The Astrophysical Journal Supplement Series, 254:25 (14pp), 2021 June https://doi.org/10.3847/1538-4365/abe6a0

Dynamical Modeling of Galaxies and Supermassive Black Holes: Axisymmetry in Triaxial Schwarzschild Orbit Superposition Models

Matthew E. Quenneville[1,2], Christopher M. Liepold[1,2], and Chung-Pei Ma[1,2]
[1] Department of Astronomy, University of California, Berkeley, CA 94720, USA; mquenneville@berkeley.edu
[2] Department of Physics, University of California, Berkeley, CA 94720, USA
Received 2020 April 30; revised 2021 January 29; accepted 2021 February 14; published 2021 May 17

Abstract

We present a detailed analysis of the behavior of the triaxial Schwarzschild orbit superposition method near the axisymmetric limit. Orbit superposition modeling is the primary method used to determine dynamical masses of supermassive black holes ($M_{\rm BH}$) in nearby galaxies; however, prior studies have reported conflicting results when comparing the outcome from axisymmetric orbit codes with that from a triaxial orbit code in the axisymmetric limit. We show that in order to achieve (oblate) axisymmetry in a triaxial code, care needs to be taken to axisymmetrize the short-axis tube orbits and to exclude both the long-axis tube and box orbits from the orbit library. Using up to 12 Gauss–Hermite moments of the line-of-sight velocity distributions as constraints, we demonstrate the effects of orbit types on the best-fit $M_{\rm BH}$ in orbit modeling of the massive elliptical galaxy NGC 1453 reported in Liepold et al. In addition, we verify the efficacy of our updated code on a mock galaxy data set. We identify a subset of slowly precessing quasi-planar orbits for which the typical integration times can be insufficient to fully capture the equilibrium orbital behavior in both axisymmetric and triaxial systems with central black holes. Further investigation is needed for a more reliable treatment of these orbits.

Unified Astronomy Thesaurus concepts: Elliptical galaxies (456); Galaxies (573); Galaxy dynamics (591); Galaxy evolution (594); Galaxy kinematics (602); Galaxy structure (622); Black holes (162); Supermassive black holes (1663)

1. Introduction

The orbit superposition method of Schwarzschild (1979) enables efficient construction of self-consistent and equilibrium mass models of galaxies. The basic procedure consists of two steps: integrating a representative set of orbits in a static triaxial gravitational potential, and finding weights for these orbits such that their superposition reproduces the assumed mass distribution.

The orbit superposition method has been extended to include kinematic information and used to determine mass distributions in real galaxies, starting with studies such as Pfenniger (1984), Richstone & Tremaine (1984, 1985), and Rix et al. (1997). From the quality of the fit to both kinematic and photometric data, this method can be used to assess the relative likelihood of a range of mass models and to determine best-fit mass parameters such as $M_{\rm BH}$, stellar mass-to-light ratios, galaxy shapes, and dark matter halo parameters.

Due to the large number of orbits needed to sample the relevant phase space, the orbit superposition method is computationally expensive. To reduce the number of orbits and the dimensions of the model parameter space, a few orbit-based numerical codes have been developed for axisymmetric systems (e.g., Cretton et al. 1999; Gebhardt et al. 2000; Thomas et al. 2004; Valluri et al. 2004; Cappellari et al. 2006). Many dynamical measurements of $M_{\rm BH}$ from stellar kinematics have been obtained using these axisymmetric orbit codes.

Triaxiality allows for more general galaxy shapes and additional orbit types, but modeling orbits in triaxial potentials comes at the cost of increased complexity and computation time. van den Bosch et al. (2008) presented a triaxial orbit-based code capable of comparing directly to observations, using an orbital sampling scheme based on Schwarzschild (1993). van de Ven et al. (2008) performed recovery tests of this code for analytically tractable triaxial potentials (excluding central black holes). Only a handful dynamical determinations of $M_{\rm BH}$ have been obtained using triaxial models from this code (van den Bosch & de Zeeuw 2010; Walsh et al. 2012; Feldmeier-Krause et al. 2017). Several additional $M_{\rm BH}$ were determined using this code in the (nearly) axisymmetric limit (Seth et al. 2014; Walsh et al. 2015, 2016, 2017; Ahn et al. 2018). This code has also been used to construct axisymmetric and triaxial galaxy models to determine stellar dynamics and dark matter distributions for a wide range of galaxies (e.g., Zhu et al. 2018; Poci et al. 2019; Jin et al. 2020). Vasiliev & Valluri (2020) recently presented a new triaxial orbit-based code using a different method for phase space sampling and orbit initialization; the method was tested on mock data but had not been applied to real data.

An important test of the orbit superposition codes is the ability to produce consistent results between an axisymmetric code and a triaxial code in the axisymmetric limit. We note that the code by van den Bosch et al. (2008) is written for triaxial potentials and "is not capable of making a perfectly axisymmetric model" (van den Bosch & de Zeeuw 2010). Studies that attempt to run it near axisymmetry and then compare with results from axisymmetric codes have reached conflicting conclusions. For instance, van den Bosch & de Zeeuw (2010) used their triaxial code to construct (nearly) axisymmetric models for M32 and NGC3379, and found the mass-to-light ratios and $M_{\rm BH}$ to be consistent with those from earlier studies using axisymmetric codes (van der Marel et al. 1998; Joseph et al. 2001; Verolme et al. 2002 for M32; Gebhardt et al. 2000; Shapiro et al. 2006 for NGC 3379). Ahn et al. (2018), on the other hand, found a puzzling global χ^2 minimum at $M_{\rm BH} = 0$ while using this triaxial code to perform axisymmetric modeling of the ultracompact dwarf galaxy M59-UCD3. They found this minimum to be inconsistent with the best-fit nonzero

M_{BH} from Jeans modeling and the axisymmetric orbit code of Cappellari et al. (2006).

It is the purpose of our recent work (Liepold et al. 2020) and this paper to investigate how to modify the van den Bosch et al. (2008) code to enable it to handle properly both axisymmetric and triaxial systems. Since no galaxy in nature is likely to be exactly axisymmetric, it may appear that we are taking a step backwards in examining the axisymmetric limit of a triaxial code. While our next goal is indeed to adopt the more realistic triaxial potentials, we believe that one critical test of a triaxial code is its behavior in the simpler, axisymmetric limit. Such a study—the main goal of this paper—is a particularly important step in the quest for dynamical M_{BH} measurements in view of the fact that almost all existing M_{BH} measurements have been obtained assuming exact axisymmetry, and that the aforementioned recent comparison of axisymmetric and triaxial codes has led to unresolved conflicting results.

In Liepold et al. (2020), we described a set of recipes and code changes for achieving axisymmetry. We then performed proper axisymmetric orbit modeling using the revised code to obtain a new M_{BH} measurement for the massive elliptical galaxy NGC 1453, a fast rotator in the MASSIVE survey (Ma et al. 2014) well suited for axisymmetric orbit modeling. Similar to Ahn et al. (2018), we had encountered difficulties in constraining M_{BH} in NGC 1453 when we used the original code with comparable settings. Through extensive testing, we came to two main conclusions: (1) higher Gauss–Hermite moments (beyond the typically used h_4) of the line-of-sight velocity distributions (LOSVDs) are needed to fully constrain the orbital weights, and (2) the orbit libraries need to be modified to satisfy axisymmetry. The use of higher moments is described in detail in Liepold et al. (2020). Here, we focus on the construction of axisymmetric orbit libraries in a triaxial orbit code.

In this paper, we provide a full discussion of the required steps to axisymmetrize the model and the various modifications that we have implemented to the triaxial code by van den Bosch et al. (2008). The code was never given a name; we will refer to it as the TriOS ("Triaxial Orbit Superposition") code from this point on. In Section 2, we provide some background information about the implementation of the orbit superposition method in this code. We focus on four topics that are pertinent to subsequent discussions: the three major orbit types in a triaxial potential (Section 2.1), orbital sampling and initialization (Section 2.2), orbital integration (Section 2.3), and parameters used to quantify triaxial shapes (Section 2.4).

In Section 3, we give an in-depth discussion of the three main ingredients for axisymmetry listed in Section 4.1 of Liepold et al. (2020): axisymmetrization of short-axis tube orbits (Section 3.1), criteria for how to exclude long-axis tube orbits (Section 3.2), and exclusion of box orbits (Section 3.3).

We have made additional improvements and corrections to the code (Section 4). We identify a subset of slowly precessing quasi-planar orbits that are misclassified and are "mirrored" improperly in the orbit library (Section 4.1). We correct an issue with the zero-point of the logarithmic potential for the dark matter halo that would otherwise render energy conservation checks ineffective in the code (Section 4.2). We are able to speed up the total run time of a mass model by a factor of two to three by a simple modification to how the point-spread function (PSF) convolution is implemented in the code (Section 4.3). An improvement in setting the intrinsic mass grid used to constrain stellar density profiles is described in Section 4.4. Finally, we illustrate the effects of these changes in the case of NGC 1453 (Section 5).

Three appendices are included as well. Appendix A derives a simple analytic criterion for the existence of long-axis tube orbits within a model. Appendix B outlines a change in the thin orbit finding algorithm that must be made to the TriOS code in order to generate the correct orbital sampling. Finally, Appendix C presents a mock recovery test demonstrating the ability of our revised TriOS code to recover the input mass parameters.

2. Orbit Modeling Background

A summary of the implementation of the Schwarzschild orbit superposition method in the TriOS code is given in Section 4 of Liepold et al. (2020). Here we focus on the topics relevant for subsequent discussions of axisymmetry (Section 3) and code modifications (Section 4).

In this paper, we use a Cartesian coordinate system in which the x-, y-, and z-axes are directed along the intrinsic major, intermediate, and minor axes of the galaxy, respectively. The z-axis is therefore the symmetry axis of an oblate axisymmetric potential, and the x-axis is the symmetry axis of a prolate axisymmetric potential. We focus on oblate axisymmetric systems in this paper, although our discussions can be easily modified for the prolate axisymmetric case.

2.1. Orbit Types in a Triaxial Potential

In a static triaxial gravitational potential, time invariance is the only global continuous symmetry of the Hamiltonian, H. By Noether's theorem, this symmetry gives rise to conservation of energy as the only "classical" integral of motion. This conservation law restricts the allowed phase space for a given orbit from the full six phase space dimensions to a 5D subspace defined by the energy $H = E$. An integral that reduces the allowed phase space dimension in this way is referred to as an isolating integral.

Numerical studies have revealed that orbits in many potentials often conserve two additional "nonclassical" isolating integrals of motion (Schwarzschild 1979), which we refer to as I_2 and I_3. These additional integrals do not typically have simple analytical expressions nor correspond to global symmetries of H. Orbits that conserve three (or more) isolating integrals of motion are referred to as regular. These regular orbits often fall into one of three main orbit types: short-axis tubes, long-axis tubes, and boxes.

