Kepler's Sieve:
Learning Asteroid Orbits from Telescopic Observations

A dissertation presented

by

Michael S. Emanuel

to

Institute of Applied Computational Science

in partial fulfillment of the requirements

for the degree of

Master of Science

in the subject of

Data Science

Harvard University

Cambridge, Massachusetts

May 2020

Dissertation Advisors:
Professor Pavlos Protopapas
Professor Christopher H. Rycroft

Author:
Michael S. Emanuel

Kepler's Sieve:
Learning Asteroid Orbits from Telescopic Observations

Abstract

A novel method is presented to learn the orbits of asteroids from a large data set of telescopic observations. The problem is formulated as a search over the six dimensional space of Keplerian orbital elements. Candidate orbital elements are initialized randomly. An objective function is formulated based on log likelihood that rewards candidate elements for getting very close to a fraction of the observed directions. The candidate elements and the parameters describing the mixture distribution are jointly optimized using gradient descent. Computations are performed quickly and efficiently on GPUs using the TensorFlow library.

The methodology of predicting the directions of telescopic detections is validated by demonstrating that out of approximately 5.69 million observations from the ZTF dataset, 3.75 million (65.71%) fall within 2.0 arc seconds of the predicted directions of known asteroids. The search process is validated on known asteroids by demonstrating the successful recovery of their orbital elements after initialization at perturbed values. A search is run on observations that do not match any known asteroids. I present orbital elements for [5] new, previously unknown asteroids.

Exact number of new asteroids presented

All code for this project is publicly available on GitHub at github.com/memanuel/kepler-sieve.

Contents

Acknowledgments

I would like to thank my advisor, Pavlos Protopapas, for suggesting this topic and for his consistent support, advice and encouragement.

I would like to thank Chris Rycroft, my secondary advisor, for guiding me through a paper in Applied Math 225 in which I explored numerical integrators for solving solar system orbits.

I would like to thank Matt Holman and Matt Payne from the Center for Astrophics (CFA) for their advice on state of the art solar system integrators.

Most importantly, I would like to thank my wife Christie for her love and support. The Covid-19 crisis struck just as my work on this thesis kicked into high gear. It would not have been possible to complete it without extraordinary understanding and support from her.

To my my children Victor and Renée

Chapter 1

Analysis of ZTF Asteroid Detections

1.1 Introduction

Zwicky Transient Facility (ZTF) is a time-domain survey of the northern sky that had first light at Palomar Observatory in 2017. It is run by CalTech. My advisor Pavlos suggested it as a data source for this project. The ZTF dataset has two major advantages for searching for asteroids:

- ZTF gives a wide and fast survey of the key, covering over 3750 square degrees an hours to a depth of 20.5 mag
- A machine learning pipeline has been developed to classify a subset of ZTF detections that are classified as probable asteroids

The data set I analyze here consists of all ZTF detections that were classified as asteroids. Data on each detection include:

- **ObjectID** an identifier of the likely ojbect associated with this detection; multiple detections often share the same ObjectID
- **CandidateID** a unique integer identifier of each detection
- **MJD** The time of the detection as an MJD
- **RA** The right ascension of the detection
- **Dec** The declination of the detection

- **mag** The apparent magnitude of the detection

Available data also includes a number of additional fields that were not used in the analysis.

ALeRCE (Automatic Learning for the Rapid Classification of Events) is an astronomical data broker. ALeRCE provides a convenient API to access the ZTF asteroid data, which can be installed with `pip`. I used ALeRCE on this project to download the ZTF asteroid data set.

1.2 Exploratory Data Analysis of ZTF Asteroid Data

Before plowing into the search for new asteroids, I conducted an exploratory data analysis (EDA) of the ZTF asteroid dataset. This can be followed interactively in the Jupyter notebook `05_ztf_data.ipynb`. I took a download of the data running through 26-Feb-2020. The first detection is on 01-Jun2018. The dataset contains 5.69 total detections. The volume of detections increases very significantly beginning in July 2019; for practical purposes the dataset consists of 8 months of detections spanning July 2019 through February 2020.

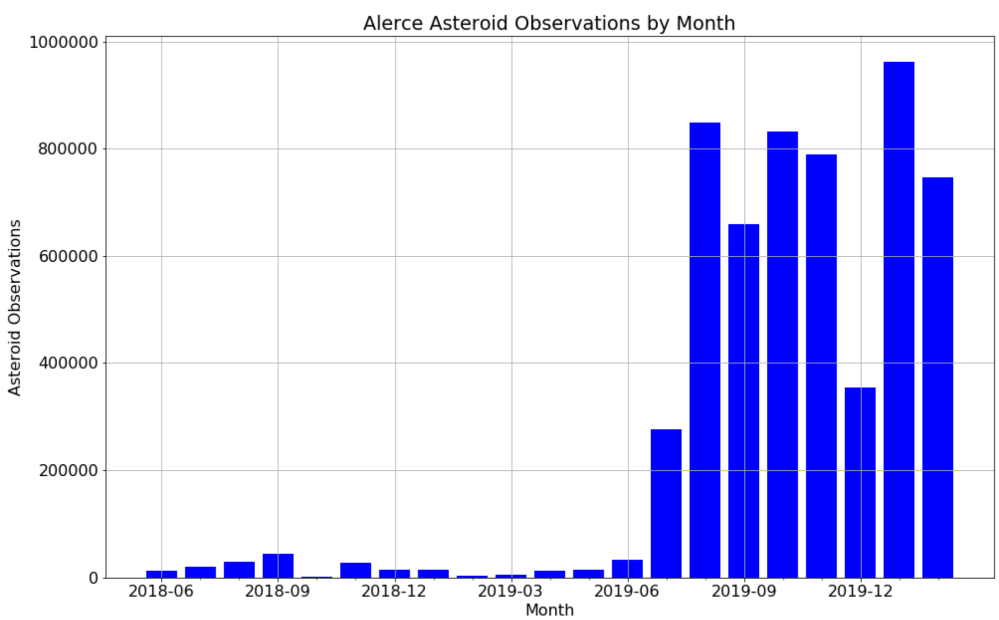

Figure 1.1: *The topographic coordinate system, courtesy of* Wikipedia

1.3 Finding the Nearest Asteroid to Each ZTF Observation

1.4 Analyzing the Distribution of Distance to the Nearest Asteroid

1.5 Setup

References

[1] Hanno Rein, Shang-Fei Liu
REBOUND: An open-source multi-purpose N-body code for collisional dynamics.
Astronomy & Astrophysics. November 11, 2011.
arXiv: 1110.4876v2

[2] Hanno Rein, David S. Spiegel
IAS15: A fast, adaptive high-order integrator for gravitational dynamics, accurate to machine precision over a billion orbits.
Monthly Notices of the Royal Astronomical Society. Printed 16 October 2014.
arXiv: 1405.4779.v2

[3] Murray, C. D.; Dermott, S.F.
Solar System Dynamics
Cambridge University Press. 1999