Two-point angular correlation function (TPACF)

Robert J. Brunner, Volodymyr Kindratenko, Dave Semeraro
National Center for Supercomputing Applications, University of Illinois

Introduction

Correlation functions of the positions of cosmological sources are one the fundamental statistical tools in cosmology. They are used to address one of the fundamental questions in cosmology: how the matter is distributed in the Universe? They describe the probability that for a given object, such as a galaxy or star, another object will be found within a given distance.

Prior to the Sloan Digital Sky Survey (SDSS), the astronomical community was lacking sufficient amount of observational data to make precise measurements of this statistic. With the SDSS, the astronomical community has gone from data "starvation" to data "saturation", and as a result, there is currently a need for techniques and software that are capable of processing large volumes of observed data in a practical amount of time.

Description

The application computes the TPACF for a given dataset: the frequency distribution of angular separations between coordinate positions on the celestial sphere, as compared to randomly distributed coordinate positions across the same space. The particular TPACF equation that should be implemented in the application is

$$\omega(\theta) = \frac{n_R \cdot DD(\theta) - 2\sum_{i=0}^{n_R-1} DR_i(\theta)}{\sum_{i=0}^{n_R-1} RR_i(\theta)} + 1$$

where DD(θ) is the frequency distribution between the points in the observed data, DR(θ) is the frequency distribution between the points in the observed data and the random data, and RR(θ) is the frequency distribution between the points in the random data, and n_R is the number of random datasets used in the calculations. In this case, we have explicitly assumed that the number of data and random points in each sample are the same.

Determining DD(θ), DR(θ), and RR(θ) requires computing separation distances and binning them into separation distributions at some angular resolution Δθ. There are two binning schemas implemented: direct computation of the bin number in the angular space, and binary search in the dot product space.

The overall application organization is as follows:

```
load observed data dataset
compute DD(θ)
for i from 1 to n_R
        load random data dataset
        compute RR_i(θ)
        compute DR_i(θ)
compute ω(θ)
```

compute {$DD(\theta)$ | $RR_i(\theta)$ | $DR_i(\theta)$} is implemented using a brute-force technique in which distances between all pairs of points are computed and mapped into the corresponding bin.

Objective

We expect the students to implement the computational kernel of the application (compute {$DD(\theta)$ | $RR_i(\theta)$ | $DR_i(\theta)$} subroutine) on the CUDA/G80 platform. We expect the students to achieve the highest possible performance the G80 platform can offer for this application.

Background

Basic Math

Resources

https://netfiles.uiuc.edu/kindrtnk/www/2pacf/

An easy-to-read introduction about how to compute TPACF in practice
https://netfiles.uiuc.edu/kindrtnk/www/2pacf/2point4Vlad.pdf

A more detailed description of the technique implemented in the provided source code
https://netfiles.uiuc.edu/kindrtnk/www/2pacf/2pacf.doc

The source code and few datasets necessary to get started
https://netfiles.uiuc.edu/kindrtnk/www/2pacf/2pacf.tgz

Contact Information

Volodymyr (Vlad) Kindratenko
kindr@ncsa.uiuc.edu
265-0209