# Department of Physics and Astronomy

## University of Heidelberg

Master Thesis

in Physics

submitted by

Jamie Alexander Grieser

born in Illertissen

2021

# Neural Potentials

# -

# Exploring Astrophysics and Cosmology

# with Machine Learning

This Master Thesis has been carried out by

Jamie Alexander Grieser

at the

Institute for Theoretical Astrophysics

at

Heidelberg University

under the supervision of

Björn Malte Schäfer

The purpose of this thesis is to introduce a novel approach to analyze physical systems using machine learning. In particular, we combine the relatively new method of neural ordinary differential equations with prior knowledge about the systems to implement a new machine learning method called neural potential.

The method enables us to learn quantities like the potential or equation of state from experimental data. This helps to make predictions for similar experiments and may gives hints about unknown underlying physical laws.

The concept of neural potentials is introduced by learning the potential of a harmonic oscillator. We continue by analyzing the motions of the star $S2$ around Sagittarius $A^*$ and try to learn Newton's gravitational law and its relativistic corrections. Then, we turn to the problem of dark energy and first attempt to learn a dynamic equation of state of the dark energy component which may be used to select from the vast amount of competing dark energy models. Finally, the quintessence model is analyzed to demonstrate how neural potentials can help investigating dark energy models by learning the potential from the observation of supernovae of type Ia.

Das Ziel dieser Arbeit ist die Entwicklung einer neuartigen Methode zur Analyse physikalischer Systeme, die auf Techniken des maschinellen Lernens beruht. Um dies umzusetzen wird die erst kürzlich entwickelte Methodik der neuronalen gewöhnlichen Differentialgleichungen mit physikalischem Vorwissen über das betrachtete System kombiniert. Die Methode erlaubt es bestimmte Größen eines physikalischen Systems, wie zum Beispiel das Potential oder eine thermodynamische Zustandsgleichung mit Hilfe experimenteller Beobachtungen zu lernen.

Dies ermöglicht es, Vorhersagen über ähnliche Experimente zu treffen und möglicherweise neue Hinweise auf zugrundeliegende physikalische Gesetzmäßigkeiten zu erhalten. Das Konzept der neuronalen Potentiale wird anhand des harmonischen Oszillators demonstriert. Danach werden die Orbits des Sterns $S2$ um Sagittarius $A^*$ analysiert und es wird versucht, Newton's Gravitationsgesetz und etwaige relativistische Korrekturterme von den Daten zu erlernen.

Im Anschluss wenden wir uns dem Problem der dunklen Energie zu und versuchen die Zustandsgleichung eines Dunkle-Energie-Fluids zu lernen. Diese Zustandsgleichung kann dann benutzt werden, um aus einer Vielzahl existierender Modelle für dunkle Energie das Passende auszuwählen. Zum Schluss wird das Quintessenzmodell analysiert und es wird demonstriert, wie neuronale Potentiale benutzt werden können um Modelle für Dunkle Energie zu untersuchen. Insbesondere wird untersucht, wie Potentiale für Skalarfelder aus Supernova-Ia-Daten erlernt werden können.

*In memory of Gerrit Scholzen, friend, best man and fellow physicist,*
*who left this plane of existence way too soon.*

# Contents

# Introduction

Machine learning and artificial intelligence are undoubtly the most important emerging technologies of the decade. Their ability to capture hidden structures and systematics in datasets, which would otherwise be missed even by experts are unmatched. They allowed for the development of a variety of new technologies as well as to accelerate scientific progress in many research areas.

However, their application to problems in physics rarely goes beyond the use for data analysis. The reason for this is that most machine learning models are not able to capture and represent the intricate priors that physics puts upon our experiments. Consider for example a simple swinging pendulum. We can measure its position and velocity at each time-step and use it to train a recurrent neural network.

It turns out that this model generalizes poorly to times beyond the time interval of the training data (see figure 0.1). This can be explained with the fact that the motion of a pendulum follows a simple law which is described by a differential equation. The differential equation contains all of the physical knowledge about the system and thus tightly constrains the possible trajectories.

A recurrent neural network is not able to capture these constraints due to its mathematical structure. Thus, to use machine learning as an explorative tool in physics that is able to yield new insights and laws by analyzing experimental data, it is imperative to have a model that is able to work with differential equations. The method of choice in this thesis is called neural ordinary differential equations. They are applied to various physical and astrophysical problems, like the aforementioned pendulum, the celestial two-body problem and the dark energy problem. The hope is to demonstrate the power of this approach and how it might help to advance our knowledge about existing physical laws and maybe even give hints towards new ones.

The thesis starts by giving an overview over neural differential equations and how they are related to neural networks. After that, we will present the necessary background to understand the problems that will be investigated. Then, the different examples will be discussed, each becoming subsequently harder to solve. Starting with the simple example of a pendulum it will be explained how to transfer a given problem into the language of neural ordinary differential equations.

After that, the Sagittarius $A^*$ dataset is analyzed to check if we are able to reconstruct Newtons gravitational law and its relativistic corrections from the motions of the $S$-stars around the galaxy's central black hole. Finally, the dark energy problem is discussed in the light of scalar fields and quintessence. We start by investigating the properties of a possible dynamic dark energy fluid by learning its equation of state from redshift-luminosity data. Then, we proceed by replacing the fluid with a quintessence field and try to reconstruct its potential using the same data. The thesis concludes with a discussion of possible improvements of the method as well as future prospects and ideas for the analysis of other interesting systems.
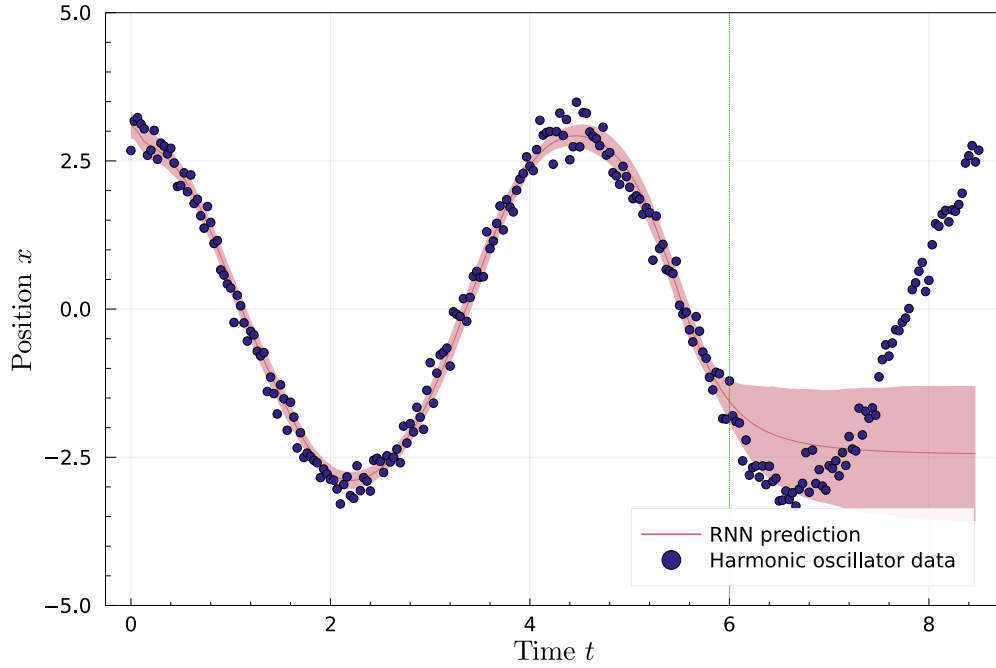
Figure 0.1: This figure shows a the output of a recurrent neural network that has been trained on a time-series dataset generated by a harmonic oscillator. The data in the interval $[0.0, 6.0]$ has been used for training. We see, that the model captures the data very well in this interval and with high confidence. However, when generalizing to the interval $(6.0, 8.5]$, the model fails to reproduce the oscillatory behavior and the uncertainty grows very fast. This is because the recurrent neural network is not able to model priors like energy conservation. The confidence intervals were constructed using an ensemble of 512 trained models. The model uses 4 dense layers with 16 nodes each and an input and output layer of 5 nodes. Using 5 input and 5 output layers, the model can process 5 consecutive time-steps and is able to predict the next 5 steps. It was trained using the ADAM optimizer with a learning rate of 0.001 for 10000 epochs and a mean squared error cost function.

# 1 Theoretical Foundations

This chapter aims to provide an overview over all the topics relevant for this thesis. It starts by introducing neural ordinary differential equations as a method to analyze experimental data and extract information about the responsible physical processes.

We continue by describing how the model is implemented and fitted to the data. Additionally, we elaborate on the construction of confidence intervals using bootstrapping methods. After this short introduction, different physical experiments and datasets are discussed, especially in the light of their governing differential equations.