Both types of tubes have a fixed sense of rotation. For short-axis tubes, the component of angular momentum along the potential's minor axis, L_z, does not change sign. Similarly, for long-axis tubes, the component of angular momentum along the potential's major axis, L_x, does not change sign. For box orbits, all three components of angular momentum change sign, leaving no fixed sense of rotation. Box orbits also have the property of touching the equipotential surface, $\Phi(x, y, z) = E$, at some point during their trajectory. Intermediate axis tube orbits are typically unstable in triaxial models (Heiligman & Schwarzschild 1979).

A triaxial system generally admits all three of these main orbit types. For oblate axisymmetric systems, the orbital structure is simpler because L_z is an integral of motion, and only short-axis tubes are present. Similarly, for prolate

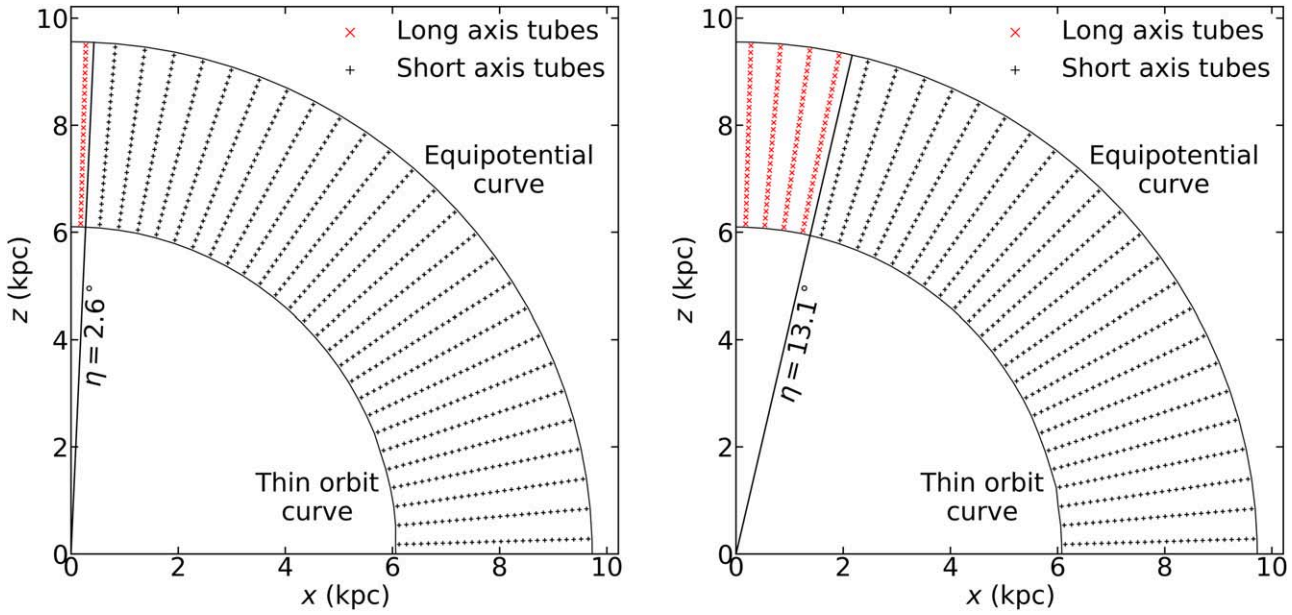

Figure 1. Two examples of the initial orbit locations in the x–z start space. Two nearly axisymmetric models for massive elliptical galaxy NGC 1453 are shown: (left) triaxiality parameter $T = 0.002$, (luminosity weighted) axis ratio $p = 0.9997$, and viewing angles $(\theta, \phi, \psi) = (89°, 45°, 90.°001)$; (right) $T = 0.05$, $p = 0.993$, and $(\theta, \phi, \psi) = (89°, 45°, 90.°026)$. Both models have the best-fit $M_{\rm BH}$, mass-to-light ratio, and dark matter halo from Liepold et al. (2020) and assume the orbital sampling parameters $(N_\Theta, N_R, N_{\rm Dither}) = (9, 9, 3)$ (see Section 3.2). In each panel, one energy is shown, where the energy is chosen such that the potential is dominated by the stellar mass. Each symbol represents the initial location for a single trajectory, which is bundled with adjacent trajectories to form one dithered orbit. The long-axis tubes (red crosses) are all contained within the angle η of the z-axis for both values of T, where η and T are related by Equation (2). In general, more triaxial potentials contain a larger fraction of long-axis tubes in the x–z start space.

axisymmetric systems, L_x is an integral of motion, and only long-axis tubes are present.

2.2. Orbital Sampling and Initialization

The set of initial conditions (referred to as a start space) should sample over all orbit types supported by the potential. Even though regular orbits in a triaxial potential conserve energy plus two additional integrals of motion, the nonclassical integrals of motion, I_2 and I_3, may not be the same quantities for each orbit type (Binney & Spergel 1984; Binney & Tremaine 2008). Thus, for a given energy, each orbit type can be sampled by a 2D start space, but the start spaces for the different orbit types cannot necessarily be combined into a single 2D start space.

Schwarzschild (1993) argued that a 4D space can guarantee that all orbit types of a given energy are sampled, and further suggested that a pair of 2D start spaces is sufficient for sampling phase space in realistic galaxy potentials. The first of these start spaces, the x–z start space, is defined by sampling over a grid of points in the x–z plane, and setting $y = v_x = v_z = 0$ and v_y from $v_y^2 = 2[E - \Phi(x, 0, z)]$ for a given E. For simplicity, v_y is taken to be positive, and a second copy is added to the orbit library with the velocity direction flipped. Two examples of this x–z start space are shown in Figure 1.

Typically, tube orbits will pass through the positive quadrant of the x–z plane perpendicularly at two points, separated by the thin-orbit curve (see Figure 1). Orbits launched along that curve will perpendicularly pass through the plane at a single point, so the curve can be found by iteratively launching orbits at different radii to identify those that pass through the x–z plane in a thin curve (see Appendix B). Each orbit in the x–z start space passes once inside and once outside the thin-orbit radius, so the code avoids double counting by initializing orbits only between the thin-orbit curve and the equipotential where $E = \Phi(r)$, as shown by the crosses in the examples in Figure 1. All three main orbit types pass through this start space.

The second 2D start space proposed by Schwarzschild (1993) is referred to as the stationary start space. In this start space, orbits are started from rest on the equipotential surface and are sampled over a solid angle. Since tube orbits never come to rest, box orbits will be the only main orbit family in this start space. By combining the x–z start space that samples mainly tube orbits with the stationary start space that samples mainly box orbits, Schwarzschild (1993) suggested that any remaining unsampled region of phase space is likely to be small.

The TriOS code is designed for static triaxial potentials that possess reflection symmetry along each of the three principal axes. Under this assumption, any orbital property only needs to be calculated in one octant; it can then be "mirrored" into the other seven octants by symmetry. Taking advantage of this symmetry, the code initializes orbits only in one octant (x, y, $z > 0$) and integrates only these orbits. Seven additional copies of each orbit are then created by simply mirroring along the three axes. The details are described in Section 4.5 of van den Bosch et al. (2008), and the mirroring scheme is given in their Table 2. A key feature to note in their Table 2 is that the exact mirror procedure (i.e., how the signs of the velocity components are flipped in each octant) depends on whether the orbit is a short-axis tube, long-axis tube, or box. The orbits therefore must be classified first.

To classify an orbit, the code determines how the angular momentum components change sign over the course of its integrated trajectory and uses these rules: (1) short-axis tubes, if L_x and L_y flip signs while L_z does not, (2) long-axis tubes, if L_y and L_z flip signs while L_x does not, and (3) box orbits, if all three angular momentum components change signs. The velocities are mirrored in order to maintain the orbit's sense

of rotation. If an orbit does not fall into any of these categories, its velocity is mirrored to have zero angular momentum.

2.3. Orbital Integration

The TriOS code uses the `DOP853` explicit Runga–Kutta integrator with order 8(5,3). The integrator performs adaptive time stepping to ensure that the relative error in the positions and velocities is below a set threshold, typically 10^{-5}. After each orbit is integrated, a relative energy tolerance is used to check energy conservation. If the change in energy exceeds this tolerance (typically set to 10%), it is re-integrated with a smaller integration error threshold.

The default integration time for each orbit is 200 dynamical times, where a dynamical time is set to the period of a closed elliptical orbit of the same energy. To enforce smoothness of the recovered distribution function, the orbital initial conditions can be "dithered" by combining $N^3_{\rm Dither}$ trajectories corresponding to nearby initial conditions. By merging trajectories in this way, each orbit represents a small volume of the start space rather than a single point. This results in smoother orbital properties without a significant memory increase, since only the bundled orbital properties are stored.

After integration, the trajectory of each orbit is interpolated onto a set of points (typically 50,000) that are uniformly spaced in time. These interpolated points are then stored and used for computing orbital properties. Once the orbit libraries are constructed, weights are found for each orbit to reproduce the observed surface brightness distribution, the LOSVDs, and intrinsic 3D mass distribution.

2.4. Viewing Angles, Axis Ratios, and Triaxiality

Three viewing angles (θ, ϕ, ψ) can be used to relate the intrinsic and projected coordinate systems of a triaxial galaxy (Binney 1985). The two angles θ and ϕ describe the orientation of the observer's line of sight with respect to the intrinsic axes of the galaxy. The angle ψ specifies the remaining degree of freedom—rotation of the galaxy around the line of sight. The angle $\psi = 90^\circ$ corresponds to an oblate axisymmetric potential. In the oblate axisymmetric limit, θ is the inclination with $\theta = 90^\circ$ corresponding to edge-on, and ϕ describes rotations about the symmetry axis.

These three viewing angles are related to the intrinsic axis ratios p and q, where $p = b/a$ is the intrinsic intermediate-to-major axis ratio, $q = c/a$ is the intrinsic minor-to-major axis ratio, and *a, b,* and *c* are the lengths of the three principal axes of a triaxial system (with $c \leqslant b \leqslant a$). A third parameter, $u = a'/a$, represents a compression factor due to projection, where a' is the major axis of the projected shape on the sky; $u = 1$ corresponds to the intrinsic major axis lying in the plane of the sky, while $u = p$ corresponds to the intrinsic intermediate axis lying in the plane of the sky. These quantities obey the inequality $0 \leqslant q \leqslant p \leqslant u \leqslant 1$. The relationship between the viewing angles, intrinsic axis ratios, and observed axis ratio is given in Equations (7)–(10) of van den Bosch et al. (2008). In addition, a triaxiality parameter is often used:

$$T = \frac{1 - p^2}{1 - q^2}. \tag{1}$$

This parameter ranges from 0 for oblate axisymmetry to 1 for prolate axisymmetry, with values in between indicating a triaxial shape.

The oblate axisymmetric limit can be achieved by setting either $p = 1$ or $\psi = 90^\circ$, but for numerical reasons, the code does not run when ψ is exactly 90°. As we discussed in Liepold et al. (2020) and elaborate below (Section 3.2), axisymmetry in the code can be achieved only with carefully chosen values of ψ or p.

