Harmonic-oscillator M2M

Jo Bovy[1,2,3]

ABSTRACT

This note presents some enhancements to the standard *made-to-measure* (M2M) approach for fitting steady-state distribution functions to kinematic data in the context of a simple one-dimensional harmonic-oscillator potential. Specifically, this note aims to do the following:

- Fit for nuisance parameters in addition to the orbital weights; specifically, fit for the Sun's height $z_\odot$ above the midplane;
- Fit for the parameters of the potential; specifically, fit for the period ω of the harmonic oscillator;
- Run MCMC for the weights using Hybrid Monte Carlo;
- Run MCMC for the nuisance parameters using Hybrid Monte Carlo as well;
- Run MCMC for the potential parameters.

With these improvements, M2M can be used to fully fit observable data and return proper uncertainties on all parameters.

1. Online material

These notes accompany an online ipython notebook that contains experiments and figures using the methodology below. This notebook is located at

`https://github.com/jobovy/simple-m2m/blob/master/py/HOM2M.ipynb`

The few movies that are part of this notebook can only be viewed on nbviewer at

`http://nbviewer.jupyter.org/github/jobovy/simple-m2m/blob/master/py/HOM2M.ipynb` .

[1] Department of Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON, M5S 3H4, Canada; bovy@astro.utoronto.ca

[2] Center for Computational Astrophysics, Flatiron Institute, 162 5th Ave, New York, NY 10010, USA

[3] Alfred P. Sloan Fellow

2. Introduction

For now, the reader is assumed to be familiar with the standard made-to-measure modeling technique. If not, Syer & Tremaine (1996) and Dehnen (2009) are useful references to get started.

3. Setup

We consider a simple one-dimensional system where the potential is a simple harmonic oscillator. The phase-space coordinates are (z, v_z) and the potential is

$$\Phi(z) = \frac{\omega^2\,z^2}{2}\,. \tag{1}$$

In this potential, we attempt to match a population drawn from the following distribution function

$$f(z, v_z) \propto e^{-E/\sigma_t^2}\,, \tag{2}$$

where $E = \omega^2\,z^2/2 + v_z^2/2$ is the energy and $\sigma_t = 0.1$ is the true velocity dispersion. This distribution function is isothermal, it has the same velocity dispersion at all heights.

To fit this distribution function using M2M, we start with $(z_i, v_{z,i})$ drawn with uniform weights w_i from a similar isothermal distribution function, but with a larger σ: $f(z, v_z) \propto e^{-E/\sigma_{\mathrm{in}}^2}$, with σ_{in} typically 0.2. It is then easy to see that the correct output weights should be

$$z_i = \exp\left(-E_i\left[1/\sigma_t^2 - 1/\sigma_{\mathrm{in}}^2\right]\right)\,. \tag{3}$$

Orbit integration in the harmonic-oscillator potential is analytic and we simply have that

$$z_i(t) = \quad A_i \quad \cos\left(\omega\,t + \phi_i\right)\,, \tag{4}$$

$$v_{z,i}(t) = -A_i\,\omega\,\sin\left(\omega\,t + \phi_i\right)\,, \tag{5}$$

where

$$A_i = \frac{\sqrt{2\,E_i}}{\omega} = \sqrt{z_i^2(0) + v_{z,i}^2(0)/\omega^2}\,, \tag{6}$$

$$\phi_i = \mathrm{arctan2}(-v_{z,i}(0)/\omega, z_i(0))\,, \tag{7}$$

in which $(z_i(0), v_{z,i}(0))$ is the initial phase-space position and arctan2 is the arc-tangent function that chooses the quadrant correctly. Sampling initial $(z_i(0), v_{z,i}(0))$ from $f(z, v_z) \propto e^{-E/\sigma^2}$ is simple: sample E_i from the exponential distribution and convert it to A_i; sample ϕ_i uniformly between 0 and 2π.

4. Standard M2M

We first describe the standard M2M case. Standard M2M models a steady-state distribution function as a set of N particles $(z_i, v_{z,i})$ indexed by i orbiting in a fixed potential. Each particle has a weight w_i that is adjusted on-the-fly during orbit integration to fit a set of constraints, like the density in bins, or the velocity dispersion. By only adjusting the weights w_i on timescales $\gg$ the orbital timescale, an approximate equilibrium distribution is obtained[1].

