galaxies; see Figure 2).

We outline and implement a classical-statistics approach to Orbital Torus Imaging and apply this to the vertical kinematics of a subsample of RGB stars from the APOGEE surveys. The fundamental concept of this method is that stars do not change their abundances appreciably over timescales comparable to Galactic orbital timescales. With that, and under the assumption that the stellar populations in the Galactic disk are phase-mixed, this implies that the stellar abundance distribution at any place in phase space must be independent of phase or conjugate angles and can only be a function of dynamical invariants, like orbital actions. In our current implementation, we explicitly try to find a setting of the gravitational potential parameters disk mass $M_{\rm disk}$ and scale height $h_{z,\odot}$ (and reference frame parameters, the solar position and velocity in z) that minimize the dependence of the abundance distribution on the vertical angle θ_z (see Figure 5). Turning this into an objective function (Section 6.2), we then optimize over the four Milky Way parameters using eight independent element-abundance ratios (see Figure 7). Individually, the constraints are already encouragingly precise, but combining the different abundance ratios provides joint constraints on the four parameters that are precise to within a few percent (see Table 1).

Ongoing and near future spectroscopic surveys (e.g., SDSS-V, GALAH, DESI, 4MOST, WEAVE) will provide samples of stars that are a factor of ≈10–100 times larger than that used in this work. Combining these stellar data with kinematic measurements from upcoming data releases from the Gaia mission will enable extremely precise constraints on the Galactic mass distribution using Orbital Torus Imaging.

It is a pleasure to thank Jo Bovy (Toronto), Anna-Christina Eilers (MIT), José G. Fernández-Trincado (U de Atacama), Suroor S. Gandhi (NYU), Matt Shetrone (UCO/Lick), David Spergel (Flatiron), Eugene Vasiliev (Cambridge), Adam Wheeler (Columbia), the Dynamics and Astronomical Data groups at the Flatiron Institute, and the Galaxy group at the MPIA for valuable discussions and input. We thank the anonymous referee for valuable, constructive feedback that improved this paper. This research was conducted in part at the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-1607611. K.V.J.'s contributions were enabled by NSF grant AST-1715582. S.H. is supported by an NSF Astronomy and Astrophysics Postdoctoral Fellowship under award AST-1801940. DAGH acknowledges support from the State Research Agency (AEI) of the Spanish Ministry of Science, Innovation and Universities (MCIU) and the European Regional Development Fund (FEDER) under grant AYA2017-88254-P.

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions.

SDSS-IV acknowledges support and resources from the Center for High Performance Computing at the University of Utah. The SDSS website is www.sdss.org.

SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, Center for Astrophysics | Harvard & Smithsonian, the Chilean Participation Group, the French Participation Group, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU)/University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatário Nacional/MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

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.

Software: Astropy (Astropy Collaboration et al. 2013, 2018), *gala* (Price-Whelan 2017), *IPython* (Pérez & Granger 2007), *matplotlib* (Hunter 2007), *numpy* (Harris et al. 2020), *schwimmbad* (Price-Whelan & Foreman-Mackey 2017), *scipy* (Virtanen et al. 2020).

Appendix
Re-weighting the K Nearest Neighbors to Account for Steep Gradients

In standard K-nearest-neighbor regression, the estimate of the (let us say scalar) label y_* for a vector test point $\boldsymbol{x}_*$ is the naïve mean $\langle y \rangle$ of the training labels y_k for the K nearest training-set neighbors k in the vector space $\boldsymbol{x}$:

$$y_* \leftarrow \langle y \rangle \equiv \frac{1}{K} \sum_{k=1}^{K} y_k. \tag{A1}$$

Issues arise when the test point $\boldsymbol{x}_*$ is near the edge of the training set or at a location of strong gradients in the density in

the training set. In these cases, the naïve mean position $\langle \boldsymbol{x} \rangle$ of the K neighbors

$$\langle \boldsymbol{x} \rangle \equiv \frac{1}{K} \sum_{k=1}^{K} \boldsymbol{x}_k \tag{A2}$$

will be substantially displaced from the test point $\boldsymbol{x}_*$, and the naïve mean of the labels y_k will not be appropriate for the test position.

