In 2023, if all we have is a **very precise mass** (or circular velocity, or core radius, or etc.) for each Local-Group member, we will have **failed miserably**.

- This is not like cosmological parameter estimation!
- Better data must bring **qualitatively new insights**.
Mapping the dark matter and inferring the formation history of the Local Group

David W. Hogg

Center for Cosmology and Particle Physics, Department of Physics, New York University
Center for Data Science, New York University
Max-Planck-Institut für Astronomie

2015 June 01

work in collaboration with: Bovy, Johnston, Ness, Price-Whelan, Rix, Sanderson
Inference of the potential $\Phi$ given phase-space points $X_n$ requires enormous marginalizations (or profiles).

- By “potential $\Phi$” I mean the entire formation history of the Local Group.
- There might be a tiny subset of the stars that contains almost all of the information.

Extending $X_n$ to include more “tags” (especially chemical tags) is expected to be informative:

- (But vastly increases the number of nuisances.)
- We aren’t ready to execute this program.

All projects so far have just been toys.
challenges

▶ our prior beliefs are generated by **non-linear physical cosmology**
  ▶ (plus small modifications and adjustments)
  ▶ physical or effective models of star formation?
  ▶ prior and likelihood function both intractable?
▶ data sets are vast and **noisy**
  ▶ delivering probabilistic outputs from telescopes
  ▶ writing inferences with posteriors in and posteriors out
  ▶ non-parametric models get huge (\(N^{2.6}\)-ish or worse)
▶ **stars** are non-trivial objects
  ▶ consistent parameters across surveys
  ▶ consistent chemical abundances across \(T_{\text{eff}}, \log g, v \sin i\), etc.
dynamical inference

- The positions and velocities $X_n$ are **initial conditions**.
- The potential $\Phi$ and its history describe the dynamics.
- **You can’t infer one from the other.**
  - (cf., undergraduate dynamics)
The positions and velocities $X_n$ are **initial conditions**.

The potential $\Phi$ and its history describe the dynamics.

**You can’t infer one from the other.**

- (cf., undergraduate dynamics)
- but what if we make other assumptions?

The Milky Way is long-lived.

- and we aren’t seeing it at a special time
- and there are no significant resonances (stars are independent)

Then we know what to do!
The positions and velocities $X_n$ are **initial conditions**.

The potential $\Phi$ and its history describe the dynamics.

**You can’t infer one from the other.**

- (cf., undergraduate dynamics)
- but what if we make other assumptions?

The Milky Way is long-lived.

- and we aren’t seeing it at a special time
- and there are no significant resonances (stars are independent)

Then we know what to do!

- (too bad all of these assumptions are wrong!)

---

dynamical inference
Can we learn the gravitational force law in the Solar System given only a snapshot of the planet phase-space positions?

Rough idea:

- For a long-lived, integrable, non-resonant system, angles are uniformly distributed, actions are conserved.
- The action distribution matters and is completely unknown.
- Model needs flexibility to discover action-space structure
- Many assumptions (spherical, power-law, time-invariant, integrable)!

\[ p(x, v \mid I, \theta, \Phi) = \int p(x, v \mid I, \theta, \Phi) \, p(I \mid \alpha) \, p(\theta) \, dI \, d\theta \, d\alpha \]

- \( p(x, v \mid I, \theta, \Phi) \) is really a transformation
- \( \alpha \) is HUGE
- Jacobian of transformation from \( x, v \) to \( I, \theta \) coordinates describes the information in the data.
April Fools’ Bovy et al., 0903.5308
April Fools’  Bovy et al., 0903.5308
April Fools’ Bovy et al., 0903.5308
April Fools’ Bovy et al., 0903.5308
April Fools’ Bovy et al., 0903.5308
April Fools’ Bovy *et al.*, 0903.5308
Inference of the potential \( \Phi \) given phase-space positions \( X_n \) is possible.

Marginalization was incredibly expensive.
- \( \alpha \) is huge
- CPU-days for \( N = 8 \) particles!

