

-sometitle-

Classifying N-body simulations with and without relativistic corrections
using machine learning techniques

Johan Mylius Kroken

Computational Science: Astrophysics
60 ECTS study points

Institute of Theoretical Astrophysics
Faculty of Mathematics and Natural Sciences

Johan Mylius Kroken

-sometitle-

Classifying N-body simulations with and without
relativistic corrections using machine learning
techniques

Supervisors:

A David Fonseca Mota

B Julian Adamek

C Francisco Antonio Villaescusa Navarro

Abstract

On large scales, comparable to the horizon, relativistic effects will affect the cosmological observables. In order to solve for these effects, one need to consistently solve for the metric, velocities and densities in a particular gauge. When simulating large-scale structures we use N-body simulations, which are usually performed in the Newtonian limit. However, it is not obvious that Newtonian gravity yield a good global description of an inhomogeneous cosmology across all scales (Jeong, Schmidt and Hirata 2012). However, literature suggest that Newtonian simulations are still solving the dynamics correctly, even on large scales close to the horizon where relativistic effects are important but may be corrected for (Chisari and Zaldarriaga 2011) (Green and Wald 2012).

Recently, Adamek et al. 2016 developed a relativistic N-body code, **gevolution**, which evolves large scales structures based on the weak field expansion in GR. I investigate the differences in the gravitational dynamics between structures evolved with and without relativistic effects, with focus on the gravitational potential Φ . This is a good choice for comparison as Φ is gauge invariant and the Newtonian and relativistic simulations are performed in different gauges.

The investigation is done by running 2000 simulations using identical Λ CDM cosmologies for the two gravity theories. The simulations are run using 64^4 particles on a 256^3 grid each with dimension 5120 Mpc/h which is a compromise in order to include both large and nonlinear scales. The data analysis consist of a preliminary analysis using conventional summary statistics, with focus on the bispectrum of Φ . There is a difference in the two cases for low redshifts in the equilateral and squeezed configurations. However, the main idea is to train a Convolutional Neural Network (CNN) to classify the two cases, given snapshots of Φ . The main analysis then involves interpretability of the CNN, which may be done by considering for instance saliency maps (Alqaraawi et al. 2020) or Grad-CAM (Selvaraju et al. 2020). In either case, revealing the features separating the two cases may help us understand the differences in the gravitational dynamics between the two theories. I expect that such a network is able to find relativistic corrections to the Newtonian snapshots that are of higher order than those obtained from power spectra and bispectra analysis. Further, it may also reveal which configurations of Fourier modes \mathbf{k} yield the highest bispectral power, which for now is mainly trial and error.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Outline	1
1.3	Aim	1
1.4	Nomenclature	1
I	Cosmological Structure Formation	3
2	Preliminaries	5
2.1	General Relativity	5
2.1.1	Einstein's Field Equations	5
2.1.2	Riemann Connection and Covariant Derivatives	5
2.1.3	Geodesic Equation	5
2.1.4	The Stress-Energy Tensor	5
2.2	Useful Relations	5
3	Background Cosmology	7
3.1	The homogeneous Universe	7
3.1.1	The Cosmological Principle	7
3.1.2	The Robertson-Walker Metric	7
3.1.3	The Friedmann Equations	7
3.2	My Universe is loaded with...	7
3.3	Thermal History of the Universe	7
4	Perturbation Theory	9
4.1	Initial Conditions	9
4.2	Transfer Functions	9
4.3	Power Spectra	9
4.4	Linear Evolution	9
4.5	Non-linear Evolution	9
4.6	Bispectra	9
4.6.1	Analytical Bispectrum	9
5	Simulation theory	13
5.1	N-body simulations	13
5.1.1	Describing a box of particles	13
5.1.2	Forces and Fields	13
5.1.3	Mass Assignment Schemes	13
5.1.4	Validity of Box	13
5.2	Newtonian Approach	13
5.3	General Relativistic Approach	13