3. Ingredients for Achieving Axisymmetry

In this section, we discuss a number of steps that need to be taken to generate orbit superposition models in the oblate axisymmetric limit using the TriOS code. It is straightforward to modify these steps for the prolate axisymmetric limit. In Appendix C, we test the modified TriOS code on a mock data set showing that we can accurately recover input parameters.

3.1. Axisymmetrize Short-axis Tube Orbits

As we described in Section 2.2, a triaxial potential exhibits reflection symmetry along each principal axis, allowing the TriOS code to initialize orbits in only one octant of the x–z start space. These orbits are then mirrored via eight-fold reflections about the principal axes into each of the other seven octants. This setup is not meant for axisymmetric systems, in which the orbit library should respect azimuthal symmetry about the symmetry axis.

To enable modeling axisymmetric systems, we have implemented an axisymmetrized version of the orbit library by creating 80 copies of each short-axis tube orbit in the original loop library: 40 copies rotated evenly through an angle 2π about the short axis with velocities rotated to preserve L_z, and another 40 copies generated by flipping the sign of z and v_z in each of the 40 rotations. We choose 40 rotations, as this gives several copies per quadrant, with a comparable density to the start space grid sampling. Once we perform this operation, it is unnecessary to perform the eight-fold reflections in the original code. A similar rotation scheme was tested on mock data with no central supermassive black hole (SMBH) in Hagen et al. (2019).

The net result of our axisymmetrization process is to create a library of short-axis tube orbits in the TriOS code that samples the azimuthal angle uniformly with effectively equal orbital weights. In order for this procedure to be justified, the library should consist solely of short-axis tubes. In the next section, we show how to ensure that no long-axis tubes occur in this library.

3.2. Exclude Long-axis Tube Orbits

In an oblate axisymmetric potential, the long-axis tube orbits become unstable since there is no longer a single preferred long axis. These orbits therefore should not be present in the orbit library.[3]

As we discussed in Section 2.4, the potential is oblate axisymmetric when ψ is set to 90° exactly, and long-axis tubes should be absent in this limit. For numerical reasons, however, the code does not run when ψ is set to 90° within machine accuracy. Prior work using this code for black hole mass measurements in the axisymmetric limit chose either $|\psi - 90^\circ|$ between 0.001° and 0.01° (Walsh et al. 2016; Ahn et al. 2018), or an axis ratio of $p = 0.99$ (Seth et al. 2014; Walsh et al.

[3] Similarly, in the case of a prolate axisymmetric potential, the short-axis tube orbits become unstable and should be absent.

2015, 2017). As we first pointed out in Liepold et al. (2020), some of these values may not have been close enough to the desired axisymmetric values to exclude long-axis tubes. Here we provide a detailed explanation.

We use two examples of the x–z start space in Figure 1 to illustrate how long-axis tube orbits are initialized in the code. As shown in Appendix A, long-axis tube orbits in many realistic triaxial potentials are confined to pass through the x–z start space within an angle η from the z-axis. The angle η depends on the shape of the potential, and we find the relation between η and the triaxiality parameter T (Equation (1)) to be well approximated by

$$\eta = \tan^{-1}\sqrt{\frac{T}{1-T}}. \quad (2)$$

This is demonstrated in Figure 1 where the black line at angle η separates the short-axis tube orbits (black crosses) from the long-axis tube orbits (red crosses). As the potential becomes more oblate axisymmetric ($T = 0.05$ in the right panel versus $T = 0.002$ in the left panel), η becomes smaller, and the area in the x–z start space occupied by long-axis tubes shrinks. To effectively achieve oblate axisymmetry, η needs to be small enough so that no orbits are sampled within an angle of η of the positive z-axis. Two additional mass models with higher triaxiality, ($T = 0.25$ and 0.75) are shown in Appendix A. Equation (2) again provides an excellent approximation for the angle demarcating the long-axis and short-axis tube orbits in the x–z start space.

Whether orbits are sampled within the angle η on the x–z plane depends on the input parameters. For a given energy, the code starts the orbits on a grid of N_R radii between the inner and outer thin-orbit radii and N_Θ angles between 0° and 90° in the positive quadrant on the x–z plane (van den Bosch et al. 2008). The code further allows for dithering, where $N^3_{\rm Dither}$ nearby initial conditions, adjacent in (E, R, Θ), are bundled together to improve the phase space sampling. Orbits are therefore sampled at a total of $N_\Theta \times N_{\rm Dither}$ angles, where the first angle from the z-axis is chosen to start at half of the grid spacing (i.e., at an angle of $(\pi/2)/(2N_\Theta N_{\rm Dither})$ from the z-axis). The criterion to satisfy oblate axisymmetry is therefore

$$\frac{1}{2N_\Theta N_{\rm Dither}}\frac{\pi}{2} \gtrsim \eta. \quad (3)$$

The two examples of NGC 1453 shown in Figure 1 have $N_\Theta = N_R = 9$, $N_{\rm Dither} = 3$, and 27×27 orbits initialized in the x–z start space. The orbits closest to the z-axis are therefore at an angle of $\approx 1\overset{\circ}{.}67$ away. These orbits lie within the demarcation angle η of Equation (2) for either model in Figure 1: $\eta = 2\overset{\circ}{.}56$ for $T = 0.002$ (left) and $\eta = 12\overset{\circ}{.}9$ for $T = 0.05$ (right). Both models therefore violate Equation (3) and contain long-axis tubes. This provides the physical explanation for our statement in Liepold et al. (2020) that even $|\psi - 90°|$ as small as 0.001 (left panel) is not sufficiently close to 90° to achieve axisymmetry in our models.

To extend the discussion beyond the two specific mass models shown in Figure 1, we illustrate in Figure 2 the relation between T and ψ for nearly axisymmetric models of NGC 1453 (top panel), and the corresponding fraction of long-axis tubes that are initialized in the x–z plane (bottom panel). The inclination angle θ is assumed to be 89° here, and the shaded band indicates the additional dependence of T on ϕ. Figure 2

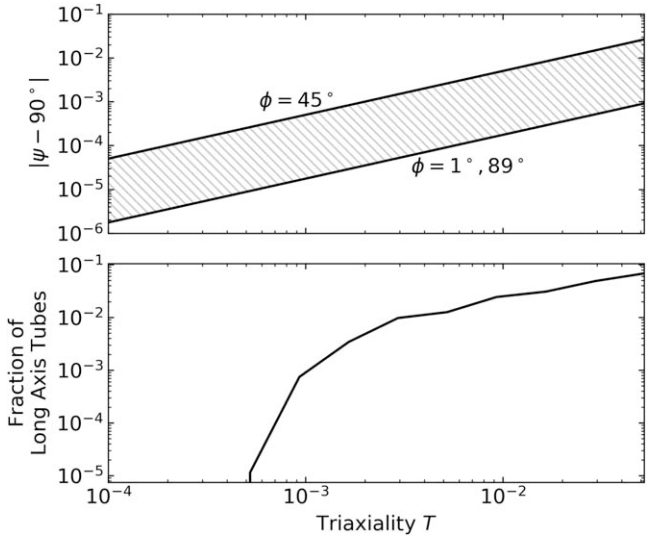

Figure 2. (Top panel) Relationship between the viewing angle ψ and the triaxiality of the deprojected stellar density. Exact oblate axisymmetry has $T = 0$ and $\psi = 90\overset{\circ}{.}0$. The other viewing angle θ is taken to be 89°, and ϕ is varied from 1° to 89°. (Bottom panel) Fraction of long-axis tube orbits in the x–z start space as a function of the triaxiality of the stellar density near the oblate axisymmetric limit. The same mass model and orbital sampling parameters for NGC 1453 shown in Figure 1 is assumed here. In this example, long-axis tube orbits begin to appear when T is as small as $\sim 5 \times 10^{-4}$, or $|\psi - 90°|$ as small as $\sim 9 \times 10^{-6}$, and the fraction of these orbits increases monotonically as the potential becomes more triaxial, reaching ~6% at $T = 0.05$.

shows that $T \lesssim 5 \times 10^{-4}$ is needed to exclude long-axis tube orbits in this case. The corresponding requirement on ψ is $|\psi - 90°| \lesssim 8.7 \times 10^{-6}$ for $\phi \sim 1°, 89°$ and $|\psi - 90°| \lesssim 2.5 \times 10^{-4}$ for $\phi \sim 45°$. We advocated $|\psi - 90°| = 10^{-9}$ in Liepold et al. (2020), which safely excluded all long-axis tube orbits.

Earlier work using the code in the near axisymmetric limit does not typically satisfy the criterion in Equation (3). For M59-UCD3, Ahn et al. (2018) used (θ, ϕ, ψ) = (85°, −49°.99, 89°.99), which we find to correspond to $T = 0.004$ and $\eta = 3\overset{\circ}{.}64$. The orbital sampling parameters were not explicitly given for the runs using the triaxial code. Assuming the same parameters used in their runs with the axisymmetric orbit code ($N_\Theta = 8$, $N_{\rm Dither} = 6$), we find that the innermost ray would be at an angle of $0\overset{\circ}{.}94$ from the z-axis, which is well inside $\eta = 3\overset{\circ}{.}64$, and therefore violates the criterion in Equation (3).

For M60-UCD1 (Seth et al. 2014), NGC 1271 (Walsh et al. 2015), and Mrk 1216 (Walsh et al. 2017), each paper quoted an axis ratio of $p = 0.99$. The minimum possible triaxiality with this value of p is $T = 1 - p^2 = 0.0199$ (in the unrealistic limit of a razor-thin disk with $q = 0$), leading to a minimum η of $8\overset{\circ}{.}1$. For NGC 1271 and Mrk 1216, $N_{\rm Dither} = 5$ was used, while N_Θ was set to 8 and 9, respectively. Thus, orbits were sampled starting at $1\overset{\circ}{.}125$ and 1° away from the z-axis, indicating that neither satisfies the criterion in Equation (3). For M60-UCD1, not enough information is given about the orbital sampling to determine whether the criterion is satisfied. However, for typical orbital sampling parameters quoted above, the criterion in Equation (3) would not be satisfied.

The modeling of the NGC 1277 black hole used $N_\Theta = 9$ and $N_{\rm Dither} = 5$ (Walsh et al. 2016); the innermost ray of initial orbits therefore lies at 1° from the positive z-axis. The complete shape information was not given in the paper, but private communication indicated that (θ,ϕ,ψ) = ($75\overset{\circ}{.}3$,$71\overset{\circ}{.}6$,$90\overset{\circ}{.}001$)

was used. We find this set of viewing angles to correspond to $T = 0.0002$ and $\eta = 0.^{\circ}85$, narrowly satisfying the criterion in Equation (3).

We note that the presence of the long-axis tube orbits in the orbit library does not necessarily imply that they receive significant weights after fitting to observational constraints for a given galaxy. Direct tests would need to be performed for each galaxy to assess the impact of these orbits on previous work.