Beginning with the simple example of a harmonic oscillator, the models become increasingly more advanced. The Kepler problem and the corresponding dataset, i.e. the $S$-stars orbiting Sagittarius $A^*$ are discussed. Proceeding to the area of cosmology, the dark energy problem and dynamic equations of state are presented. Finally, quintessence is described as a simple model for dark energy.

## 1.1 Differential Equations and Machine Learning

Differential equations are the language of modern physics. They are able to quantify physical priors and constraints like feedbacks, damping effects and couplings between different observables. If we want to apply machine learning techniques in an explorative manner such that we are able to gain new insights on unsolved physics problems, it is necessary to include those priors.

Most of the time, this prior knowledge is expressed through differential equations. The method has to be able to reproduce this, but also has to provide a learnable and adaptable structure that can be trained to reflect physical laws. But what exactly do we learn from the data? And what structure is used to model it?

The dynamics of a problem are often described through different orders of spatial and temporal derivatives, for example whether or not a system is accelerated. The physics on the other hand, that is particular laws that separate one experiment from another, are described through potentials. As an example, take the difference between throwing a ball and a swinging pendulum. For both problems, Newton's second law is used to model the dynamics, that is acceleration through some force, but the physical constraints on the ball and the pendulum are completely different.

Those constraints are reflected by different shapes of the potential and the use of different coordinates. This is especially true for energy-conserving systems where there are no dissipative forces [Bartelmann, 2015]. Therefore if we want to explore the physics of a system from experimental data it makes sense to start by modeling the potential. To make the approach as flexible as possible, the potential or its gradient is represented by a neural network.

Using common training techniques like backpropagation and minimization of a cost function, the neural network is adjusted such that it approximates the potential of the problem. However, potentials are not the only possible part of a differential equation that can be replaced by a neural network. Other possibilities include, but are not limited to the equation of state of a fluid, contributions from friction terms and even entire Lagrange functions [Cranmer et al., 2020, Lutter et al., 2019].

The key method to combine the power of neural networks with the concept of potentials and differential equations is called *neural ordinary differential equations*. As the name already suggests, this approach is applicable only to ordinary differential equations.

### 1.1.1 Neural Differential Equations

Neural ordinary differential equations are a very recent development in the area of machine learning and artificial intelligence [Chen et al., 2018]. They are the continuous generalization of residual neural networks, which use a hidden state $\mathbf{h}_t$ to store information about the current state of the network and its inputs to be able to reuse it later within the network. Residual neural networks are equivalent to recurrent neural networks that are often used in time-series analysis, is also the problem we are trying to solve [Liao and Poggio, 2016]. Both can be written as a sequence of discrete transformations that is applied to the hidden state:

$$\mathbf{h}_{t+1} = \mathbf{h}_t + f(\mathbf{h}_t, \theta_t) \tag{1.1}$$

This equation is similar to the Euler discretization of an ordinary differential equation (ODE) of the form

$$\frac{\mathrm{d}\mathbf{h}(t)}{\mathrm{d}t} = f(\mathbf{h}(t), \theta, t). \tag{1.2}$$

If we take $\mathbf{h}(0)$ to be the input layer, then we can compute the output layer $\mathbf{h}(T)$ by solving of the initial value problem (1.2) using a numerical differential equation solver. By using an adaptive solver, the model evaluates the hidden dynamics $f$ where necessary by itself with the desired accuracy. In particular, this allows the model evaluation to scale with the problem complexity. When training a machine learning model, it is necessary to have an algorithm that allows to adjust the parameters with respect to some cost function. The method of choice for this is usually backpropagation [Linnainmaa, 1970].

For a differential equation it is possible to back-propagate through the operations of the solver, but then memory cost and evaluation time scale directly with the number of model evaluations and it is necessary to know the internal workings of the ODE solver. Another method to compute the gradients is called the *adjoint sensitivity method* [Pontryagin, 1962]. This method defines a second, augmented ODE that is solved backwards in time and gives the gradient of the cost function and model with respect to the hidden state and parameters.

2

This approach has multiple advantages:

- The ODE solver is treated as a black box, i.e. internal workings are no longer relevant.

- Forward pass and backward pass can be done using the same ODE solver.

- It scales linearly with the problem size and has low memory cost.

- The numerical error is controlled explicitly.

- Initial conditions can also be included as trainable parameters.

## Adjoint Sensitivity or How to Calculate Gradients

The forward pass of the model consists of plugging the solution of the ODE into the cost function $\mathcal{L}$:

$$\mathcal{L}(\mathbf{h}(T)) = \mathcal{L}\left(\mathbf{h}(0) + \int_0^T f(\mathbf{h}(t), \theta, t)\mathrm{d}t\right). \tag{1.3}$$

To construct the augmented ODE, we introduce the *adjoint* $\mathbf{a}(t) = \frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\mathbf{h}(t)}$ that describes how the cost function changes with the hidden state. Differentiating (1.3) with respect to the time $t$ and rewriting the terms as derivatives with respect to the hidden state $\mathbf{h}$ yields the ODE for the adjoint:

$$\frac{\mathrm{d}\mathbf{a}(t)}{\mathrm{d}t} = -\mathbf{a}(t)^\top \frac{\partial f(\mathbf{h}(t), \theta, t)}{\partial \mathbf{h}}. \tag{1.4}$$

The interesting quantity is $\mathbf{a}(0) = \frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\mathbf{h}(0)}$, which is equivalent to gradient at the input layer after back-propagating through all intermediate layers. To compute this, (1.3) has to be solved backwards with the initial value $\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\mathbf{h}(T)}$. However, to solve this ODE the values of $\mathbf{h}(t)$ needs to be known at all intermediate evaluation times.

They can be obtained by solving (1.2) backwards in time with the initial value $\mathbf{h}(T)$ that is known from the forward pass. Last but not least, we need to include the gradients of the cost function with respect to the model parameters, that is $\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\theta}$. Deriving (1.3) with respect to the parameters $\theta$ gives the corresponding differential equation:

$$\mathbf{g}(t) = \frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\theta} = -\int_T^0 \mathbf{a}(t)^\top \frac{\partial f(\mathbf{h}(t), \theta, t)}{\partial \theta}\mathrm{d}t. \tag{1.5}$$

Taking the three quantities together allows us to construct an augmented ODE with the state vector $(\mathbf{h}(t), \mathbf{a}(t), \mathbf{g}(t))$:

$$
\begin{aligned}
\frac{\mathrm{d}\mathbf{h}}{\mathrm{d}t} &= f(\mathbf{h}(t), \theta, t) \\
\frac{\mathrm{d}\mathbf{a}}{\mathrm{d}t} &= -\mathbf{a}(t)^\top \frac{\partial f(\mathbf{h}(t), \theta, t)}{\partial \mathbf{h}} \\
\frac{\mathrm{d}\mathbf{g}}{\mathrm{d}t} &= -\mathbf{a}(t)^\top \frac{\partial f(\mathbf{h}(t), \theta, t)}{\partial \theta}
\end{aligned}
\tag{1.6}
$$

The solution of the initial value problem with the initial conditions $(\mathbf{h}(0), \frac{d\mathcal{L}}{d\mathbf{h}(T)}, \mathbf{0})$ then gives the gradients $\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\mathbf{h}(0)}$ and $\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\theta}$, that in turn can be used to adjust the model parameters accordingly. It should be mentioned that this approach of constructing an adjoint for gradient computations works only for ordinary differential equations. The most pressing question now is how to choose $f$.

## 1.1.2 Neural Networks

In principle, $f$ can be chosen to be any differentiable function. Most of the time, $f$ will be chosen to be a neural network of some sort, for example a fully connected network or a convolutional neural network or a combinations of these. Doing so constructs a continuous-depth residual/recurrent neural network where the ODE solver determines how many residual blocks it needs for the training epoch. This is also what makes neural ordinary differential equations "neural".

Here, we present a hybrid approach where $f$ contains parts that are closed mathematical expressions while others are realized through neural networks. The idea is to see whether these parts can be learned from the data. They are constructed with a network of fully connected dense layers with corresponding non-linear activation functions like *ReLU* or *tanh*. Neural networks are used because they offer great flexibility due to Kolmogorovs theorem, which is also sometimes referred to as *Universal Approximation Theorem* [Kurkova, 1992]. It states that a neural network can approximate any function with arbitrary precision if it satisfies the following conditions:

1. It is least two layers deep.

2. The activation functions is non-linear.

3. It has to have enough nodes per layer.

The number of nodes per layer will of course scale with the complexity of the function in question. As the problems that will be analyzed are comparatively easy, two layers with up to 32 nodes will suffice most of the time. For more information on neural networks, see Schmidhuber [2014], Emmert-Streib et al. [2020].