In practice, M2M maximizes an objective function F that represents a balance between reproducing the constraints and smoothness of the distribution function (through a maximum-entropy constraint)

$$F = \mu S - \frac{1}{2}\sum_j \chi_j^2 \tag{8}$$

where $S = -\sum_i w_i \ln(w_i/\hat{w}_i)$ is the entropy and χ_j^2 are constraints expressed in a χ^2 type manner. Constraints are expressed as a kernel applied to the distribution function $f(z, v_z)$:

$$Y_j = \int \mathrm{d}z\mathrm{d}v_z\, K_j(z, v_z) f(z, v_z) \tag{9}$$

which for the N-body relation is computed as

$$y_j = \sum_i w_i K_j(z_i, v_{z,i})\,. \tag{10}$$

To illustrate the standard M2M case, we use the density observed at a few points, so

$$\rho(\tilde{z}_j) = \frac{1}{N}\sum_i w_i K^\rho(|\tilde{z}_j + z_\odot - z_i|; h)\,, \tag{11}$$

where $K^\rho(r;h)$ is a kernel function that integrates to one ($\int \mathrm{d}r\, K^\rho(r;h) = 1$), we have assumed that $\sum_i w_i = N$, and we assume that the observations are done as a function of $\tilde{z}$, which is measured with respect to the Sun's position, located at $z_\odot$ from the $z = 0$ midplane. We form χ_j^2 as

$$\chi_j^2 = [\Delta_j^\rho]^2 = \left(\rho(\tilde{z}_j)/\rho_j^{\mathrm{obs}} - 1\right)^2\,. \tag{12}$$

---

[1] Unlike in Schwarzschild modeling, where orbits are integrated for hundreds of dynamical times and the observables are fit using these orbits, this is a more approximate equilibrium. But in some sense it is *exactly* the correct state to fit, because stellar populations are only in a quasi-stationary state for a few orbital periods (on longer timescales they will evolve due to interactions with clouds, spiral arms, satellite galaxies, etc. or due to the evolution of the underlying gravitational potential.

The M2M *force of change* equation is then given by

$$\begin{aligned}\frac{\mathrm{d}w_i}{\mathrm{d}t} &= \epsilon w_i \,\frac{\partial F}{\partial w_i} \\ &= -\epsilon w_i \left[\mu\left(\ln\left[w_i/\hat{w}_i\right]+1\right)+\sum_j \Delta_j^\rho\, K_j^\rho(z_i;h)/\rho_j^{\mathrm{obs}}\right]\,.\end{aligned} \tag{13}$$

We solve this equation using a simple Euler method with a fixed step size, computing the orbital evolution as we go along using equations (4) and (5). This method for optimizing the objective function can be thought of as a sort of gradient ascent. In this interpretation, ϵ is adjusted such that substantial changes to the orbital weights only happen on timescales $\gg$ the orbital timescale, which pushes the weights to an equilibrium distribution.

Syer & Tremaine (1996) propose to lessen the impact of Poisson noise due to the finite number of N-body particles by smoothing the Δ_j^ρ deviation that appears in equation (13) with a smoothed version $\tilde{\Delta}_j$. In the end, this leads one to solve for $\tilde{\Delta}_j$ using the differential equation

$$\frac{\mathrm{d}\tilde{\Delta}_j}{\mathrm{d}t} = \alpha\left(\Delta_j - \tilde{\Delta}_j\right)\,, \tag{14}$$

where α is another timescale parameter. Because we only want to smooth on shorter timescales than that over which we substantially change the weights, we typically need $\alpha > \epsilon$.

5. Fitting for the nuisance parameter $z_\odot$

In the observed density above, we have assumed that we know the Sun's distance from the midplane from which the stars are observed. Now suppose that we do not know $z_\odot$. Then when we compare the density to the observed density $\rho(\tilde{z})$ as a function of $\tilde{z}$, we do not know how to shift the model z_i to $\tilde{z}$. The standard approach would be to run multiple instances of the standard M2M algorithm for different values of $z_\odot$ and to then determine the maximum of $F(z_\odot)$. This becomes expensive if there are many nuisance parameters.

Here we propose to simply update $z_\odot$ in tandem with the weights. To do this, we compute the force of change for $z_\odot$, which in the case of fitting the density is

$$\frac{\mathrm{d}z_\odot}{\mathrm{d}t} = -\epsilon_z \sum_j \Delta_j^\rho/\rho_j^{\mathrm{obs}} \sum_i w_i \left.\frac{\mathrm{d}K^\rho(r)}{\mathrm{d}r}\right|_{|\tilde{z}+z_\odot-z_i|} \mathrm{sign}(\tilde{z}+z_\odot-z_i)\,, \tag{15}$$

where we have introduced a different ϵ_z parameter to control how fast $z_\odot$ is changed during the maximization. We expect that we want $\epsilon_z < \epsilon$, because we want to change $z_\odot$ only on timescales that are longer than the timescales over which we adjust the weights.