3.3. Exclude Box Orbits

As we discussed in Section 2.1, all orbits in the (oblate) axisymmetric limit conserve L_z. Box orbits in this limit have $L_z = 0$ and therefore have similar properties to the tube orbits with small L_z. In this case, as long as angular momentum is sufficiently sampled by the tube orbits, there is no need to include box orbits explicitly.

The TriOS code devotes an entire library to box orbits and initializes them in the stationary start space (Section 2.2). One can modify the code to exclude this library when needed. We use a simpler approach without changing the code itself: we skip running the orbital integration routine `orblib_f.f90` for the stationary start space, and replace the box library with a copy of the x–z library in the input file for the weight-finding routine `triaxnnls.f90`. These modifications typically reduce the total computation time of the original code by more than half.

While box orbits are unnecessary in the axisymmetric limit, they also should be harmless and not affect the results if included. As a test, we have run our revised code including the box library for comparison. Since the box orbits launched at different azimuthal angles are allowed to have different weights in the triaxial code, we have to impose an additional constraint of equal weights to enforce axisymmetry in the box library. Once these weights are forced to be equal, we indeed find similar results as the case when the box library is excluded altogether. The case where the box library weights are free to differ between azimuthal angles is discussed in Section 5. To reduce computational cost, we recommend excluding the stationary start space for axisymmetric models.

For a triaxial potential, we note that box orbits can also occur in the x–z start space (e.g., Figure 1 of Schwarzschild 1993). However, the region in the x–z start space that would generate box orbits shrinks as the potential becomes increasingly axisymmetric. When exact axisymmetry is reached, only the orbits that begin exactly on the equipotential surface in the x–z start space have $L_z = 0$ (since they have zero initial velocities) and are box orbits. The TriOS code does not sample orbits lying exactly on the equipotential curve in the x–z start space, so the number of box orbits will shrink to zero as axisymmetry is approached. In other orbit-based codes that assume axisymmetry from the start, the $L_z = 0$ orbits also are not usually sampled, as they are presumed to be represented by the tube orbits with small but nonzero L_z (e.g., Cretton et al. 1999; Thomas et al. 2004).

4. Additional Code Fixes and Improvements

We have made several modifications in the TriOS code in addition to those described in Section 3. These modifications include corrections, improvements, and speedups that are general to the code regardless of the issue of axisymmetry. We describe these changes in this section.

4.1. Correct Orbit Misclassifications

As we described in Section 2.2, the TriOS code assumes the triaxial potential to possess reflection symmetry along each of the three principal axes and integrates only orbits that are initialized in one octant of space to save computation time. It then uses an eight-fold reflection scheme to generate seven more copies of each orbit. How the orbits are "mirrored" depends on whether the orbit is classified as a short-axis tube, long-axis tube, or box orbit.

We have discovered that the mirroring scheme in the original code misclassifies a subset of orbits for which the angular momenta vary on timescales slower than the integration time. We find this to happen in at least two situations. First, in nearly oblate axisymmetric models, many box orbits in the stationary start space tend to be misclassified as short-axis tubes due to the near conservation of L_z. Because L_z varies slowly, it may not change sign throughout the integration time. However, these orbits have very low angular momentum, so it is unlikely that mirroring these orbits to preserve L_z would cause significant issues in the models themselves.

The second situation occurs in regions of space where the potential is nearly spherical, e.g., deep within the sphere of influence (SOI) of an SMBH, or in the outer part of a galaxy where a (spherical) dark matter dominates the potential. Some orbits in these regions follow quasi-planar rosettas or Keplerian-like ellipses with nearly constant angular momentum vectors, consistent with prior studies of orbits near a central point mass (Sridhar & Touma 1999, 2016; Valluri et al. 2016). For the subset of orbits with precession time longer than the integration time, no component of their angular momentum changes sign over the entire integrated trajectories. These orbits therefore do not fall into any of the categories listed above and are mirrored incorrectly to have no net angular momentum.

These quasi-planar orbits will not be significant in most Schwarzschild models, as they are only present at extreme radii. We checked this in our models of NGC 1453, with the properly axisymmetrized code as described in Section 3 using the lowest four Gauss–Hermite kinematic moments as constraints. In this model, we find that ∼10% of the total weight after orbital weight minimization is assigned to orbits that would have been quasi-planar in the original version of the code (∼10% of the mass within the Mitchell apertures and ∼2% of the mass within the Gemini Multi-Object Spectrograph apertures). These relatively low percentages suggest a minimal effect on the model for NGC 1453.

We expect the issues with orbital integration time and misclassification to be more severe for galaxies with data that resolve well within the black hole's SOI, or well beyond the stellar half-light radius, e.g., M87 and the Milky Way black hole. The effect is also likely to be more significant if the galaxy has a net rotation at these radii.

We find a further issue with orbit classification in the orbital composition information outputted in the file `intrinsic_moments.out`. This file reports the mass fraction of box orbits for each bin in the intrinsic spatial grid described in Section 4.4. In this case, however, all orbits that are neither long-axis tubes or short-axis tubes are grouped together as box orbits. Since this includes the quasi-planar orbits discussed above, the reported fraction of true box orbits may be overestimated.

In our revised code for axisymmetric systems, these orbit misclassification issues are not present because we manually

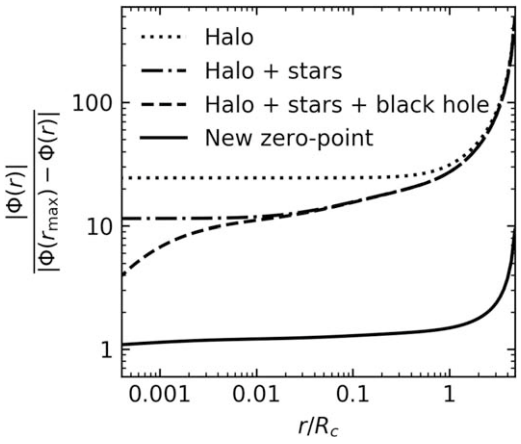

Figure 3. Illustration of the issue with setting the zero-point of the logarithmic potential to $\Phi_0 = 0$ in Equation (4), as is assumed in the original code. As an example, we use the best-fit mass model for NGC 1453 in Liepold et al. (2020) with a logarithmic dark matter halo of $R_c = 15$ kpc and $V_c = 633\text{km s}^{-1}$. The ratio of the potential energy to the maximum kinetic energy is plotted for this halo (dotted), halo plus stars (dotted–dashed), and all three mass components (dashed). When this ratio is much larger than one, as is shown for a large range of radii, even large errors in the kinetic energy would have little effect on the total energy. Energy conservation is therefore effectively not enforced in the original code for a logarithmic potential. The solid line shows the same ratio with all three mass components included, but with the halo zero-point set according to Equation (6).

assign all orbits as short-axis tubes and exclude all other orbit types. We will discuss further these quasi-planar orbits in triaxial systems in Section 6.

4.2. Fix Zero-point Issues with the Logarithmic Halo

A logarithmic potential has often been used to approximate the dark matter halo in prior orbit modeling work. The spherical version of a logarithmic halo is given by

$$\Phi(r) = \frac{1}{2} V_c^2 \ln(R_c^2 + r^2) + \Phi_0, \tag{4}$$

where R_c is the core radius, and V_c is the circular velocity at large r:

$$V_c(r) = \frac{V_c\, r}{\sqrt{R_c^2 + r^2}}. \tag{5}$$

The zero-point Φ_0 can in principle be chosen arbitrarily; the original code set $\Phi_0 = 0$. In practice, we find the choice of $\Phi_0 = 0$ and the use of physical units such as kilometers for all distances to create numerical problems. The cause is simple: unlike other commonly used dark matter potentials such as Hernquist (1990) and Navarro et al. (1996) that are negative at all locations and approach zero at large r, the logarithmic potential with $\Phi_0 = 0$ is positive everywhere and grows unbounded at large r. Thus, for the other potentials, $|\Phi(r)|$ can be interpreted as the local maximum kinetic energy for a bound orbit, but the orbital binding energy is infinite in the logarithmic potential. Furthermore, with the choice of $\Phi_0 = 0$, $|\Phi(r)|$ is much larger than the kinetic energy for all orbits in a logarithmic halo. This is because the central potential energy value, $\Phi(0) = V_c^2 \ln(R_c)$, is much larger than the maximum possible kinetic energy sampled by the orbits, which is $\Phi(r_{max}) - \Phi(0)$, where r_{max} is the largest equipotential radius of any orbit in a model.

To illustrate this point, we plot the ratio of $|\Phi(r)|$ and $|\Phi(r_{max}) - \Phi(0)|$ for the best-fit logarithmic dark matter halo of NGC 1453 (Liepold et al. 2020) in Figure 3 (dotted curve). Additional contributions to the potential from the stars and black hole reduce the value of the potential energy and help lower this ratio (dotted–dashed and dashed solid curves), but the ratio is well above unity for all relevant radii in all cases.

An unintended consequence of this large central offset is that even a ∼100% change in the kinetic energy would contribute to only a tiny fraction of the total energy and would be difficult to detect. The energy conservation checks in the code are therefore effectively not performed for most orbits. While these numbers are worrying, we did not find the choice of $\Phi_0 = 0$ to affect significantly the best-fit mass parameters of NGC 1453 in Liepold et al. (2020). The reason for this particular case is that the orbital integrator happened to be accurate enough to satisfy the energy conservation tolerance (set to the default 10%) even when this conservation criterion was unchecked. There is, however, no guarantee that this would be true for other galaxies or for parameters outside the ranges that we had explored.

To ensure that energy conservation is checked in the code for the logarithmic potential, we choose a different zero-point

$$\Phi_0 = -\Phi(r = 2r_{max}), \tag{6}$$

so that $\Phi(r)$ is negative for the entire allowed radial range of the orbits and approaches zero outside the largest equipotential radius r_{max}. The resulting ratio of $|\Phi(r)|$ to $|\Phi(r_{max}) - \Phi(0)|$ for the best-fit model of NGC 1453 is shown by the solid line in Figure 3.

Our choice of Φ_0 in Equation (6) also removes another issue that we have encountered with the original code: the orbit start space was sometimes not calculated correctly for mass models in which the black hole is either absent or has small mass compared to the stellar component and the logarithmic halo. As discussed in Section 2.2 and shown in Figure 1, the x–z start space of Schwarzschild (1993) requires finding equipotential curves in the x–z plane. The code locates it by finding the equipotential radius for each of a series of angles in the plane. For each angle, the equipotential radius is found via bisection with a relative tolerance that is typically taken between 10^{-7} and 10^{-5}. For $\Phi_0 = 0$, this tolerance again is not enforcing the intended accuracy level due to the large central value of Φ. For NGC 1453, this issue exists only for a few central equipotential radii, and thus it did not have a significant impact on our results.

4.3. Speed Up Point-spread Function Implementation

The PSF of the relevant observations needs to be incorporated into a mass model before the model is fitted to data to determine the orbital weights. The TriOS code approximates the effect of the PSF by perturbing each trajectory at every stored time step with a pair of δx and δy randomly drawn from the PSF, which is assumed to be a single or multiple Gaussian functions. This scheme involves a large number of operations since an orbit is typically stored at 50,000 points along the trajectory (see Section 2.3), and up to $\sim 10^6$ orbits can be used to represent a single mass model.