### 1.1.3 Machine Learning Techniques

**Cost Functions**

The task we try to solve is essentially a regression problem. In this case, one commonly uses the $L^2$-norm, that is the mean squared error (MSE) as a cost function [Zhang et al., 2019b]. Assuming a model makes predictions $\mathbf{h}(t_i)$ about some process that has produced observations $O_i$, the MSE cost function reads

$$\mathcal{L}(\mathbf{h}) = \sum_{i \in I} (\mathbf{h}(t_i) - O_i)^2. \tag{1.7}$$

This cost function penalizes large errors quadratically. Thus, it is sensitive towards outliers and given several examples with the same input feature values, the optimal prediction will be their mean target value . This can be compared to the mean absolute error, i.e. the $L^1$-norm, where the optimal prediction will be their median.

A generalization of this cost function is the $\chi^2$-statistic, that also takes into account the measurement error $\sigma_i$ of the data [Cochran, 1952, Rinne, 2008]:

$$\chi^2(\mathbf{h}) = \sum_{i \in I} \frac{(\mathbf{h}(t_i) - O_i)^2}{\sigma_i^2} \tag{1.8}$$

With the $\chi^2$-statistic, it is possible to give a weight to the different observations based on their measurement accuracy. More precise measurements will have stronger influence on the model. This approach assumes that the error of the measurements follows a Gaussian distribution. For some experiments this might not be the case. Then, one can fall back to the MSE cost function (1.7) or use the $\chi^2$-statistic as an approximation, but the interpretation of weighing the measurements by their errors looses its meaning.

**Optimizers**

When using backpropagation to adjust the parameters of a model, it has become a custom to use a gradient descent based optimization algorithms [Cauchy, 1847]. Today, there exist many different realizations with distinct features and applications. Although the datasets discussed in this thesis come from different sources they have one common property, that is all of them are time-series. This suggests the use of optimization algorithms from time-series analysis, especially when performed with recurrent neural networks. After all, a neural ordinary differential equation is nothing more than a continuous-depth recurrent neural network.

Some of the popular optimization algorithms of time-series analysis, which also have been empirically proven to perform very well, are ADAM and its Nesterov momentum version NADAM [Kingma and Ba, 2015]. They use concepts like a moving exponential average of past gradients and per-parameter learning rates to increase convergence speed. Both employ the second moments to adapt the learning rate, but NADAM additionally incorporates the current "velocity" of the optimizer to increase precision. For strongly oscillating gradients, NADAM is the better choice, but in all other cases ADAM converges faster.

## Julia and Automatic Differentiation

All of the numerical experiments in this thesis have been implemented using the fairly new programming language *Julia* [Bezanson et al., 2012]. It is said to combine the speed of *C* with the syntax of *Python*. The language has been specifically designed for scientific computing and offers a variety of packages for machine learning applications. This includes the packages *Zygote*, *Flux* and *DiffEqFlux* [Innes, 2019].

*Zygote* is a package that allows the computation of derivatives using a method called *automatic differentiation* [Linnainmaa, 1976, Baydin et al., 2017]. This allows the computation of gradients without the use of finite-differencing methods and thus reduces numerical error and increases computational speed. This allows to compute derivatives of very complex expressions with high numerical precision and even differentiate if-statements and loops. It should be mentioned however, that it is cumbersome to find higher order derivatives using this approach.

*Flux* is essentially the Julia equivalent of Python packages like Tensorflow and Pytorch. It provides functionalities for neural network construction, optimization routines and data loading to build deep learning applications. Those functionalities are built around the automatic differentiation features of Zygote.

*DiffEqFlux* provides an interface to construct neural ordinary differential equations by using Julia's ODE solvers to build the adjoint sensitivity method and connect it with the Flux package. With these three packages, it is possible to build up the entire pipeline of a neural ordinary differential equation as it will be proposed in this thesis.

## Bootstrapping and Confidence Intervals

Bootstrapping is a resampling technique that allows to generate new samples from a population inferred from experimental data [Efron and Tibshirani, 1986]. The new samples are then used to investigate and estimate certain statistics, for example mean and variance. Assume that we have a data set of $n$ points $x_i$ with standard errors $\sigma_i$. For many experiments, it is reasonable to assume that for each measurement the error follows a Gaussian distribution. Thus, to resample from the dataset, we resample at each point $i$ from a Gaussian with mean $x_i$ and variance $\sigma_i^2$.

The model is then fitted to all datasets and gives a sample of different possible model parameters. This sample allows then to estimate the mean of the model parameters, which is often a good estimate for the best fit model parameters. It is also possible to construct a crude notion of a confidence interval. For a large enough sample, that is several hundered or thousand resamplings, quantiles provide a simple way of constructing confidence bounds. Taking the 0.025 and 0.975-quantiles of the parameters and fitting curves gives a notion of a $2\sigma$-confidence interval on those quantities. This is called the percentile confidence interval.

## 1.2 Periodic Motions - The Harmonic Oscillator

This section is a short review of the harmonic oscillator. The simplest version of a harmonic oscillator is a mass-spring system where the force generated by the spring is proportional to the elongation. This is known as Hooke's law and is only true for small excitations. Using Newton's second law gives us a simple ordinary differential equation, which reads

$$m\ddot{x} = -kx \Rightarrow \ddot{x} + \omega^2 x = 0 \text{ where } \omega = \sqrt{\frac{k}{m}} \tag{1.9}$$

where $k$ is called the spring constant. This system is analytically solvable in many ways, and the solutions are sinusoidal motions of frequency $\omega$. For an energy-conserving system like the mass-spring system, the force exerted may be written as the gradient of a potential, i.e.

$$m\ddot{x} = -\frac{\mathrm{d}V}{\mathrm{d}x}. \tag{1.10}$$

In the case of a simple harmonic oscillator we have a potential $V(x) = \frac{k}{2}x^2$. Quadratic potentials are the easiest way to realize harmonic oscillations, but are only an approximation to nature. By employing more complicated potentials, it is possible to account for non-linear behavior of the force exerted by the oscillator. However, often it is no longer possible to solve such systems analytically and one has to fall back to numerical methods. One example for this is a pendulum in a gravitational field. The differential equation for the deflection angle $x$ of the mass $m$ can be written as

$$ml\ddot{x} = -mg\sin(x). \tag{1.11}$$

In this case, the potential reads $V(x) = -mg\cos(x)$. Here the trajectories are no longer sinusoidal, but they are still periodic. For small angles $x$, we can use the approximation $\sin(x) \approx x$ which gives us back the linear behavior of the force, similar to the mass-spring system. In the same manner, we can see that for small $x$, the potential will become quadratic again. It is possible to further increase the complexity of the harmonic oscillator by adding damping terms, for example a term proportional to $\dot{x}$. Then, the motion of the mass will be an exponentially damped sinusoid as can be seen by solving the differential equation

$$m\ddot{x} + 2\gamma\dot{x} + \omega^2 x = 0. \tag{1.12}$$

The term proportional to the velocity implements friction and thus energy is no longer conserved and the formulation using a potential breaks down. We will see that the presented approach does often rely on the existence of a potential and thus we will not discuss this case any further.

## 1.3 Celestial Mechanics - The Kepler Problem

Solving the problem of celestial motions is probably one of the biggest achievements in human history as it allowed mankind to use the universe itself as a laboratory. In its simplest form, the Kepler problem consists of two bodies orbiting each other due to attraction by gravitational forces. Usually, one transforms into the barycentric coordinate system and introduces the reduced mass $\frac{1}{\mu} = \frac{1}{m} + \frac{1}{M}$ to acquire an equation of motion for the distance vector $\mathbf{r}$ of the two bodies.

Thus, the problem is reduced to a single body moving in a potential. It is possible to show that the problem is inherently two-dimensional Bartelmann [2015]. For an arbitrary central potential $V(r)$ it is common to use polar coordinates. The Lagrange function of this problem has the following shape:

$$L = \frac{1}{2}\mu\dot{r}^2 + \frac{1}{2}\mu r^2\dot{\phi}^2 - V(r). \tag{1.13}$$

Using the Euler-Lagrange equations, we find that the angular momentum in the $z$-direction, $L_z = \mu r^2 \dot{\phi}$, is conserved and the equation of motion of the radial coordinate. Combining both equations gives the final differential equation containing an additional contribution from a centrifugal potential:

$$\mu\ddot{r} = \frac{L_z^2}{\mu r^3} - \frac{dV}{dr}. \tag{1.14}$$

It is important to mention that this is not a physical potential in the sense that there is no physical process behind it but is rather due to the choice of our coordinate system. A change of variables $u = \frac{1}{r}$ is performed and the trajectories are parametrized through the polar angle $\phi$. A result, the following differential equation is derived:

$$\frac{d^2u}{d\phi^2} = -u - \frac{\mu}{L_z^2}\frac{d}{du}V\left(\frac{1}{u}\right). \tag{1.15}$$

In the case of a simple Newtonian potential $V(r) = -\frac{k}{r}$ where $k = GMm$, the equation reduces to a simple harmonic oscillator driven by an additional constant force, that is

$$\frac{d^2u}{d\phi^2} + u = \frac{k\mu}{L_z^2}. \tag{1.16}$$

The solutions are the very well known conic sections $r(\phi) = \frac{L_z^2}{k\mu}\frac{1}{1+e\cos(\phi-\phi_0)}$. Binet's equation is the generalization of this problem to arbitrary central potentials [Binet, 1813]. For an in-depth treatment, the reader is referred to Bartelmann [2015].