We can correct for this problem by replacing the naïve mean of the labels with a more sophisticated weighted mean. We begin by defining scalar displacements ξ_k of the neighbors away from the test point

$$\xi_k \equiv (\boldsymbol{x}_k - \boldsymbol{x}_*) \cdot (\langle \boldsymbol{x} \rangle - \boldsymbol{x}_*), \tag{A3}$$

which are the individual neighbor vector displacements projected onto the displacement between the mean position and the test position. Then we fit (by linear least squares) a linear model of the form

$$y_k = a + b\, \xi_k + \text{noise}, \tag{A4}$$

where a is an intercept and b is a slope. The intercept a is the linear fit interpolated to the position $\boldsymbol{x}_*$ of the test point. In detail because this linear fit is just a two-parameter least-square fit, it has a simple closed form:

$$y_* \leftarrow \hat{a} = \frac{\sum_{k=1}^{K} w_k\, y_k}{\sum_{k=1}^{K} w_k} \tag{A5}$$

$$w_k \equiv \xi_k \sum_{j=1}^{K} \xi_j - \sum_{j=1}^{K} \xi_j^2, \tag{A6}$$

where $\hat{a}$ is the least-squares estimate of the intercept a, and that estimate is itself a nontrivial weighted sum of the data with weights w_k. One implementation note: in the (rare) edge case that the displacement $\langle \boldsymbol{x} \rangle - \boldsymbol{x}_*$ underflows the floating-point representation, the weighted mean should be replaced with the naïve mean $\langle y \rangle$.

This new estimate $y_* = \hat{a}$ is much more accurate than the naïve mean $\langle y \rangle$ in the presence of gradients in the training set. It is, in some sense, the first-order correction of naïve K-nearest neighbors. It is the first in a series of corrections to account for nontrivial training-set distributions.

ORCID iDs

Adrian M. Price-Whelan https://orcid.org/0000-0003-0872-7098
David W. Hogg https://orcid.org/0000-0003-2866-9403
Kathryn V. Johnston https://orcid.org/0000-0001-6244-6727
Melissa K. Ness https://orcid.org/0000-0001-5082-6693
Hans-Walter Rix https://orcid.org/0000-0003-4996-9069
Rachael L. Beaton https://orcid.org/0000-0002-1691-8217
Joel R. Brownstein https://orcid.org/0000-0002-8725-1069
D. A. García-Hernández https://orcid.org/0000-0002-1693-2721
Sten Hasselquist https://orcid.org/0000-0001-5388-0994
Christian R. Hayes https://orcid.org/0000-0003-2969-2445
Matthew Shetrone https://orcid.org/0000-0003-0509-2656
Gail Zasowski https://orcid.org/0000-0001-6761-9359