The code generates each orbit perturbation by drawing two independent numbers, k_x and k_y, from a uniform distribution over the interval $(-1, 1)$ repeatedly until a pair with $k \equiv |\boldsymbol{k}| < 1$ is found. The perturbations $\delta x = \frac{k_x}{k}\sqrt{-2\ln(k^2)}$ and

$\delta y = \frac{k_y}{k}\sqrt{-2\ln(k^2)}$ are then normally distributed. This large number of operations is not easily vectorized and is computed sequentially.

We are able to speed up this process significantly using instead the Box–Muller transform, which is easily vectorized. In this scheme, we draw a pair of independent numbers A and B from the uniform distribution over (0, 1) and then construct the normal distribution with $\delta x = \sqrt{-2\ln A}\cos(2\pi B)$ and $\delta y = \sqrt{-2\ln A}\sin(2\pi B)$. We have tested that the resulting distributions of displacements are consistent with analytical PSFs to within the counting error from the finite number of time steps, and the consistency increases as expected when the number of time steps increases.

To benchmark the amount of speedup gained by our scheme, we note that PSF convolution is one of several operations performed in the orbit library construction subroutine `orblib_f.f90` in the code. This subroutine first integrates the orbits and generates the necessary reflected or rotated copies of the orbits about the symmetry axes (see Section 3.1). It then computes each orbit's contribution to the 3D mass grid and projects each orbit onto the sky plane. The projected trajectories are then perturbed according to the PSF as described above. Finally, the subroutine determines each orbit's contribution to each observed kinematic aperture on the sky and stores the associated LOSVDs. The tasks performed in this subroutine consume the bulk (>90%) of the total run time of the code (for one mass model); much of the remaining time is spent on performing minimizations to find optimal orbital weights.

To our surprise, our timing analysis of the various tasks executed in this subroutine (using $N_{\rm Dither} = 5$ and NGC 1453 as a test case) shows that the PSF portion of the code (before implementing orbit axisymmetrization in Section 3.1) takes up ∼55% of the run time, while the orbital integration itself only contributes ∼20%, and sky projections contribute the remaining ∼25%. When we switch to the vectorized Box–Muller transform, the computation time for the PSF step becomes negligible. We are therefore able to reduce the total run time of the code by a factor of ∼2 in this test.

The speedup is even more dramatic in our axisymmetrized version when the orbits are copied azimuthally (Section 3.1). In this case, 80 (instead of eight) copies of each orbit are projected onto the sky and perturbed by the PSF. We find ∼70% of the run time is spent on the PSF portion with the original scheme, while our new scheme reduces the run time by a factor of ∼3.

4.4. Improve Intrinsic 3D Mass Grid

The TriOS code uses an intrinsic 3D spatial grid to constrain the stellar component in a model to reproduce the 3D stellar density profile deprojected from the photometry of a galaxy. The code calculates the mass contributed by each orbit as it passes through a spatial bin and records this information during the stage of orbit library construction. At the subsequent stage of orbital weight optimization, the superposition of the orbits is required to match the input mass profile within a pre-specified precision (typically 1%) in each bin.

In each octant of this 3D spatial grid, the code uses azimuthal and polar bins for the two angles, each linearly spaced between 0° and 90°. The radial bins are logarithmically spaced between $r_{\rm min}$ and $r_{\rm max}/2$, where $r_{\rm min}$ and $r_{\rm max}$ are the innermost and outermost equipotential radii used to determine the orbital energies sampled in the model. The innermost bin is then extended down to $r = 0$, and the outermost is extended out to $100 r_{\rm max}$.

For the outer boundary of the innermost mass bin, we find it preferable not to base the value on $r_{\rm min}$, which is used for a different purpose of specifying the innermost equipotential radius for sampling orbital energy. Instead, we modify the code to make it an independent parameter, which we set to be of similar scale as the PSF of the photometric data since these are the data used to constrain the deprojected 3D mass density. To ensure that sufficient orbits are used to represent the innermost mass bins, we recommend that $r_{\rm min}$ be set to be smaller than the outer boundary of the innermost mass bin. In the case of NGC 1453, we set the outer boundary of the innermost mass bin to be 0.03″ and set $r_{\rm min}$ to 0.01″.

For similar reasons, we allow the outermost mass bin's edges to also be set independently from the outermost equipotential radius, $r_{\rm max}$. The remaining bin boundaries are then logarithmically sampled between the outer boundary of the innermost bin and the inner boundary of the outermost bin.

5. A Case Study: NGC 1453

We use the massive elliptical galaxy NGC 1453 reported in Liepold et al. (2020) to illustrate the effects of the modifications described thus far. In Liepold et al. (2020), we demonstrated that using more than four Gauss–Hermite moments was essential for obtaining robust constraints on the model LOSVDs. Below we examine the effects in both the four-moment and 12-moment cases, with the latter being our chosen configuration. We stress that the four-moment case is included here only for comparison purposes since this is the typical configuration used in the literature. We have found the four-moment case to lead to unconstrained higher moments and spurious features in the LOSVDs for NGC 1453 (Figures 10 and 11 of Liepold et al. 2020); the resulting χ^2 in this case should therefore not be trusted.

5.1. Fitting up to h_4

We begin with the case labeled "up to h_4" and "original Leiden version" in Figure 12 of Liepold et al. (2020). This case is run with the original code, $N_{\rm Dither} = 3$, $N_\Theta = 9$, and the viewing angles $(\theta, \phi, \psi) = (89°, 45°, 90.°001)$, corresponding to a nearly oblate axisymmetric potential with a triaxiality parameter of $T = 0.002$. As we discussed in Section 3.2, these parameters are chosen to resemble those used in earlier studies, and the models include both the x–z and stationary start spaces and contain all three major types of orbits: short-axis tubes, long-axis tubes, and box orbits. The left panel of Figure 1 illustrates the starting locations of both short- and long-axis tube orbits in the x–z start space for one energy in this configuration.

The 1D χ^2 as a function of $M_{\rm BH}$ (marginalized over the mass-to-light ratio) is shown in the left panel of Figure 4 (red dotted curve). As first shown in Liepold et al. (2020), the favored model in this case contains no black hole. The χ^2 minimum at $M_{\rm BH} = 0$ here resembles the finding for the dwarf galaxy M59-UCD3 by Ahn et al. (2018), which also used four Gauss–Hermite moments as constraints and a set of viewing angles with a similar deviation from axisymmetry.

Applying the code changes described in Section 4 results in minor changes in the χ^2 contour for NGC 1453 (purple short-dashed curve in Figure 4), but the $M_{\rm BH} = 0$ minimum remains.

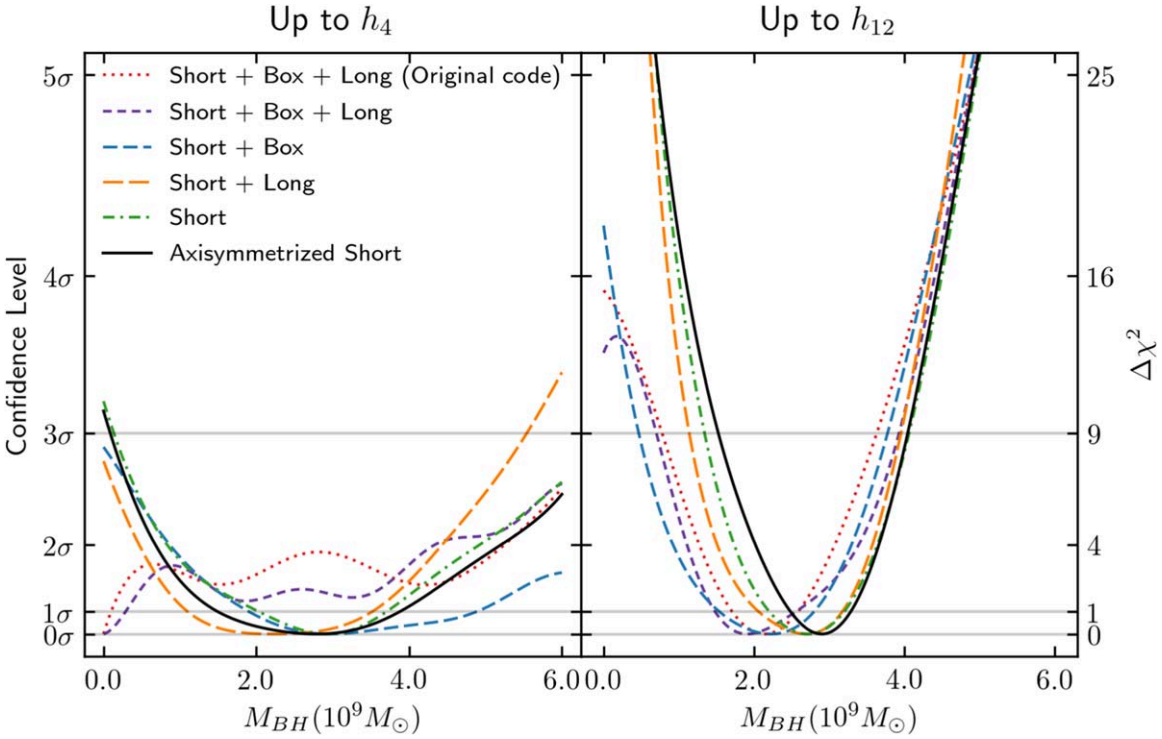

Figure 4. Illustration of the changing $M_{\rm BH}$ constraints in NGC 1453 as the orbit model goes through the step-by-step axisymmetrization procedure described in Sections 3 and 4. The starting case (red dotted) uses the original code with typical (near) axisymmetric parameters assumed in the literature ($\psi = 90.^\circ001$; see Section 5.1 for details). The end case (black solid) uses our final axisymmetrized code including all changes from Sections 3 and 4. The four intermediate curves have all of the code fixes described in Section 4, but have different combinations of orbit types according to Sections 3.2 and 3.3. The left panel is for models with orbital weights chosen by fitting to the first four Gauss–Hermite moments of the LOSVDs determined from kinematic data, as is typical in the literature. The right panel uses 12 moments as constraints and shows tighter constraints on $M_{\rm BH}$, as is reported in Liepold et al. (2020). The 1D χ^2 in $M_{\rm BH}$ is obtained by marginalizing over the stellar mass-to-light ratio using a smoothed 2D χ^2 landscape generated by Gaussian process regression with a squared-exponential covariance function (Pedregosa et al. 2011). The dark matter halo is fixed to the best-fit logarithmic halo in Liepold et al. (2020).

In the next step, we exclude the box orbits and long-axis tube orbits as described in Section 3. The box orbits are eliminated by the simple procedure in Section 3.3. To remove the long-axis tube orbits, we choose a galaxy shape that is sufficiently axisymmetric, as discussed in Section 3.2. For NGC 1453, we simply change ψ from $90.^\circ001$ to $(90 + 10^{-9})^\circ$, as was done in Liepold et al. (2020). This new value is far enough from 90.0° to avoid numerical issues in the code but is close enough to 90.0° so that all of our orbits lie outside the long-axis tube region in the x–z start space shown in Figure 1.