Assuming that one body is significantly more massive than the other, $M \gg m$, closed orbits are either circles or ellipses where the heavier body is in one of the focal points. The trajectory itself is closed, meaning that the perihelion is fixed.

### 1.3.1 Relativistic Corrections

For post-Newtonian corrections from General Relativity, this changes significantly. The curvature of spacetime created by a body of mass $M$ is given through the Schwarzschild metric [Misner et al., 1973]:

$$\mathrm{d}s^2 = c^2\left(1 - \frac{2GM}{c^2 r}\right)\mathrm{d}t^2 - \frac{\mathrm{d}r^2}{1 - \dfrac{2GM}{c^2 r}} - r^2\left(\mathrm{d}\theta^2 + \sin^2(\theta)\mathrm{d}\phi^2\right) \tag{1.17}$$

With the metric $g_{\mu\nu}$ known, it is in principle possible to derive the equation of motion by computing the Christoffel symbols and use the geodesic equation. This is a tedious process and we can speed it up by realizing that geodesics are always the shortest paths between two points on the spacetime manifold. In Riemannian geometry, the arc length $A$ on some manifold for a curve $\gamma$ with starting point $\gamma(0)$ and end point $\gamma(1)$ is given through

$$A(\gamma) = \int_0^1 \sqrt{g_{\mu\nu}\frac{\mathrm{d}\gamma^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}\gamma^\nu}{\mathrm{d}\lambda}}\,\mathrm{d}\lambda. \tag{1.18}$$

Using Hamilton's principle of stationary action on this problem (where the arc length is the quantity to be minimized), we can apply the Euler-Lagrange equations to the integrand to find the geodesics. As the square root is a monotonic function and therefore does not change the position of the extrema, we can treat $L = g_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda}$ as our Lagrange function . The Euler-Lagrange functions for fields are given through

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{\partial L}{\partial \dot{x}^\mu}\right) - \left(\frac{\partial L}{\partial x^\mu}\right). \tag{1.19}$$

where the dot signs a derivative respective to the parametrization of the curve $\lambda$. On a four-dimensional spacetime manifold, we obtain four equations [Hobson et al., 2006]:

$$\left(1 - \frac{2GM}{c^2 r}\right)\dot{t} = K$$

$$\left(1 - \frac{2GM}{c^2 r}\right)^{-1}\ddot{r} + \frac{GM}{r^2}\dot{t}^2 - \left(1 - \frac{2GM}{c^2 r}\right)^{-2}\frac{GM}{c^2 r}\dot{r}^2 - r\left(\dot{\theta}^2 + \sin^2(\theta)\dot{\phi}^2\right) = 0 \tag{1.20}$$

$$\ddot{\theta} + \frac{2}{r}\dot{r}\dot{\theta} - \sin(\theta)\cos(\theta)\dot{\phi}^2 = 0$$

$$r^2\sin^2(\theta)\dot{\phi} = H$$

The third equation is solved by $\theta = \frac{\pi}{2}$ such that we can confine our analysis to the equatorial plane. We are only interested in the motion of massive particles, i.e. stars. Then, the constant $H = mr^2\dot{\phi}$ is the specific angular momentum of the particle while $K = \frac{E}{mc^2}$. To simplify the second equation, we use the fact that for any non-null geodesic, we have

$$g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu = c^2. \tag{1.21}$$

Orbits around black holes are time-like geodesics. Therefore, the trajectories can be parameterized through the proper time $\tau$. Compiling all these insights and plugging the first and fourth equation into the second equation gives [Hobson et al., 2006]:

$$\dot{r}^2 + \frac{H^2}{r^2}\left(1 - \frac{2GM}{c^2 r}\right) - \frac{2GM}{r} = c^2\left(K^2 - 1\right) \tag{1.22}$$

We change the parametrization from the proper time $\tau$ to the angle $\phi$ and again express the equation using the inverse radial coordinate $u = \frac{1}{r}$:

$$\left(\frac{\mathrm{d}u}{\mathrm{d}\phi}\right)^2 + u^2 = \frac{c^2}{H^2}\left(K^2 - 1\right) + \frac{2GM}{H^2}u + \frac{GM}{c^2}u^3 \tag{1.23}$$

Differentiating with respect to $\phi$ gives the equation

$$\frac{\mathrm{d}^2 u}{\mathrm{d}\phi^2} + u = \frac{GM}{H^2} + 3\frac{GM}{c^2}u^2. \tag{1.24}$$

This gives us an effective potential to model the contribution of General Relativity:

$$V(r) = -\frac{GMm}{r} - \frac{G(M + m)L_z^2}{c^2 \mu r^3} \tag{1.25}$$

The inverse cubic contribution is responsible for the precession of the periapsis while the other term is the well-known Newtonian potential. Therefore, we no longer have closed trajectories for non-circular orbits. Figure B.1 in the appendix demonstrates the different possible trajectories and potentials.

## 1.3.2 Rotated Orbits

When observing trajectories of objects orbiting a large central mass outside the solar system, as for example for stars orbiting Sagittarius $A^*$ [Eckart and Genzel, 1996, Ghez et al., 1998], it often happens that the plane of motion is not orthogonal to the line of sight. The position of the orbit with respect to some reference direction can be described using three angles:

- The inclination $i$ describes the tilt of the orbital plane against the reference plane.

- The longitude of the ascending node $\Omega$ describes the angle ascending node encloses with the reference direction.

- The argument of periapsis $\omega$ describes the angle between the ascending node and the periapsis.

Inverting these three rotations allows us to reconstruct the original trajectory. For a graphical explanation of the angles, see figure 1.1.
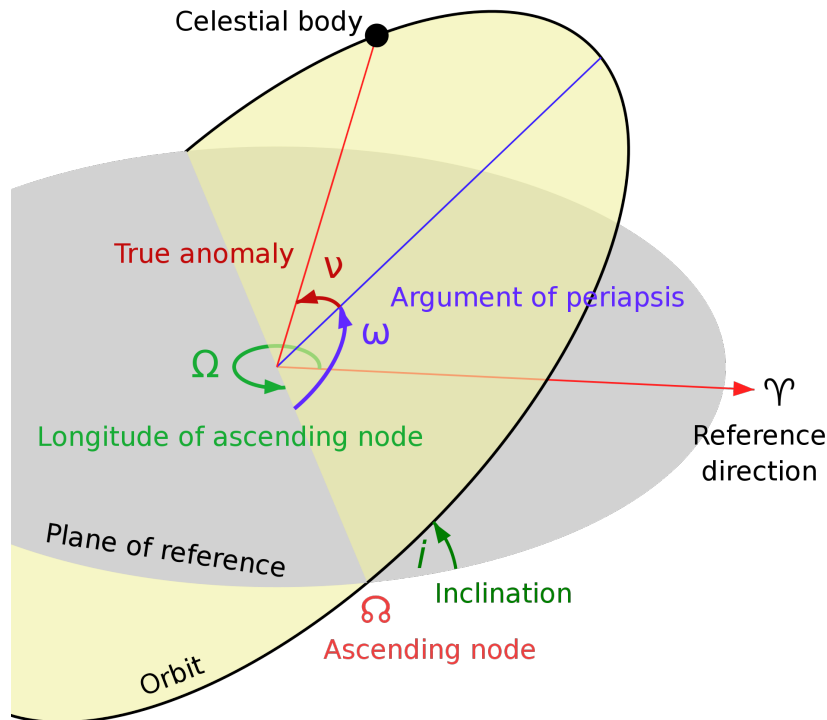
10

Figure 1.1: This figure shows the three angles $i$, $\Omega$ and $\omega$ graphically [Kep]. The orbit of the celestial body is tilted against the plane of reference by an angle $i$ about some axis that encloses the angle $\Omega$ with some reference direction within the plane of reference. The tip and end of the end of this axis are called the ascending and descending node respectively, because these are the points where the body submerges to or emerges from the plane of reference. From the point of view of the Earth, one typically chooses the vernal equinox as the reference direction and the celestial equatorial plane as the plane of reference.