6. Fitting for the potential parameter ω

Next, we want to fit for the potential parameter ω as well. To do this, we need to introduce a velocity constraint as well, because the potential cannot be constrained using the density alone. For simplicity, we choose the velocity constraint to be the density times the mean squared velocity

$$\begin{aligned}\rho \times \langle v^2\rangle(\tilde{z}_j) &= \frac{1}{N}\sum_i w_i K^v(|\tilde{z}_j + z_\odot - z_i|, v_{z,i}; h)\\ &= \frac{1}{N}\sum_i w_i\, vzi^2\, K^\rho(|\tilde{z}_j + z_\odot - z_i|; h)\,,\end{aligned} \tag{16}$$

where we have set $K^v(z, v_z) = v^2\, K^\rho(z)$. We assume that each velocity constraint adds the following to the total χ^2

$$\chi^2_{j,v} = [\Delta^v_j]^2 = \left(\frac{\rho \times \langle v^2\rangle(\tilde{z}_j)}{[\rho \times \langle v^2\rangle]^{\mathrm{obs}}_j} - 1\right)\,. \tag{17}$$

The force of change equation for the weights then gets an additional contribution due to the velocity constraint. This additional contribution (which gets added to the right-hand side of equation [13]) is

$$-\epsilon_v w_i \sum_j \Delta^v_j\, v^2_{z,i}\, K^\rho_j(|\tilde{z}_j + z_\odot - z_i|; h)/[\rho \times \langle v^2\rangle]^{\mathrm{obs}}_j\,. \tag{18}$$

Here we have kept the freedom to use a different epsilon parameter ϵ_v to adjust the weighting between the density and velocity constraints.

Similar to the force of change for $z_\odot$, we can also compute the force of change for ω. This is given by

$$\begin{aligned}\frac{\mathrm{d}\omega}{\mathrm{d}t} = -\,\epsilon_\omega \sum_j\sum_i w_i \Bigg[&\left\{\frac{\Delta^\rho_j}{\rho^{\mathrm{obs}}_j} + \frac{\Delta^v_j}{[\rho \times \langle v^2\rangle]^{\mathrm{obs}}_j}\, v^2_{z,i}\right\} \left.\frac{\mathrm{d}K^\rho(r)}{\mathrm{d}r}\right|_{|\tilde{z}+z_\odot - z_i|} \mathrm{sign}(\tilde{z} - z_\odot - z_i)\frac{\partial z_i}{\partial \omega}\\ &+ 2\frac{\Delta^v_j}{[\rho \times \langle v^2\rangle]^{\mathrm{obs}}_j}\, v_{z,i}\, K^\rho(|\tilde{z} + z_\odot - z_i|)\,\frac{\partial v_{z,i}}{\partial\omega}\Bigg]\,.\end{aligned} \tag{19}$$

The problem is now that the instantaneous phase-space position $(z_i, v_{z,i})$ does not depend on ω.

Nevertheless, we can compute the partial derivatives $\partial z_i/\partial\omega$ and $\partial v_{z,i}/\partial\omega$ using equations (4) and (5). This gives

$$\frac{\partial z_i}{\partial\omega} = \frac{t}{\omega} v_{z,i}\,, \tag{20}$$

$$\frac{\partial v_{z,i}}{\partial\omega} = -\omega\, t z_i\,. \tag{21}$$

At $t = 0$ these derivatives are zero (because the instantaneous phase-space position does not depend on the potential), but if we integrate for a short time, the phase-space positions computed in different potentials would be different, as given by these partial derivatives for short times. Thus, we can drop these into equation (19) and absorb the t parameter in the above equation in ϵ_ω. We thus change the potential parameter ω on-the-fly as well, again requiring ϵ_ω to be such that substantial changes to ω only occur on timescales $\gg$ than the timescale over which substantial changes in the weights occur. Because we change ω on the fly, each time step, we need to re-compute the (A_i, ϕ_i) parameters using equations (6) and (7).

This appears to work quite well! For more complicated potentials, these partial derivatives can be computed by solving the differential equation for the difference in z and v_z for small potential changes.

REFERENCES

Syer, D., & Tremaine, S. 1996, MNRAS, 282, 223

Dehnen, W. 2009, MNRAS, 395, 1079

This preprint was prepared with the AAS LaTeX macros v5.2.