II	Machine Learning	15
6	Fundamental Elements of Machine Learning	17
6.1	Introduction	17
6.2	Linear Algebra	17
6.3	Probability and Information Theory	17
6.4	Basic Machine Learning	17
6.4.1	Optimisation and Generalisation	17
6.4.2	Data and Fitting	18
6.4.3	Estimators, Bias, Variance and Error	18
6.4.4	Maximum Likelihood Estimation	18
6.4.5	Bayesian Statistics	18
6.4.6	Supervised Learning	18
6.4.7	Unsupervised Learning	18
7	Neural Networks	19
7.1	Forward pass - Prediction	19
7.1.1	Activation functions	19
7.1.2	Loss functions	19
7.2	Backpropagation - Training	19
7.2.1	Chain rule	19
7.2.2	Gradient descent	19
7.2.3	Optimizers	19
7.2.4	Regularization	19
8	Convolutional Neural Networks	21
8.1	Convolution	21
8.2	New Layers	21
8.2.1	Convolutional layers	21
8.2.2	Pooling layers	21
8.2.3	Dropout layers	21
III	Acquiring Data	23
9	Simulations	25
9.1	Parameters	25
9.1.1	Cosmological parameters	25
9.1.2	Primordial power spectrum	25
9.1.3	Box parameters	26
9.1.4	Seeds and gravity theories	26
9.2	Output	26
9.2.1	Datacubes	26
9.2.2	Redshift problem	27
9.2.3	Statistics of dataset	27
9.2.4	Normalisation	27

10	Data Verification	29
10.1	Slices of Datacubes	29
10.2	Powerspectra from Simulations	29
10.3	Powerspectra from Datacubes	29
10.4	Analytical Bispectra	29
10.5	Bispectra from Cube	29
10.5.1	Binning	29
10.5.2	Bispectra	29
11	Trainable Dataset	37
11.1	Gathering the cubes	37
11.2	Find image from index	37

Contents

List of Figures

4.1	Angles in arbitrary bispectrum triangle configuration where $\sum_i \mathbf{k}_i = 0$	10
10.1	Slice 0	30
10.2	Average matter power spectra at different redshifts.	31
10.3	Average potential power spectra at different redshifts.	32
11.1	Outline of data structure.	38

List of Figures

List of Tables

9.1	Cosmological parameters	25
9.2	Primordial power spectra parameters	25
9.3	Box parameters	26

List of Tables

Preface

Here comes your preface, including acknowledgments and thanks.

Part II

Machine Learning

Chapter 6

Fundamental Elements of Machine Learning

6.1 Introduction

In this chapter I will give a brief introduction into machine learning. This includes a mathematical description of some fundamental concepts common across numerous machine learning models. The more advanced models will be dealt with at a later stage. If not otherwise stated, the following chapter is based on Goodfellow, Bengio and Courville 2016 and Hastie, Tibshirani and Friedman 2009.

6.2 Linear Algebra

maybe

6.3 Probability and Information Theory

maybe

6.4 Basic Machine Learning

TODO: [Fill more here](#)

6.4.1 Optimisation and Generalisation

Optimisation Optimisation problems are problems in which we want to minimise some error, given some data. In other words, we want to optimise an algorithm or model given a specific dataset. We care about the error of our model for that specific dataset only.

Generalisation The concept of *generalisation* is what makes a machine learning model different from an optimisation model. We still *train* the machine learning model on some specific data, but we measure how good the model is based on how it performs on a different set of data, which it was not trained on. I.e. we need a generalised model, which is not restricted to the data it was trained on.

6.4.2 Data and Fitting

The key ingredient to any machine learning algorithm is the data fed into it.

6.4.3 Estimators, Bias, Variance and Error

Estimators Based on the assumption that there exists some true parameter(s) θ which remain unknown,¹ we are able to make predictions and estimations of such parameter(s). Let's say we have m independent and identically distributed (i.i.d.) random variables $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$ drawn from the same probability distribution $p(\mathbf{x})$. An *estimator* of the true values θ is any function of the data such that $\hat{\theta}_m = g(\mathbf{x}_1, \dots, \mathbf{x}_m)$, where $\hat{\theta}$ is the estimate of θ . This is known as point estimation, as we are estimating a single value. This definition does not pose any restrictions on the function g . However, a good estimator would yield an estimate $\hat{\theta}_m$ that is close to the true value θ .

Function estimators Say we want to predict a variable \mathbf{y} given some vector \mathbf{x} . We assume the true variable \mathbf{y} is given by some function approximation $f(\mathbf{x})$ plus some error ϵ : $\mathbf{y} = f(\mathbf{x}) + \epsilon$. The aim is then to estimate the function f with the estimator \hat{f} . If we then realise that \hat{f} is really just a point estimator in function space, the two above concepts are equivalent.