With the argument of periapsis $\omega$, it is possible to quantify the position of the periapsis with respect to the axis of rotation, i.e. the ascending node. If $\omega = 0$, the periapsis is either the ascending or descending node, depending on the direction of the orbit.

### 1.3.3 The Equatorial Coordinate System

In astronomy, one usually uses equatorial coordinates to specify the position of an object on the night sky. This coordinate system uses the two angles called declination $\delta$ and right ascension $\alpha$. The declination is essentially the latitude with respect to the celestial equator. The right ascension is the measured angle on the celestial equator between the vernal equinox and the intersection of the circle of declination of the objects position on the night sky. Figure 1.2 gives a graphical explanation of these angles. The position of the $S$-stars are given in these coordinates relative to Sagittarius $A^*$.
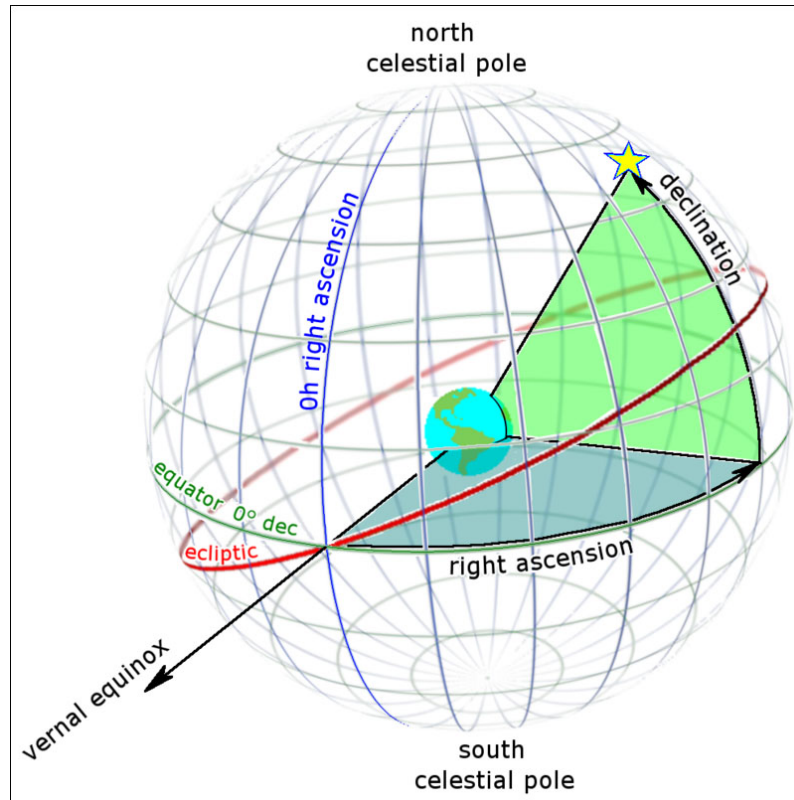


Figure 1.2: This figure is a graphical explanation of the equatorial coordinates [RaD]. In contrast to other celestial coordinate systems, the equatorial coordinates use the projection of the Earths equator on the celestial sphere as the reference plane. The vernal equinox is the intersection of celestial equator and ecliptic. It is the point on the celestial equator where the Sun passes in spring.

# 1.4 Cosmology - Quintessence and Dark Energy

It is a well-known fact that the Universe we live in is not static but instead a dynamic object undergoing expanding motion [Hubble, 1929]. Perlmutter et al. [1997] and Riess et al. [1998] discovered that the Universe is not only expanding, but the expansion is accelerating. This observation gave rise to a plenty of competing extensions and modifications of Einsteins gravitational field theory, which today are condensed into the concept dark energy.

To name a few, we have the cosmological constant, Quintessence models and Scalar-Tensor theories. The cosmological constant was first introduced by Einstein to create a static universe. It was later rediscovered and used to explain the observed acceleration and is the simplest theory of dark energy. However, its physical interpretation remains unclear. There have been attempts to relate it to the vacuum energy density predicted by quantum field theory, but the values differ by 120 orders of magnitude [Adler et al., 1995]. Therefore it is only used as a phenomenological explanation today and serves as a benchmark for other theories.

A simple extension of Einsteins gravitational field theory is to add a scalar field to the Einstein-Hilbert action to account for the dark energy component. Other theories, such as scalar-tensor theories introduce multiple fields to model the gravitational interaction. These theories try to solve several problems at once and as a consequence are very complex and sometimes their behavior is poorly understood. In this thesis, the focus lies on the quintessence field as it is a very simple theory that has undergone extensive investigation by other researchers in the past, such that our results can be compared to existing literature. This allows us the build up a baseline to evaluate the performance of our models.

## 1.4.1 The Expanding Universe

If we assume a homogeneous and isotropic universe which is approximately valid on scales larger than 100 Mpc [Dodelson and Schmidt, 2021], then we can derive a set of equations from the Einstein equations, which are called the Friedman equations. These equations describe the evolution of such a universe and are one of the cornerstones of cosmology. However, these equations are useless if we cannot relate them to observable quantities. By measuring the luminosity of distant supernovae, it is possible to calculate the distance of the supernovae and relate them to the redshift. With the distance-redshift relation known, we can relate our measurements to the expansion rate and thus use the Friedmann equations to infer cosmological parameters like the value of the cosmological constant, the Universe's curvature or the fraction of dark matter it contains.

### The Friedmann Equations

The Einstein equations describe the evolution of a 4-dimensional spacetime manifold in the presence of matter and energy. They give rise to many intriguing concepts like gravitational waves, the Schwarzschild precession, gravitational redshifts and the expansion of

the Universe [Misner et al., 1973].

The metric $g_{\mu\nu}$ is the most important quantity in general relativity as it tells us how to measure spacetime intervals and thus allows to relate theory to experimental observations. The equation for the metric in the presence of an energy-momentum tensor $T_{\mu\nu}$ and a cosmological constant $\Lambda$ can be derived from the Einstein-Hilbert action

$$S_{\text{EH}} = \frac{c^4}{16\pi G} \int \mathrm{d}^4x \sqrt{-g}(R - 2\Lambda) + S_{\text{matter}} \tag{1.26}$$

by applying the Euler-Lagrange equations for fields to the corresponding Lagrange densities. The result are the Einstein equations

$$R_{\mu\nu} - \frac{R}{2}g_{\mu\nu} = -\frac{8\pi G}{c^4}T_{\mu\nu} - \Lambda g_{\mu\nu}. \tag{1.27}$$

The Ricci tensor $R_{\mu\nu}$ and Ricci scalar $R$ are traces over the Riemann curvature tensor $R_{\alpha\beta\gamma\delta}$. For a full calculation, see Weinberg [2008]. Thus, the Einstein equations are a very complicated set of ten coupled partial differential equations. It seems to be hopeless to solve them analytically, but by assuming a homogeneous and isotropic universe we arrive at a so called Friedmann-Lemaître-Robertson-Walter (FLRW) spacetime with the line element

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - a^2(t)\left(\frac{\mathrm{d}r^2}{1 - \kappa r^2} + r^2(\mathrm{d}\theta^2 + \sin^2(\theta)\mathrm{d}\phi^2)\right). \tag{1.28}$$

Here, $t$ is what is called the cosmic time, $r$ is the radial comoving coordinate and $\kappa$ is called the curvature parameter indicates the global shape of the Universe. When $\kappa = 0$, the universe is flat and we get back the Minkowski metric. In the case that $\kappa = 1$, the universe is closed and spherical while for $\kappa = -1$, the Universe is open and of hyperbolical shape. Calculating the Christoffel symbols, Ricci tensor and Ricci scalar and plugging them into the Einstein equation severely simplifies things. Another simplification is the assumption that the Universe if filled with one or more homogeneous fluids. The energy-momentum tensor of such a fluid has the following shape:

$$T_{\mu\nu} = \left(\rho + \frac{p}{c^2}\right)u_\mu u_\nu - \frac{p}{c^2}g_{\mu\nu} \tag{1.29}$$

The contribution of the cosmological constant can be rewritten as a contribution to the energy-momentum tensor and is therefore also modeled as a fluid with negative energy density $\rho_\Lambda = \frac{\Lambda c^4}{3}$. As a result, there remain only two coupled ordinary differential equations, which are called the Friedmann equations. Introducing the rate of expansion $H = \frac{\dot{a}}{a}$, the equations can be written in the following shape:

$$H^2 = \frac{8\pi G}{3}\rho - \frac{\kappa c^2}{a^2} + \frac{\Lambda c^2}{3} \tag{1.30}$$

$$\dot{H} + H^2 = -\frac{4\pi G}{3}\left(\rho + 3\frac{p}{c^2}\right) + \frac{\Lambda c^2}{3} \tag{1.31}$$

Throughout this thesis, equation (1.30) will be referred to as *Friedmann equation*, while equation (1.31) will be referred to as *acceleration equation*.