The effect of excluding these orbits on the best-fit parameter values for NGC 1453 is significant. The preferred $M_{\rm BH}$ is changed from zero to $2.8 \times 10^9 M_\odot$ (green dotted–dashed curve in Figure 4(a)). Before their removal, box orbits generally accounted for less than 10%–35% of total mass, while long-axis tube orbits accounted for less than 2%. Removing box orbits (orange long-dashed curve in Figure 4(a)) has a significant effect on $M_{\rm BH}$ because box orbits starting at different azimuthal angles are not forced to have equal weights in the original code (Section 3.3). Removing the long-axis tubes (blue dashed curve in Figure 4(a)) has a significant impact likely due to their ability to fit minor-axis rotation in triaxial potentials.

In addition to excluding the box and long-axis tubes, we describe in Section 3.1 the need to enforce axisymmetry in the code by generating many azimuthally rotated copies of each short-axis tube in the x–z start space. For NGC 1453, we find that the main effect on the χ^2 contour of this axisymmetrization procedure is to widen the minimum (black solid curve in Figure 4(a)), as a broader range of orbital weights are able to fit the mass constraint for each mass model.

The results presented thus far with the original version of the code all assumed a viewing angle of $\phi = 45^\circ$. When the model galaxy is perfectly axisymmetric, this angle is irrelevant, and the resulting χ^2 landscape should be independent of ϕ. As a test, we have repeated the run with the original code (using four Gauss–Hermite moments) with two other values of ϕ (15° and 75°) while keeping all other parameters fixed. The resulting χ^2 as a function of $M_{\rm BH}$ for the three values of ϕ is shown in Figure 5. The dependence on ϕ indicates that the mass models are indeed not consistent with axisymmetry. All three values of ϕ exhibit the same preference for $M_{\rm BH} = 0$.

5.2. *Fitting up to* h_{12}

We now examine models in which the orbital weights are constrained to fit the first 12 Gauss–Hermite moments of the observed LOSVDs for NGC 1453. The first eight moments are measured from spectroscopic observations, while the 9th–12th moments are set to zero with an error bar based on the lower moments, as described in detail in Liepold et al. (2020). Even without any of the modifications described in this paper, Liepold et al. (2020) showed that the original code performed better when 12, rather than four, moments were used as constraints. The right panel of Figure 12 in Liepold et al. (2020) illustrated how the best-fit black hole mass moved from $M_{\rm BH} = 0$ for four moments (green curve) to $M_{\rm BH} = 2.2 \times 10^9 M_\odot$ for 12 moments (black curve). The result from the original code, however, was highly dependent on the

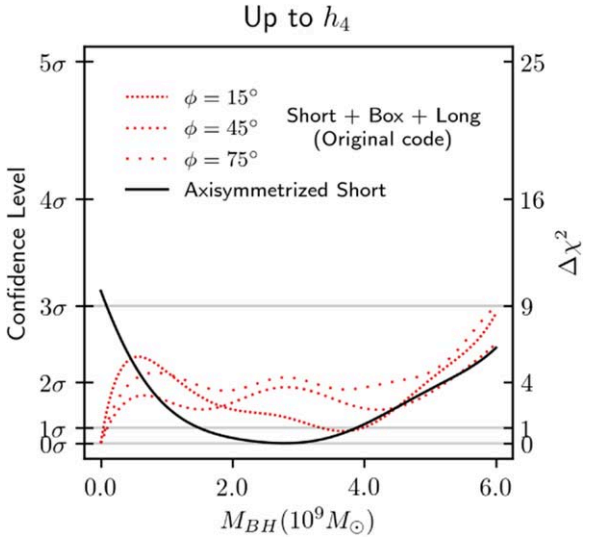

Figure 5. The same as the left panel of Figure 4 but showing the azimuthal dependence of the original code when ψ is chosen to be 90.°001, and all three main orbit types are included (red curves). Our final axisymmetrized code does not depend on ϕ and obeys azimuthal symmetry.

number of input moments and showed no convergence even at 12 moments. By contrast, after the orbit and code modifications were implemented, the main effect of increasing the constraining kinematic moments was to tighten the error bars while leaving the best-fit values largely unchanged (left panel of Figure 12 in Liepold et al. 2020).

Here we examine the progression of changes after each of the key modifications described in Sections 3 and 4 is implemented, all for the case of using 12 moments as constraints. The right panel of Figure 4 shows that implementing the code fixes described in Section 4 (purple dotted–dotted–dashed curve) and removing long-axis tubes (blue dotted–dashed curve) move the best-fit $M_{\rm BH}$ by $\sim$10% in comparison to $M_{\rm BH} \sim 2.2 \times 10^9 M_\odot$ from the original code (red dotted curve). Removing the box orbits increases $M_{\rm BH}$ to $\sim 2.9 \times 10^9\, M_\odot$ (orange dotted–dashed–dashed and green dashed curve). The subsequent axisymmetrization of short-axis tubes (Section 3.1) has essentially no effect (black solid curve).

To ensure that the number of orbits included in the modeling is sufficient, we tested the effect of increasing the number of orbits. We increased the density of energy sampling by a factor of four, from 40 energy values to 160 over the same range. With four times the number of orbits, the best-fit $M_{\rm BH}$ changed by less than 3%, and the 1σ error changed by less than 10%, demonstrating that our results do not depend on the exact number of orbits used.

6. Conclusion

We have presented a revised version of the triaxial orbit superposition code by van den Bosch et al. (2008), which we refer to as the TriOS code, which is capable of properly modeling axisymmetric systems. The original code was designed for triaxial systems with (discrete) reflection symmetry along each of the three principal axes. The setup was not capable of modeling exactly axisymmetric systems in which the orbit library should respect (continuous) azimuthal symmetry about the symmetry axis.

We have implemented two main changes needed for modeling axisymmetric systems within the triaxial code: excluding all orbit types that are not allowed in an axisymmetric model, and enforcing axisymmetry among the allowed orbits. In the case of oblate axisymmetry, our recipe involves (1) axisymmetrizing the short-axis tube orbits by creating multiple copies of the orbits rotated about the symmetry axis (Section 3.1), (2) setting the viewing angle ψ to be sufficiently close to 90° to allow no long-axis tube orbits (Section 3.2), and (3) excluding the stationary start space used to generate box orbits (Section 3.3).

We have made further improvements and corrections to the code in general. We discussed an issue with slowly precessing quasi-planar orbits that are misclassified and are mirrored improperly in the orbit library (Section 4.1). We also corrected a problem with the logarithmic halo implementation that prevented checking energy conservation of the integrated orbits (Section 4.2). We achieved a factor of two to three speedup in the run time of the code by adopting a different algorithm for modeling PSF convolution (Section 4.3). Finally, we allowed the orbital sampling and mass constraints to be set independently (Section 4.4).

For NGC 1453, we found the shape of the χ^2 contours for $M_{\rm BH}$ to vary significantly as we went through the step-by-step axisymmetrization procedure described in this paper (Figure 4). As we described in Liepold et al. (2020), the orbit models favored no black holes when we used the original code with typical (near) axisymmetric parameters in the literature and four Gauss–Hermite moments to constrain the stellar LOSVDs. In contrast, we obtained a well constrained nonzero $M_{\rm BH}$ using our final axisymmetrized code including all of the changes described in Sections 3 and 4.

One issue that warrants further investigation in triaxial models is the equilibrium behavior of quasi-planar orbits in regions where the potential is nearly spherical, e.g., well within an SMBH's SOI, or far outside the galaxy's effective radius in a spherical dark matter halo. As we discussed in Section 4.1, the subset of quasi-planar orbits with precession times longer than the integration time has a nearly constant $\boldsymbol{L}$ and is misclassified and mirrored incorrectly in the original code. Furthermore, the integration time for these orbits is not long enough to fill the allowed volume of phase space. For axisymmetric systems, we resolve these issues in our revised code described in this paper by including only short-axis tubes and enforcing axisymmetry in the orbits, while preserving L_z.

We also expect the severity of the orbital integration issue to vary from system to system: the better an SMBH's SOI is resolved by the available kinematic data, the more care is needed to test orbital integration time because quasi-planar orbits occupy a large fraction of the orbit library, and more orbits are deeper in the SMBH's potential and hence have longer precession times. For the NGC 1453 SMBH studied in Liepold et al. (2020) and here, since our kinematic data do not reach deep inside the SOI, orbits in our mass models with precession times exceeding 200 dynamical times account for less than 4% of luminosity within the central arcsecond. The integration issue (and the resulting misclassification) therefore does not significantly impact our results, as is evidenced by the similarity between the solid black and green dashed curves in Figure 4. We expect a different situation for better resolved systems such as the M87 and Milky Way SMBHs.

In future work, a straightforward solution to ensure that quasi-planar orbits are representative of their equilibrium distributions is to extend the default integration time of 200 dynamical times in the code. Our preliminary tests suggest that integrating the orbits up to ∼10 times longer is computationally feasible, but this may still be insufficient for the orbits closest to the SMBH and in the outermost part of the galaxy where the precession times are slowest. A more reliable treatment of these orbits would be needed.

We thank Jonelle Walsh for useful discussions and a critical reading of the manuscript, and Karl Gebhardt for discussions about mock tests. We also thank the anonymous referee for their helpful comments and suggestions. M.E.Q. acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), PGSD3-517040-2018. C.-P.M. acknowledges support from NSF AST-1817100, HST GO-15265, HST AR-14573, the Heising-Simons Foundation, the Miller Institute for Basic Research in Science, and the Aspen Center for Physics, which is supported by NSF grant PHY-1607611. This work used the Extreme Science and Engineering Discovery Environment at the San Diego Supercomputing Center through allocation AST180041, which is supported by NSF grant ACI-1548562.