Bias The bias of the estimator $\hat{\theta}_m$ is defined as the difference between the expected value of the estimator and the true value of the parameter: $\text{bias}(\hat{\theta}_m) = \mathbb{E}[\hat{\theta}_m] - \theta$. An unbiased estimator has zero bias, i.e. $\mathbb{E}[\hat{\theta}_m] = \theta$. An estimator is asymptotically unbiased if its bias approaches zero as the number of data points m approaches infinity, i.e. $\lim_{m \rightarrow \infty} \mathbb{E}[\hat{\theta}_m] = \theta$.

Variance

Standard Error

Mean Squared Error

6.4.4 Maximum Likelihood Estimation

6.4.5 Bayesian Statistics

6.4.6 Supervised Learning

6.4.7 Unsupervised Learning

¹This is the frequentist perspective of statistics

Chapter 7

Neural Networks

7.1 Forward pass - Prediction

7.1.1 Activation functions

7.1.2 Loss functions

7.2 Backpropagation - Training

7.2.1 Chain rule

Basics Lets say we have a function $f(x)$ that is composed of two functions $g(x)$ and $h(x)$, such that $f(x) = g(h(x))$. The chain rule states that the derivative of $f(x)$ with respect to x is the product of the derivative of $g(x)$ with respect to $h(x)$ and the derivative of $h(x)$ with respect to x :

$$\frac{df}{dx} = \frac{dg}{dh} \frac{dh}{dx}. \quad (7.1)$$

This can be generalised into the vector case where $f(\mathbf{x}) = g(h(\mathbf{x}))$ and $\mathbf{x} \in \mathbb{R}^n$ and [TODO: write more ahaha](#)

7.2.2 Gradient descent

7.2.3 Optimizers

7.2.4 Regularization

Chapter 8

Convolutional Neural Networks

8.1 Convolution

8.2 New Layers

8.2.1 Convolutional layers

8.2.2 Pooling layers

8.2.3 Dropout layers

Bibliography

- Adamek, Julian et al. (29th July 2016). ‘gevolution: a cosmological N-body code based on General Relativity’. In: *Journal of Cosmology and Astroparticle Physics* 2016.7, pp. 053–053. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2016/07/053. arXiv: 1604.06065[astro-ph,physics:gr-qc,physics:physics]. URL: <http://arxiv.org/abs/1604.06065> (visited on 23/08/2023).
- Alqaraawi, Ahmed et al. (3rd Feb. 2020). *Evaluating Saliency Map Explanations for Convolutional Neural Networks: A User Study*. arXiv: 2002.00772[cs]. URL: <http://arxiv.org/abs/2002.00772> (visited on 23/10/2023).
- Chisari, Nora Elisa and Matias Zaldarriaga (2nd June 2011). ‘Connection between Newtonian simulations and general relativity’. In: *Physical Review D* 83.12, p. 123505. ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.83.123505. arXiv: 1101.3555[astro-ph,physics:gr-qc]. URL: <http://arxiv.org/abs/1101.3555> (visited on 23/08/2023).
- Goodfellow, Ian, Yoshua Bengio and Aaron Courville (2016). *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press.
- Green, Stephen R. and Robert M. Wald (15th Mar. 2012). ‘Newtonian and Relativistic Cosmologies’. In: *Physical Review D* 85.6, p. 063512. ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.85.063512. arXiv: 1111.2997[astro-ph,physics:gr-qc]. URL: <http://arxiv.org/abs/1111.2997> (visited on 23/08/2023).
- Hastie, T., R. Tibshirani and J.H. Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer series in statistics. Springer. ISBN: 9780387848846. URL: <https://books.google.no/books?id=eBSgoAEACAAJ>.
- Jeong, Donghui, Fabian Schmidt and Christopher M. Hirata (4th Jan. 2012). ‘Large-scale clustering of galaxies in general relativity’. In: *Physical Review D* 85.2, p. 023504. ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.85.023504. arXiv: 1107.5427[astro-ph]. URL: <http://arxiv.org/abs/1107.5427> (visited on 23/08/2023).
- Selvaraju, Ramprasaath R. et al. (Feb. 2020). ‘Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization’. In: *International Journal of Computer Vision* 128.2, pp. 336–359. ISSN: 0920-5691, 1573-1405. DOI: 10.1007/s11263-019-01228-7. arXiv: 1610.02391[cs]. URL: <http://arxiv.org/abs/1610.02391> (visited on 23/10/2023).