14

## Density Parameters and Equations of State

The standard model of cosmology, called the $\Lambda$CDM-model, assumes the presence of three fluids with different equations of state. The equation of state is a dimensionless number from statistical physics that relates the the density of a perfect fluid to its pressure through $p = w\rho c^2$. It is similar to the equation of state of a ideal gas in thermodynamics in that it describes the change of two related thermodynamic state variables.

One component represents the cosmological constant, that is dark energy with a negative energy density $\rho_\Lambda$ and an equation of state $w_\Lambda = -1$.

The second component $\rho_m$ represents dark and baryonic matter. Because they are assumed to be cold and non-interacting on the relevant scales, they have an equation of state $w_m = 0$ which implies that there is no internal pressure.

Radiation is represented by a third component $\rho_\gamma$ with an equation of state $w_\gamma = \frac{1}{3}$ that can be explained through the bosonic properties of photons [Fließbach, 2010]. Usually the quantities $\rho$ and $p$ are not used in cosmology. Instead, the critical density $\rho_c = \frac{3H^2}{8\pi G}$ is introduced and all the components are expressed relative to the critical density. This gives rise to the density parameter $\Omega_i = \frac{\rho_i}{\rho_c} = \frac{8\pi G}{3H^2}\rho_i$. The Friedmann equation (1.30) then becomes

$$\Omega_m + \Omega_\gamma + \Omega_\Lambda + \Omega_\kappa = 1. \tag{1.32}$$

Observations of the cosmic microwave background indicate that our Universe is approximately flat [Stompor et al., 2001, Spergel et al., 2007]. Thus the curvature density parameter vanishes. We will adopt this result from now on and set $\Omega_\kappa = 0$.

The conservation law $\nabla_\mu T^{\mu\nu} = 0$ for the energy-momentum tensor is equivalent to

$$\partial_t\rho + 3H(1 + w)\rho = 0 \tag{1.33}$$

for a flat universe. Solving this differential equation for the three different fluids tells us how they evolve in time. From that, one can derive how they evolve with the redshift using the relation to the scale factor $a = \frac{1}{1+z}$:

$$
\begin{aligned}
\rho_\Lambda &\propto \text{const.} \\
\rho_m &\propto (1 + z)^3 \\
\rho_\gamma &\propto (1 + z)^4
\end{aligned}
\tag{1.34}
$$

In the case of dynamic equations of state $w(t)$, one has to solve equation (1.33) accordingly.

## Redshift-Luminosity Data

Now that we have defined a model that is able to describe the expansion of the Universe, we need to relate quantities like the expansion rate $H$ to physical observables. As mentioned before, the acceleration of the expansion rate was first discovered by measuring the apparent luminosity $m$ and redshift $z$ of supernovae Ia.

Supernovae Ia are a useful tool for this, because the energy emitted is expected to be approximately constant, i.e. they can be used as standard candles [Colgate, 1979, Khokhlov et al., 1993]. It is possible to correlate the absolute luminosity $M$ to the width of the light curve which is a measure for the emission time.

Using this knowledge, an absolute luminosity $M_{\mathrm{SN}}$ is defined that is corrected for light curve width. This quantity is a universal constant for all supernovae Ia. The exact procedure is explained in Guy et al. [2010] and Jha et al. [2007]. Knowing the absolute magnitude of the supernovae allows us to calculate their luminosity distance $d_L(z)$ from the formula

$$\mu = m - M_{\mathrm{SN}} = 5 \log\left(\frac{d_L(z)}{1\,\mathrm{Mpc}}\right). \tag{1.35}$$

Here, $\mu$ is called the distance modulus. For a flat universe, the comoving distance $\chi(z)$ is related to the luminosity distance through $d_L(z) = (1 + z)\chi(z)$. Finally the comoving distance can be calculated from the expansion rate that we get from solving the Friedmann equations:

$$\chi(z) = c \int_0^z \frac{\mathrm{d}z'}{H(z', \{\Omega_{i0}\})} \tag{1.36}$$

The quantities $\{\Omega_{i0}\}$ are the initial conditions of the different density parameters today at $z = 0$. With the comoving distance it is possible to connect our measurements of the distance modulus and redshift to the expansion rate by adjusting the initial density parameters accordingly. A typical redshift-luminosity dataset is displayed in figure 1.3. Fitting a flat $\Lambda$CDM model to such a dataset gives the following parameters [PLANCK-Collaboration, 2020]:

$$\Omega_{m,0} = 0.3111 \pm 0.0056, \qquad \Omega_{\Lambda,0} = 0.6889 \pm 0.0056, \qquad \Omega_\gamma \approx 0 \tag{1.37}$$

These values serve as a baseline for other more elaborate theories, because these have to be able to reproduce the results from the $\Lambda$CDM model as they already provide a very good fit to the observations.

It should be mentioned that with the Hubble constant $H_0$, the $\Lambda$CDM model has one more free parameter. The Hubble constant is the value of the expansion rate at $z = 0$ and is usually constrained by other measurements, for example using the cosmic microwave background and baryon acoustic oscillations which gives $H_0 = 67.66 \pm 0.42$ [PLANCK-Collaboration, 2020].
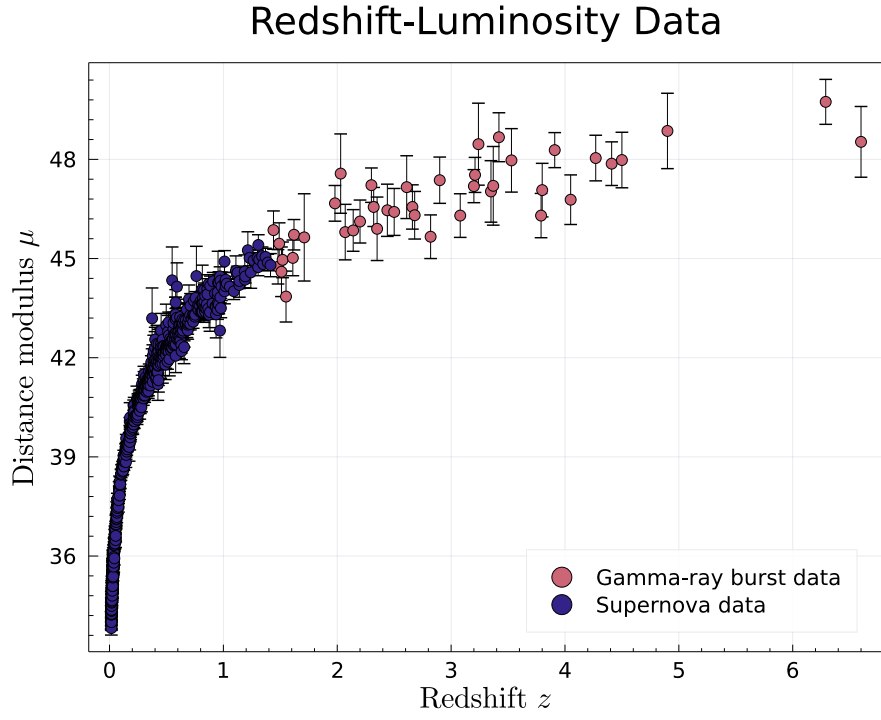
Figure 1.3: This is the redshift-luminosity dataset that will be used in this thesis to analyze the dark energy problem. It consists of two datasets. In the redshift range $[0, 1.5]$, the data is mainly comprised by supernova observations while for larger redshifts, Gamma-ray bursts are the dominant source of observations. It is also evident, that the magnitude uncertainties grow with increasing redshift. Datasets like this have also been used by Perlmutter and Riess to infer the expansion of the Universe.