Appendix A
Criterion for Existence of Long-axis Tubes

We use Stäckel potentials to gain insights into the existence of long-axis tubes. A potential is said to be in Stäckel form if it can be written as:

$$V(\lambda, \mu, \nu) = -\frac{F(\lambda)}{(\lambda-\mu)(\lambda-\nu)} - \frac{F(\mu)}{(\mu-\nu)(\mu-\lambda)} - \frac{F(\nu)}{(\nu-\lambda)(\nu-\mu)}, \tag{A1}$$

for some function $F(\tau)$ where (λ, μ, ν) are ellipsoidal coordinates defined as the roots of τ in the equation

$$\frac{x^2}{\tau+\alpha} + \frac{y^2}{\tau+\beta} + \frac{z^2}{\tau+\gamma} = 1, \tag{A2}$$

such that $-\gamma \leqslant \nu \leqslant -\beta \leqslant \mu \leqslant -\alpha \leqslant \lambda$. Here, (α, β, γ) are constants that define the coordinate system. Such a potential is said to be separable in these coordinates. When a density corresponding to a Stäckel potential is projected in any direction to give a 2D surface density, it will have no isophotal twists (Franx 1988). Thus, we can use the viewing angles (θ, ϕ, ψ) of Binney (1985) to define the relationship between the primary axes of the projected and intrinsic densities. This set of viewing angles imposes a constraint on the allowed values of (α, β, γ) given by:

$$\frac{\sqrt{\beta-\alpha}}{\sqrt{\gamma-\beta}} = \sqrt{\frac{\sin^2\theta}{\cot 2\psi \sin 2\phi \cos\theta + \cos^2\phi(\cos^2\theta+1) - 1}}. \tag{A3}$$

This expression follows from Equation (B9) of Franx (1988). Orbital structure in Stäckel potentials has been well studied (de Zeeuw 1985). This structure is what motivated the x–z start space described in Schwarzschild (1993). Long-axis tube orbits pass through the x–z start space above the focal curve, defined by

$$\frac{z^2}{\gamma-\beta} + \frac{x^2}{\alpha-\beta} = 1. \tag{A4}$$

For large x, this curve is approximately a line given by $z \approx x\frac{\sqrt{\gamma-\beta}}{\sqrt{\beta-\alpha}}$. Therefore, the angle that this line forms with the z-axis can be written simply in terms of the viewing angles as

$$\tan\eta = \sqrt{\frac{\sin^2\theta}{\cot 2\psi \sin 2\phi \cos\theta + \cos^2\phi(\cos^2\theta+1) - 1}}. \tag{A5}$$

Any orbits launched initially between the focal curve and the positive z-axis in the x–z plane will be long-axis tubes, which violate axisymmetry. To effectively achieve axisymmetry, the angle η must be small enough for no orbits to be sampled above the focal curve. Since the line defined by the angle η is a lower bound to this curve, if all initial orbits in the positive x–z quadrant are launched outside of the approximate angular region between the z-axis and the angle η, there will be no long-axis tubes in the model.

This expression is derived for Stäckel potentials. However, in the absence of isophotal twists, we expect it to apply reasonably well to more realistic models as they can often be locally approximated by a Stäckel potential (Sanders & Binney 2015). A central SMBH is inconsistent with a Stäckel potential and can thus destroy the ordered orbital structure. However, we suggest that Equation (A5) could give a rough rule-of-thumb for where the boundary between long-axis and short-axis tubes will exist in models from the code, particularly at radii far from the SMBH.

The stellar mass distribution is represented by a multi-Gaussian expansion (MGE) in our models (Cappellari 2002). Each Gaussian component is stratified on similar ellipsoids, and can thus be related to its deprojection via the equations given in Binney (1985). These equations can be rearranged to give

$$\frac{T}{1-T} = \frac{\sin^2\theta}{\cot 2\psi \sin 2\phi \cos\theta + \cos^2\phi(\cos^2\theta+1) - 1}, \tag{A6}$$

where $T = (1-p^2)/(1-q^2)$ of each MGE component. For an MGE with no isophotal twists, each MGE component has the same triaxiality parameter, T. Thus, in this case, the angle, η, can be written simply as:

$$\eta = \tan^{-1}\sqrt{\frac{T}{1-T}}, \tag{A7}$$

where T is the triaxiality parameter for each MGE component. Two examples of triaxial start spaces for NGC 1453 models are shown in Figure 6. The boundary between long-axis tubes and short-axis tubes is well approximated by the angle η for a wide range of galaxy shapes.

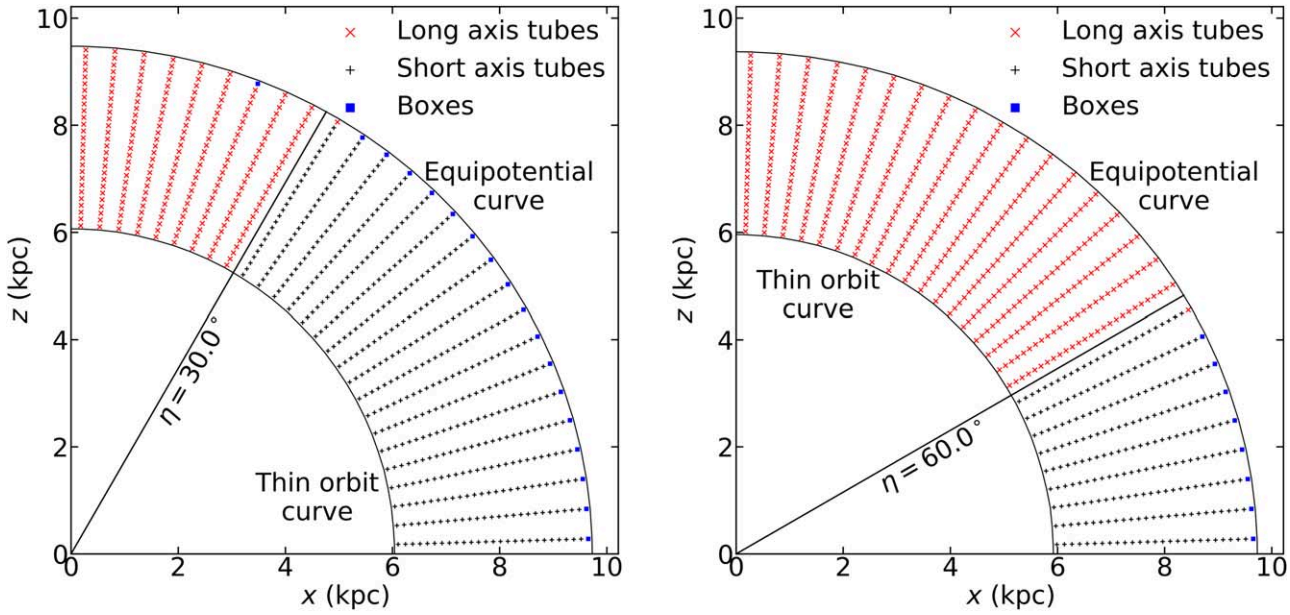

Figure 6. The same as Figure 1 but for two additional mass models with larger triaxiality: (left) triaxiality parameter $T = 0.25$, (luminosity averaged) shape parameters $(u, p, q) = (0.96, 0.95, 0.77)$, and viewing angles $(\theta, \phi, \psi) = (67.^{\circ}62, -28.^{\circ}38, 86.^{\circ}61)$; and (right) $T = 0.75$, $(u, p, q) = (0.96, 0.85, 0.79)$, and $(\theta, \phi, \psi) = (48.^{\circ}74, -51.^{\circ}33, 67.^{\circ}15)$. The diagonal black line in each panel represents the angle η given in Equation (2). As in Figure 1, this angle approximates the boundary separating long-axis (red symbols) and short-axis (black symbols) tube orbits in the x–z start space.

Appendix B
Thin Orbit Finding

The TriOS code uses the thin-orbit curve to construct its start space. This curve has to be found numerically in the x–z plane. For a given angle in this plane, the thin-orbit radius is found by integrating test orbits starting at different radii. For each orbit, the radius of the orbit is recorded each time it passes through the x–z plane. The thin-orbit radius is found by minimizing the difference between the maximum and minimum of these radii.

This algorithm should work for triaxial models but needs some revision in the axisymmetric case, particularly when there is no central density cusp or mass concentration. In this case, when close enough to the center, the potential should be well approximated by a harmonic oscillator. When the potential is axisymmetric, the motion can be regarded as two separate contributions: an oscillation in the z-axis and a closed elliptical orbit about the z-axis. Since the x–y motion constitutes a closed ellipse centered on the z-axis, all orbits will pass through the x–z plane at a fixed x-value, with some z-value. The orbital width is then simply set by the maximum and minimum z-values. Thus, for a given ray in the x–z plane, the orbital width in this plane can be minimized by simply taking the initial radius to be as small as possible. To solve this issue when running an axisymmetric model, we instead record radii when passing through the x–y plane. Closed ellipses will have a finite width in this plane while all thin orbits should pass through this plane in a circle of zero width.

It is unclear how much this issue should affect the resulting orbit libraries. If orbits are sampled starting at the origin instead of the thin orbit, the result should be a less uniform sampling of angular momentum. There should also be some range of energies where the thin-orbit radius is not estimated to be zero or the correct value, but rather somewhere in between. This would result in a significantly nonuniform sampling of angular momentum since orbits passing through the x–z plane within this radius will be undersampled relative those that do not. This issue should be essentially resolved outside of the axisymmetric limit, or if a black hole or density cusp is included. However, axisymmetric studies that use this code with no central cusp may be affected (Hagen et al. 2019).

Appendix C
Mock Recovery Tests

In the above sections, we demonstrated that our changes to the TriOS code result in a consistent, nonzero SMBH mass estimate for our NGC 1453 data set. Here, we show that the changes correctly recover the SMBH mass in a mock data set with known parameter values. Mock tests have been performed within various other Schwarzschild codes (e.g., Valluri et al. 2004; Cretton & Emsellem 2004; Magorrian 2006; Siopis et al. 2009; Vasiliev & Valluri 2020).

For our mock galaxy, we use a flattened version of the spherical potential introduced in Siopis et al. (2009). These models have an axisymmetric gravitational potential given by

$$\Phi(R, z) = \frac{1}{2} V_c^2 \ln\left(\frac{R^2 + z^2/q_\Phi^2}{1\ \mathrm{pc}^2}\right) - \frac{GM_{\mathrm{BH}}}{\sqrt{R^2 + z^2}}, \quad \text{(C1)}$$

where q_Φ is the flattening of the potential due to the extended mass distribution. The stellar distribution function (DF) is chosen to have a Michie-like form:

$$f = A \exp\left\{-\left[\frac{E + L_z^2/(2r_a^2)}{\sigma^2}\right]\right\} L_z^{2N} \sqcap (E_1, E, E_2), \quad \text{(C2)}$$

where A is the normalization, and r_a, N, σ, E_1, and E_2 are parameters of the model: r_a is an anisotropy distance, N controls the L_z dependence, σ is a characteristic velocity dispersion, and E_1 and E_2 are energy cutoffs. The symbol $\sqcap$