References

Ahumada, R., Allende Prieto, C., Almeida, A., et al. 2020, ApJS, 249, 3
Antoja, T., Helmi, A., Romero-Gómez, M., et al. 2018, Natur, 561, 360
Astropy Collaboration, Price-Whelan, A. M., Sipócz, B. M., et al. 2018, AJ, 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Bahcall, J. N. 1984, ApJ, 276, 169
Bailer-Jones, C. A. L. 2015, PASP, 127, 994
Beloborodov, A. M., & Levin, Y. 2004, ApJ, 613, 224
Bennett, M., & Bovy, J. 2019, MNRAS, 482, 1417
Binney, J. 2012, MNRAS, 426, 1324
Binney, J., Burnett, B., Kordopatis, G., et al. 2014, MNRAS, 439, 1231
Binney, J., & Sanders, J. L. 2016, AN, 337, 939
Binney, J., & Tremaine, S. 2008, Galactic Dynamics, Second Edition (Princeton, NJ: Princeton Univ. Press)
Bland-Hawthorn, J., & Gerhard, O. 2016, ARA&A, 54, 529
Blanton, M. R., Bershady, M. A., Abolfathi, B., et al. 2017, AJ, 154, 28
Bonaca, A., & Hogg, D. W. 2018, ApJ, 867, 101
Bovy, J. 2015, ApJS, 216, 29
Bovy, J., Murray, I., & Hogg, D. W. 2010, ApJ, 711, 1157
Bovy, J., Nidever, D. L., Rix, H.-W., et al. 2014, ApJ, 790, 127
Bovy, J., & Rix, H.-W. 2013, ApJ, 779, 115
Bovy, J., Rix, H.-W., Liu, C., et al. 2012, ApJ, 753, 148
Bovy, J., Rix, H.-W., Schlafly, E. F., et al. 2016, ApJ, 823, 30
Bowen, I. S., & Vaughan, A. H. J. 1973, ApOpt, 12, 1430
Buch, J., Leung, J. S. C., & Fan, J. 2019, JCAP, 2019, 026
Buckley, M. R., & Peter, A. H. G. 2018, PhR, 761, 1
Buder, S., Asplund, M., Duong, L., et al. 2018, MNRAS, 478, 4513
Buist, H. J. T., & Helmi, A. 2015, A&A, 584, A120
Bullock, J. S., & Boylan-Kolchin, M. 2017, ARA&A, 55, 343
Burger, J. D., Peñarrubia, J., & Zavala, J. 2020, arXiv:2012.00737
Coronado, J., Rix, H.-W., Trick, W. H., et al. 2020, MNRAS, 495, 4098
Das, P., & Binney, J. 2016, MNRAS, 460, 1725
Deng, L.-C., Newberg, H. J., Liu, C., et al. 2012, RAA, 12, 735
Drimmel, R., & Poggio, E. 2018, RNAAS, 2, 210
Eilers, A.-C., Hogg, D. W., Rix, H.-W., et al. 2020, ApJ, 900, 186
Eilers, A.-C., Hogg, D. W., Rix, H.-W., & Ness, M. K. 2019, ApJ, 871, 120
Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, AJ, 142, 72
Evans, N. W., An, J., & Walker, M. G. 2009, MNRAS, 393, L50
Eyre, A., & Binney, J. 2011, MNRAS, 413, 1852
Freeman, K., & Bland-Hawthorn, J. 2002, ARA&A, 40, 487
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018a, A&A, 616, A1
Gaia Collaboration, Katz, D., Antoja, T., et al. 2018b, A&A, 616, A11
Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1
Gao, F., & Han, L. 2012, Comput Optim Appl, 51, 259
García Pérez, A. E., Allende Prieto, C., Holtzman, J. A., et al. 2016, AJ, 151, 144
Gravity Collaboration, Abuter, R., Amorim, A., et al. 2018, A&A, 615, L15
Grillmair, C. J., & Carlin, J. L. 2016, ASSL, 420, 87
Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, AJ, 131, 2332
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Natur, 585, 357
Hayden, M. R., Bovy, J., Holtzman, J. A., et al. 2015, ApJ, 808, 132
Helmi, A., & White, S. D. M. 1999, MNRAS, 307, 495
Hernquist, L. 1990, ApJ, 356, 359
Holtzman, J. A., Hasselquist, S., Shetrone, M., et al. 2018, AJ, 156, 125
Hunt, J. A. S., Hong, J., Bovy, J., Kawata, D., & Grand, R. J. J. 2018, MNRAS, 481, 3794
Hunter, J. D. 2007, CSE, 9, 90
Iben, I. J. 1965, ApJ, 142, 1447
Iorio, G., & Belokurov, V. 2021, MNRAS, 502, 5686
Jeans, J. H. 1919, Problems of Cosmogony and Stellar Dynamics (Cambridge: Cambridge Univ. Press)
Jeans, J. H. 1922, MNRAS, 82, 122
Johnston, K. V., Zhao, H., Spergel, D. N., & Hernquist, L. 1999, ApJL, 512, L109
Jönsson, H., Holtzman, J. A., Prieto, C. A., et al. 2020, AJ, 160, 120
Kamdar, H., Conroy, C., Ting, Y.-S., et al. 2019, ApJL, 884, L42
Kepler, J. 1609, Astronomia Nova (1st ed.; Heidelberg: G. Voegelinus)
Khanna, S., Sharma, S., Tepper-Garcia, T., et al. 2019, MNRAS, 489, 4962
Koppelman, H., Helmi, A., & Veljanoski, J. 2018, ApJL, 860, L11
Laporte, C. F. P., Belokurov, V., Koposov, S. E., Smith, M. C., & Hill, V. 2020, MNRAS, 492, L61