## 1.4.2 The Quintessence Field

One possibility to go beyond the phenomenological explanation of dark energy through a cosmological constant is to introduce an additional scalar field $\phi$ to the Einstein-Hilbert action [Amendola and Tsujikawa, 2010]:

$$S_Q = \int d^4x \sqrt{-g}\left(\frac{c^4}{16\pi G}R - \frac{1}{2}g^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi - V(\phi)\right) + S_{\text{matter}}. \tag{1.38}$$

Once again applying the Euler-Lagrange formalism and assuming an FLRW spacetime gives an equation of motion and an energy-momentum tensor for the scalar field. Due to the requirement of a homogeneous and isotropic universe, the scalar field $\phi$ can only depend on the cosmic time $t$. The equation of motion of the scalar field is similar to the one-dimensional Klein-Gordon equation:

$$\ddot{\phi} + 3H\dot{\phi} + c^2\frac{\partial V}{\partial\phi} = 0 \tag{1.39}$$

The additional term $3H\dot{\phi}$ is called the *Hubble friction*. It suggests that there are two possible types of solutions, damped oscillating solutions and solutions where the friction term is dominant. From the energy-momentum tensor, we can derive an energy density and pressure for the quintessence field that looks like [Kolb and Turner, 1990]

$$\rho_\phi = \frac{\dot{\phi}^2}{2c^2} + V \quad\text{and}\quad p_\phi = \frac{\dot{\phi}^2}{2} - c^2V. \tag{1.40}$$

This leads to an equation of state that depends only on $\dot{\phi}$ and $V$:

$$w_\phi(t) = \frac{p_\phi}{c^2\rho_\phi} = \frac{\dot{\phi}^2 - 2c^2V}{\dot{\phi}^2 + 2c^2V} \tag{1.41}$$

It is required for the equation of state $w_\phi < -\frac{1}{3}$ to have late-time cosmic acceleration [Weinberg, 2008]. This results in the condition $\frac{\dot{\phi}^2}{c^2} < V(\phi)$ for the potential which implies that is has to be shallow enough for the scalar field to evolve slowly. In a similar manner, $w_Q$ should not be smaller than $-1$ for longer periods of time due to the the possibility of a big rip [Caldwell et al., 2003].

For a flat universe with pressureless dark matter and a negligible radiation component, the Friedmann and acceleration equations (1.30) and (1.31) become

$$H^2 = \frac{8\pi G}{3}\left(\rho_m + \frac{\dot{\phi}^2}{2c^2} + V\right),$$
$$\dot{H} = -\frac{4\pi G}{3}\left(\rho_m + \frac{\dot{\phi}^2}{c^2}\right). \tag{1.42}$$

The quintessence density parameter $\Omega_\phi$ is defined by

$$\Omega_\phi = \frac{8\pi G\rho_\phi}{3H^2} = \frac{8\pi G}{3H^2}\left(\frac{\dot{\phi}^2}{2c^2} + V\right). \tag{1.43}$$

18

Similar to cosmic inflation models, the scalar field is considered "slow-rolling", if

$$\dot{\phi}^2 \ll V(\phi), \quad \left|\ddot{\phi}\right| \ll \left|3H\dot{\phi}\right|. \tag{1.44}$$

Using these conditions, one can define the slow-rolling parameters

$$\epsilon_V = \frac{1}{16\pi G}\left(\frac{1}{V}\frac{\mathrm{d}V}{\mathrm{d}\phi}\right)^2, \eta_V = \frac{1}{16\pi V}\frac{\mathrm{d}^2 V}{\mathrm{d}\phi^2}. \tag{1.45}$$

When the field is slow-rolling, these parameters satisfy $\epsilon_V \ll 1$ and $\eta_V \ll 1$. Also, in that case we can approximate the equation of state through $1 + w_\phi \cong \frac{2\epsilon_V}{3}$, i.e. for perfectly slow-rolling field, quintessence is indistinguishable from the $\Lambda$CDM model with $w_\phi = -1$.
It is obvious that the quintessence potential shapes the way in which a universe evolves. The possible potentials can roughly divided into two groups [Caldwell and Kamionkowski, 2009]:

- *Thawing Models* For such models, the potential has a minimum $V(\phi) = 0$ that is accessible in a finite range of values for the scalar field $\phi$. Such models are said to be frozen by the Hubble friction at $w_\phi = -1$ for high redshifts and start to roll down the potential as the Hubble constant decays. They evolve towards $w_\phi = 0$.

- *Freezing Models* These models have no minimum that is accessible in a finite range. As a result the field rolls down the potential indefinitely but decelerates so that the equation of state evolves towards $w_\phi = -1$.

By looking at the equation of state, one can simply classify a given model into one of these two groups.

# 2 Introducing the Model - Harmonic Oscillations

This chapter aims to introduce and explain the method of neural potentials using the example of a harmonic oscillator. Beginning with data generated using a quadratic potential, the concepts and setup are explained as well as its advantages and disadvantages. Then, more complex potentials are analyzed to show that the method generalizes to this case as well. It is also shown that the model can capture small local features, which indicates that it can be applied to a vast amount of different problems.

## 2.1 A Simple Harmonic Oscillator

A simple harmonic oscillator with a quadratic potential is often observed in nature. For example a pendulum exhibits this behavior for small deflection angles. The differential equation that describes such and experiment is given through equation (1.9) with a potential $V(x) = \frac{1}{2}kx^2$. Measuring the displacements or angles at different times produces a time-series of observations $\{(t_i, x_i)\}_{i \in I}$.

We want to reconstruct the potential $\frac{1}{m}V(x)$ on the interval $[x_{\min}, x_{\max}]$ from this data by modeling it through a neural network. However this is not directly possible in Julia because automatic differentiation through derivatives of functions is sometimes not well-defined and thus prone to errors. Instead, we model the gradient of the potential, $\frac{1}{m}\frac{dV(x)}{dx}$ through a neural network $\mathrm{NN}(x, \theta)$ with parameters $\theta$:

$$\frac{d^2x(t)}{dt^2} = -\mathrm{NN}(x, \theta) \tag{2.1}$$

For one-dimensional problems, the potential has one input neuron and one output neuron and several hidden layers of fully connected nodes. The right number of nodes per layer has to be determined empirically, but often a choice between 8 and 32 neurons per layer is more than sufficient for the model to capture the gradient of the potential. For the model to work properly, that means to be able to capture non-linear behavior, it needs at least two hidden layers with non-linear activation functions.

By integrating the neural network, the potential can be reconstructed. Comparing our differential equation (2.1) to equation (1.2) shows us that it is necessary to convert this second-order ODE into a system of first-order ODEs. Thus, we introduce the quantity $v(t)$, which is essentially the velocity of the mass on the spring:

$$\begin{aligned}
\frac{dx(t)}{dt} &= v(t) \\
\frac{dv(t)}{dt} &= -\mathrm{NN}(x, \theta)
\end{aligned} \tag{2.2}$$

This is a step that will have to be performed for any other model in the future. The function $f$ that defines the dynamics of the hidden layers in section 1.1.1 is now a two-dimensional mapping that has a closed expression in the first row and a neural network in the second row. It is obvious, that this will strongly impact the behavior of the model compared to a continuous-depth recurrent neural network where the entire right-hand vector is replaced by a single neural network.

At this point, it is a good idea to stop and investigate manifestations of physics-induced priors in our model. The closed expression in the first row is an artifact of Newton's law, which tells us that we only have second-order derivatives. More generally, one could use Ostrogradsky's theorem to argue why we only need to use second-order time derivatives [Motohashi and Suyama, 2015]. It states that time-derivatives of third order or higher often cause instabilities in the equations of motion.

Another example is the use of a potential $V(x)$, or rather its gradient, which is a direct manifestation of energy conservation in the model. This also already shows the limits of this approach of neural potentials, because it requires the experiment to be at least approximately energy-conserving or to know a closed expression for the dissipative terms to be able to build them into the model.

### 2.1.1 Training the Model

The neural potential approach allows to learn not only the parameters of the neural network and subsequently the potential, but also to infer initial conditions as well as parameters of the closed expressions. In the following experiments, we will focus on the first two possibilities. All models will be trained using a $\chi^2$-statistic as cost function and an ADAM optimizer. The neural network has two fully connected hidden layers with 8 nodes and *ReLU* activation functions and no activation function on the output layer to put no constrains on the potential.

#### Training Algorithm

We start with a random initialization of the parameters and initial conditions using a uniform distribution. All of the models have been trained with an initial learning rate of 0.2 for 1500 epochs. The differential equation (2.2) is solved in the forward pass using an adaptive Runge-Kutta 5/4 scheme [Tsitouras, 2011]. The result is plugged into a $\chi^2$-statistic. With the same solver, the adjoint ODE is solved to compute the gradients with respect to the models parameters.