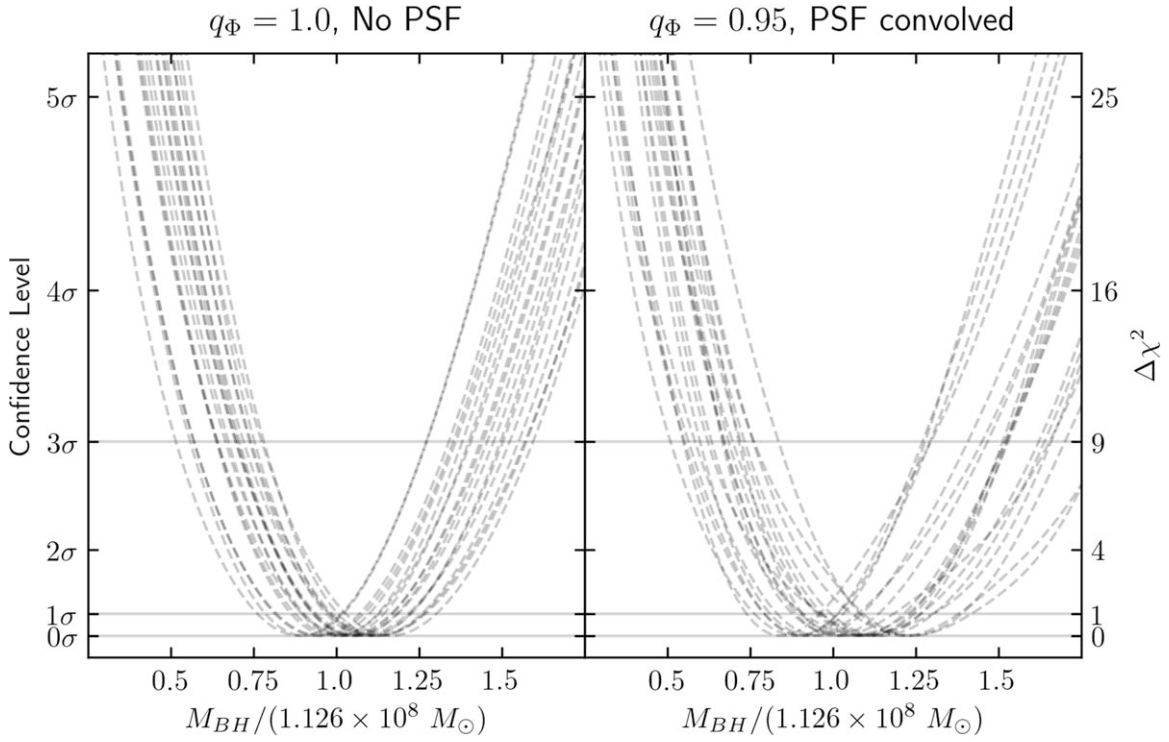

Figure 7. Illustration of the $M_{\rm BH}$ constraints for the mock data sets described in the text. Each dashed curve represents a separate realization of the noise. In the left panel, the potential is spherical, no PSF convolution is performed, and each DF component has 5×10^8 stars. In the right panel, the model is flattened, projected stellar positions are convolved with a circular Gaussian PSF with a standard deviation of 5 pc, and each DF component has 5×10^9 stars. The 1D χ^2 curves are obtained by marginalizing over V_c in the smoothed 2D χ^2 landscape generated by Gaussian process regression with a squared-exponential covariance function (Pedregosa et al. 2011).

denotes a step function defined by

$$\sqcap(E_1, E, E_2) = \begin{cases} 1, & \text{if } E_1 \leqslant E \leqslant E_2 \\ 0, & \text{otherwise.} \end{cases} \quad \text{(C3)}$$

Because L_z only enters Equation (C2) in even powers, there is additional freedom in how f differs for positive and negative values of L_z. Here, we set a fixed fraction of stars to rotate in each positive direction. In this model, the stars are essentially regarded as massless tracers of the underlying potential in Equation (C2). Even when the potential is chosen to be spherical, the stellar distribution function can be axisymmetric.

We use the same potential parameters as Siopis et al. (2009), with $V_c = 220\,\mathrm{km\,s^{-1}}$ and $M_{\rm BH} = 1.126 \times 10^8\,M_\odot$. We generated two models: one model with a spherical potential ($q_\Phi = 1$) to compare with Siopis et al. (2009), and one model with a flattened potential ($q_\Phi = 0.95$). The models both have a sphere of influence of about 10 pc. We also use the same two component DF parameters as Siopis et al. (2009): the first component is a nonrotating nearly spherical bulge-like component, which has $\sigma = 160\,\mathrm{km\,s^{-1}}$, $r_a = 600$ pc, $N = 0$, $E_1 = \Phi(10\,\mathrm{pc})$, and $E_2 = \Phi(1000\,\mathrm{pc})$ with equal numbers of stars having positive and negative L_z; the second component is a rotating disk-like component, which has $\sigma = 120\,\mathrm{km\,s^{-1}}$, $r_a = 200$ pc, $N = 2$, $E_1 = \Phi(10\,\mathrm{pc})$, and $E_2 = \Phi(1000\,\mathrm{pc})$ with three-fourths of the stars having positive L_z and one-fourth having negative L_z. The two components have equal numbers of stars.

We draw points in phase space from this distribution function for each star to generate mock data. We use a nearly edge-on projection, with an inclination of $\theta = 89°$. For the model with the flattened potential, we convolve the projected positions with a circular Gaussian PSF with a standard deviation of 5 pc. We bin the stars into mock integral field unit data with a resolution of 10 pc, with a square field of view of 1000 pc. We fit an MGE to the projected surface brightness. We then run Voronoi binning on all bins with central radii >20 pc, resulting in 12 inner unbinned kinematic points and 108 larger outer bins. In order to keep the bins between the two models fixed, we use the Voronoi bins derived from the spherical potential. Each LOSVD is fit with a Gauss–Hermite expansion up to h_{12}. Gaussian noise is added to each LOSVD bin, resulting in a scatter of about 0.03 in each moment and about $0.03\sqrt{2}\,\sigma$ in the average velocity and velocity dispersion for each bin. We draw 20 realizations of this noise, and run the updated TriOS code for each realization.

Figure 7 shows the resulting constraint on $M_{\rm BH}$ for each noise realization. The left panel is for the mock in the spherical potential, while the right panel is for the mock in the flattened potential. The kinematic contribution to the reduced χ^2 in these realizations ranges from 0.81 to 0.94 for the spherical potential and 0.71 to 0.84 for the flattened potential, indicating a good fit to the projected kinematics for all realizations.

The average SMBH masses and corresponding sample standard deviations from these combined 20 runs are $M_{\rm BH} = (1.17 \pm 0.09) \times 10^8 M_\odot$ for the spherical potential, and $M_{\rm BH} = (1.18 \pm 0.13) \times 10^8 M_\odot$ for the flattened potential. In both test cases, the estimated $M_{\rm BH}$ values are in excellent agreement with the true value of $M_{\rm BH} = 1.126 \times 10^8 M_\odot$.

ORCID iDs

Matthew E. Quenneville https://orcid.org/0000-0002-6148-5481

Christopher M. Liepold https://orcid.org/0000-0002-7703-7077
Chung-Pei Ma https://orcid.org/0000-0002-4430-102X

References

Ahn, C. P., Seth, A. C., Cappellari, M., et al. 2018, ApJ, 858, 102
Binney, J. 1985, MNRAS, 212, 767
Binney, J., & Spergel, D. 1984, MNRAS, 206, 159
Binney, J., & Tremaine, S. 2008, Galactic Dynamics (2nd ed.; Princeton, NJ: Princeton Univ. Press), 75
Cappellari, M. 2002, MNRAS, 333, 400
Cappellari, M., Bacon, R., Bureau, M., et al. 2006, MNRAS, 366, 1126
Cretton, N., de Zeeuw, P. T., van der Marel, R. P., & Rix, H.-W. 1999, ApJS, 124, 383
Cretton, N., & Emsellem, E. 2004, MNRAS, 347, L31
de Zeeuw, T. 1985, MNRAS, 216, 273
Feldmeier-Krause, A., Zhu, L., Neumayer, N., et al. 2017, MNRAS, 466, 4040
Franx, M. 1988, MNRAS, 231, 285
Gebhardt, K., Richstone, D., Kormendy, J., et al. 2000, AJ, 119, 1157
Hagen, J. H. J., Helmi, A., & Breddels, M. A. 2019, A&A, 632, A99
Heiligman, G., & Schwarzschild, M. 1979, ApJ, 233, 872
Hernquist, L. 1990, ApJ, 356, 359
Jin, Y., Zhu, L., Long, R. J., et al. 2020, MNRAS, 491, 1690
Joseph, C. L., Merritt, D., Olling, R., et al. 2001, ApJ, 550, 668
Liepold, C. M., Quenneville, M. E., Ma, C.-P., et al. 2020, ApJ, 891, 4
Ma, C.-P., Greene, J. E., McConnell, N., et al. 2014, ApJ, 795, 158
Magorrian, J. 2006, MNRAS, 373, 425
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, J. Mach. Learn. Res., 12, 2825, https://dl.acm.org/doi/10.5555/1953048.2078195
Pfenniger, D. 1984, A&A, 141, 171
Poci, A., McDermid, R. M., Zhu, L., & van de Ven, G. 2019, MNRAS, 487, 3776
Richstone, D. O., & Tremaine, S. 1984, ApJ, 286, 27
Richstone, D. O., & Tremaine, S. 1985, ApJ, 296, 370
Rix, H.-W., de Zeeuw, P. T., Cretton, N., van der Marel, R. P., & Carollo, C. M. 1997, ApJ, 488, 702
Sanders, J. L., & Binney, J. 2015, MNRAS, 447, 2479
Schwarzschild, M. 1979, ApJ, 232, 236
Schwarzschild, M. 1993, ApJ, 409, 563
Seth, A. C., van den Bosch, R., Mieske, S., et al. 2014, Natur, 513, 398
Shapiro, K. L., Cappellari, M., de Zeeuw, T., et al. 2006, MNRAS, 370, 559
Siopis, C., Gebhardt, K., Lauer, T. R., et al. 2009, ApJ, 693, 946
Sridhar, S., & Touma, J. 1999, MNRAS, 303, 483
Sridhar, S., & Touma, J. 2016, MNRAS, 458, 4129
Thomas, J., Saglia, R. P., Bender, R., et al. 2004, MNRAS, 353, 391
Valluri, M., Merritt, D., & Emsellem, E. 2004, ApJ, 602, 66
Valluri, M., Shen, J., Abbott, C., & Debattista, V. P. 2016, ApJ, 818, 141
van de Ven, G., De Zeeuw, P. T., & Van Den Bosch, R. C. E. 2008, MNRAS, 385, 614
van den Bosch, R. C. E., & de Zeeuw, P. T. 2010, MNRAS, 401, 1770
van den Bosch, R. C. E., van de Ven, G., Verolme, E. K., Cappellari, M., & de Zeeuw, P. T. 2008, MNRAS, 385, 647
van der Marel, R. P., Cretton, N., de Zeeuw, P. T., & Rix, H.-W. 1998, ApJ, 493, 613
Vasiliev, E., & Valluri, M. 2020, ApJ, 889, 39
Verolme, E. K., Cappellari, M., Copin, Y., et al. 2002, MNRAS, 335, 517
Walsh, J. L., van den Bosch, R. C. E., Barth, A. J., & Sarzi, M. 2012, ApJ, 753, 79
Walsh, J. L., van den Bosch, R. C. E., Gebhardt, K., et al. 2015, ApJ, 808, 183
Walsh, J. L., van den Bosch, R. C. E., Gebhardt, K., et al. 2017, ApJ, 835, 208
Walsh, J. L., van den Bosch, R. C. E., Gebhardt, K., et al. 2016, ApJ, 817, 2
Zhu, L., van de Ven, G., van den Bosch, R., et al. 2018, NatAs, 2, 233