Laporte, C. F. P., Minchev, I., Johnston, K. V., & Gómez, F. A. 2019, MNRAS, 485, 3134
Lindegren, L., Hernández, J., Bombrun, A., et al. 2018, A&A, 616, A2
Mackereth, J. T., & Bovy, J. 2020, MNRAS, 492, 3631
Magorrian, J. 2014, MNRAS, 437, 2230
Magorrian, J. 2019, MNRAS, 484, 1166
Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, AJ, 154, 94
Marrese, P. M., Marinoni, S., Fabrizio, M., & Altavilla, G. 2019, A&A, 621, A144
Martell, S. L., Sharma, S., Buder, S., et al. 2017, MNRAS, 465, 3203
Martig, M., Fouesneau, M., Rix, H.-W., et al. 2016, MNRAS, 456, 3655
McMillan, P. J., & Binney, J. J. 2013, MNRAS, 433, 1411
Miyamoto, M., & Nagai, R. 1975, PASJ, 27, 533
Monari, G., Famaey, B., Siebert, A., Wegg, C., & Gerhard, O. 2019, A&A, 626, A41
Myeong, G. C., Evans, N. W., Belokurov, V., Sand ERS, J. L., & Koposov, S. E. 2018, ApJL, 856, L26
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563
Newton, I. 1687, Philosophiæ Naturalis Principia Mathematica (1st ed.; London: Joseph Streater for the Royal Society)
Nidever, D. L., Holtzman, J. A., Allende Prieto, C., et al. 2015, AJ, 150, 173
Oort, J. H. 1932, BAN, 6, 249
Peñarrubia, J. 2013, MNRAS, 433, 2576
Pérez, F., & Granger, B. E. 2007, CSE, 9, 21
Poggio, E., Drimmel, R., Andrae, R., et al. 2020, NatAs, 4, 590
Price-Whelan, A. M. 2017, JOSS, 2, 388
Price-Whelan, A. M., & Foreman-Mackey, D. 2017, JOSS, 2, 357
Price-Whelan, A. M., Hogg, D. W., Johnston, K. V., & Hendel, D. 2014, ApJ, 794, 4
Price-Whelan, A. M., Johnston, K. V., Sheffield, A. A., Laporte, C. F. P., & Sesar, B. 2015, MNRAS, 452, 676
Queiroz, A. B. A., Anders, F., Chiappini, C., et al. 2020, A&A, 638, A76
Read, J. I. 2014, JPhG, 41, 063101
Reid, M. J., & Brunthaler, A. 2004, ApJ, 616, 872
Romanowsky, A. J., Douglas, N. G., Arnaboldi, M., et al. 2003, Sci, 301, 1696
Sanders, J. 2012, MNRAS, 426, 128
Sanders, J. L., & Binney, J. 2013, MNRAS, 433, 1813
Sanders, J. L., & Binney, J. 2014, MNRAS, 441, 3284
Sanders, J. L., & Binney, J. 2015, MNRAS, 449, 3479
Sanders, J. L., & Binney, J. 2016, MNRAS, 457, 2107
Schönrich, R., & Dehnen, W. 2018, MNRAS, 478, 3809
Shipp, N., Drlica-Wagner, A., Balbinot, E., et al. 2018, ApJ, 862, 114
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261
Walker, M. G., & Peñarrubia, J. 2011, ApJ, 742, 20
Watkins, L. L., Evans, N. W., & An, J. H. 2010, MNRAS, 406, 264
Wheeler, A., Ness, M., Buder, S., et al. 2020, ApJ, 898, 58
Wilson, J. C., Hearty, F. R., Skrutskie, M. F., et al. 2019, PASP, 131, 055001
Xu, Y., Liu, C., Tian, H., et al. 2020, ApJ, 905, 6
Zasowski, G., Cohen, R. E., Chojnowski, S. D., et al. 2017, AJ, 154, 198
Zasowski, G., Johnson, J. A., Frinchaboy, P. M., et al. 2013, AJ, 146, 81
Zhai, M., Xue, X.-X., Zhang, L., et al. 2018, RAA, 18, 113