Some “stars” were far more valuable than others.

Lessons for 2023?
- Computational expense daunting.
- Key stars are worth finding (if they exist).
- Things won’t be integrable, stationary, or symmetric.
extended distribution functions

- stellar orbital actions are conserved “tags” on stars.
  - (Only if the potential is **integrable**, which it isn’t!)
- so are stellar “birth dates”
- surface (birth-environment) abundances (maybe 30-dimensional?)
  - (But no-one has ever demonstrated this.)
- Naively these birth indicators must be exceedingly informative, if we can measure them.
there is no hope of good, consistent point-estimates of chemical abundances
  (not for 2023, anyway, and not good enough for “tagging”)
  prove me wrong!
  (and keep working; this is God’s work)
we have to see the chemical tagging as being done simultaneously with the dynamical fitting.
  either effective models of stellar spectra, or else
  parameterizing all stellar model uncertainties and treating them as nuisances
  (these models will get enormous)
data-driven model of stellar spectra
  ▶ stellar parameter estimation and chemical abundances
  ▶ get different data sets onto the same “system”
  ▶ capitalize on enormous data sets (APOGEE demo)

train-and-test framework
  ▶ trained on cluster stars will well understood chemical abundances
  ▶ (not large enough) range of $T_{\text{eff}}$ and log $g$ and chemistry
  ▶ but contains a realistic spectral noise model

designed for computational tractability
  ▶ fast
  ▶ generalizable and extensible
\[
\ln p(f_n \mid \ell_n, \theta) = \sum_{\lambda=1}^{L} \ln p(f_{\lambda n} \mid \ell_n, \theta_{\lambda}, s_{\lambda}^2)
\]

\[
\ln p(f_{\lambda n} \mid \ell_n, \theta_{\lambda}, s_{\lambda}^2) = -\frac{1}{2} \frac{[f_{\lambda n} - \theta_{\lambda}^T \cdot \ell_n]^2}{\sigma_{\lambda n}^2 + s_{\lambda}^2} + \ln(\sigma_{\lambda n}^2 + s_{\lambda}^2)
\]

\[
\ell^T \equiv \{1, T_{\text{eff}}, \log g, [\text{Fe/H}], T_{\text{eff}}^2, T_{\text{eff}} \log g, \cdots, [\text{Fe/H}]^2\}
\]

\[
\theta^T \equiv \{\theta_{\lambda}, s_{\lambda}^2\}_{\lambda=1}^{L}
\]

**training step:** optimize parameters \(\theta\) at fixed labels \(\ell\) using training-set data

- linear least squares
- every wavelength \(\lambda\) treated independently

**test step:** optimize labels \(\ell\) at fixed parameters \(\theta\) using test-set (survey) data

- non-linear optimization
The Cannon Ness et al., 1501.07604

Teff = 4750, log g = 3.0, [Fe/H] = 0.15

Teff = 4849, log g = 2.2, [Fe/H] = -1.0

Teff = 3614, log g = 0.4, [Fe/H] = -0.68

Teff = 5003, log g = 2.8, [Fe/H] = -0.71
The Cannon Ness et al., 1501.07604
The Cannon Ness et al., 1501.07604
The Cannon Ness et al., 1501.07604
We can transfer labels from a training set to all APOGEE data.

Our labels are as accurate as those based on physical models.

The system is still a toy; we need to:

- demonstrate that we can go beyond \([\text{Fe/H}]\)
- generalize to simultaneous fitting of training and test data
- generalize to **simultaneous fitting of stellar kinematics and spectra**
- (and the ISM)
Inference of the potential $\Phi$ given phase-space points $X_n$ requires enormous marginalizations (or profiles).

- By “potential $\Phi$” I mean the entire formation history of the Local Group.
- There might be a tiny subset of the stars that contains almost all of the information.

Extending $X_n$ to include more “tags” (especially chemical tags) is expected to be informative:

- (But vastly increases the number of nuisances.)
- We aren’t ready to execute this program.

All projects so far have just been toys.