The ADAM optimizer then adjusts the parameters and initial conditions accordingly and uses past gradients to modify the learning rate. Evaluation and training of the model are depicted graphically in figure 2.1. The described procedure is repeated until the model is fitted properly. A simple termination criterion can be defined by requiring $\chi^2$ to be of order $O(1)$ or by checking if the cost function stalls at a certain value.
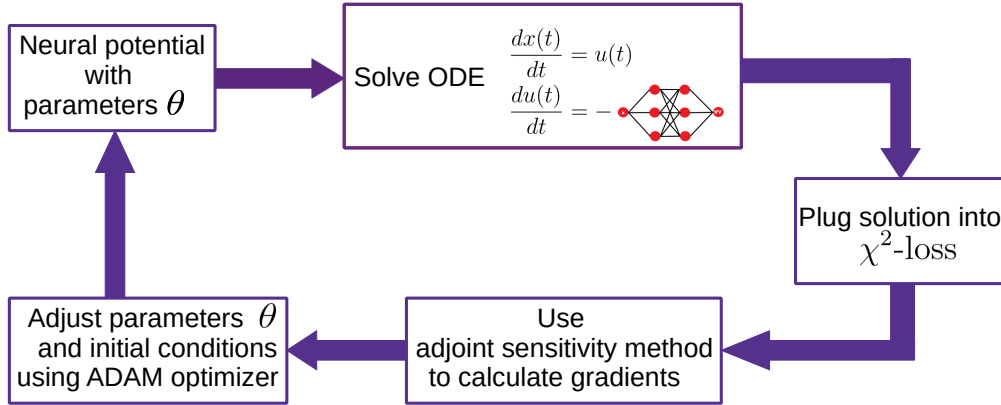
Figure 2.1: Graphical depiction of the algorithm. It runs until the $\chi^2$-loss is small enough or begins to stall at a certain value.

## 2.1.2 Running the First Experiment

To run the first experiment, synthetic data of a harmonic oscillator with initial conditions $x(0) = 3.0$ and $\dot{x}(0) = 0.0$ as well as parameters $k = 2$ and $m = 1$ is generated. The dataset contains 256 observations with additional Gaussian noise added with a standard deviation of $\sigma = 0.1$ to create a more realistic sample. Only positional data is used during the learning process as the velocity data is implicitly constrained through the shape of the model. When starting the training procedure, the models initial conditions are randomly set to values in the interval $[-5, 5]$ and the neural network parameters are randomly sampled from a normal distribution.

The initialization procedure is the same for all following experiments. The result of the learning procedure is depicted in figure 2.2. Running the model was a success as the potential is approximated very well within the relevant interval. Slight deviations from the original potential are due to the additional artificial noise. The inferred initial conditions have the values $x(0) = 2.986$ and $\dot{x}(0) = -0.006$ and are therefore also very close to the original ones.

### Confidence Intervals

However, it would be nice to have confidence interval on the predictions of the potential and the initial conditions. As described in section 1.1.3, they can be constructed using bootstraps and percentiles. The model is evaluated 1024 times on the same dataset to construct the $1\sigma$ and $2\sigma$-confidence intervals, i.e. the 68% and 95% confidence limits. It is a good practice to do this using multi-threading because the model is small and a single modern CPU suffices to solve the problem efficiently.

Figure 2.3 shows the resulting prediction. The uncertainties grow for larger values of $x$, in particular for the region $-2 > x > -3$. This is because there is only one minimum of the oscillations included. As a result the model sees only few observations is this region. The initial conditions have values $x(0) = 3.004 \pm 0.012$ and $\dot{x}(0) = -0.037 \pm 0.012$ where we gave the 68% confidence limits.

**Predictions**

The model can also be used to make predictions for later times. Consider 434 observations evenly distributed in the interval $[0.0, 8.5]$. Taking the subset $[0.0, 6.0]$ for training and then solving the ODE on the entire interval gives us a way to test the performance. The data in the interval $(6.0, 8.5]$ provides a test dataset that the neural network has not seen before. Comparing the predictions in figure 2.4 to figure 0.1, where a RNN was used to make a prediction, our model performs much better for the same dataset even though a smaller interval was used for the training.

The confidence into the predictions is much higher and the periodic motions are reproduced with high accuracy. This is a direct result of the implementation of physics-induced priors like energy conservation and suggests that our model maybe outperforms RNNs in other physics-related scenarios as well. It is important to mention here that the model is very good at generalizing in the $t$-direction, i.e. predict how a certain trajectory will continue. However, it will fail at generalizations into the $x$-direction, meaning that generalizations for trajectories that include points with $x$-values outside the trained interval are not possible. If for example the model is trained on data ranging from $-3.0$ to $3.0$, any attempts to evaluate a trajectory that starts with initial conditions outside this area are in vain as the neural network was not trained with respective samples.

## 2.2 Generalization to more elaborate Potentials

Analyzing the simple harmonic oscillator is not so much of a difficulty as the gradient of the potential, i.e. the force is just a linear function that is easily learned by a neural network. What happens if we increase the complexity of the potential, for example by putting a small "bump" in the middle of the two wings of it? A potential like this can be described through

$$V(x) = \frac{1}{4}x^4 - 2x^2. \tag{2.3}$$

This potential is used again to generate data with initial conditions $x(0) = 3.0$, $\dot{x}(0) = 0.0$ and additional noise with variance $\sigma = 0.01$. The variance has to be reduced so that the small features induced by the bump are still visible and learnable by the model. As the potential has become more complicated, it is useful to increase the number of neurons per hidden layer to 16. The neural network can then be trained with a NADAM optimizer with a learning rate of 0.01 and 10000 epochs. It turns out that even with these minor adjustments, the model is still able to reconstruct the potential as well as the initial conditions $x(0) = 2.8571 \pm 0.016$, $\dot{x}(0) = -0.300 \pm 0.038$.

The accuracy has decreased significantly, but the model was able to capture the "bump" and its position very well. This is a highly promising result as the model is able to capture local features and thus suggests its usage for more complicated problems.

The predictions may be improved by using a smaller initial learning rate and more iterations as well as using more observations. Another option would be to using additional trajectories with different initial conditions to learn specific features of the potential with a higher accuracy.
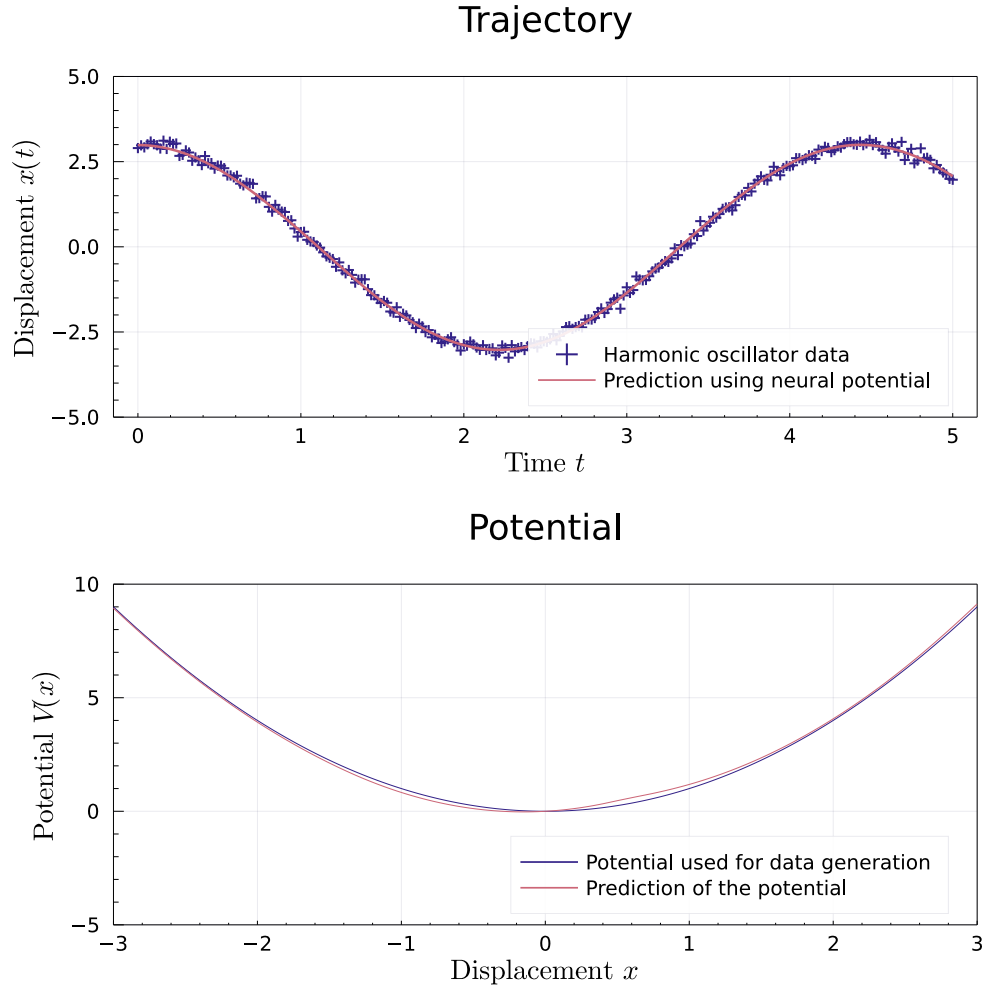


Figure 2.2: Result of the neural potential approach yielded a nearly quadratic potential that approximates the actual potential with high accuracy. The initial conditions were predicted to be $x(0) = 2.986$ and $\dot{x}(0) = -0.006$.
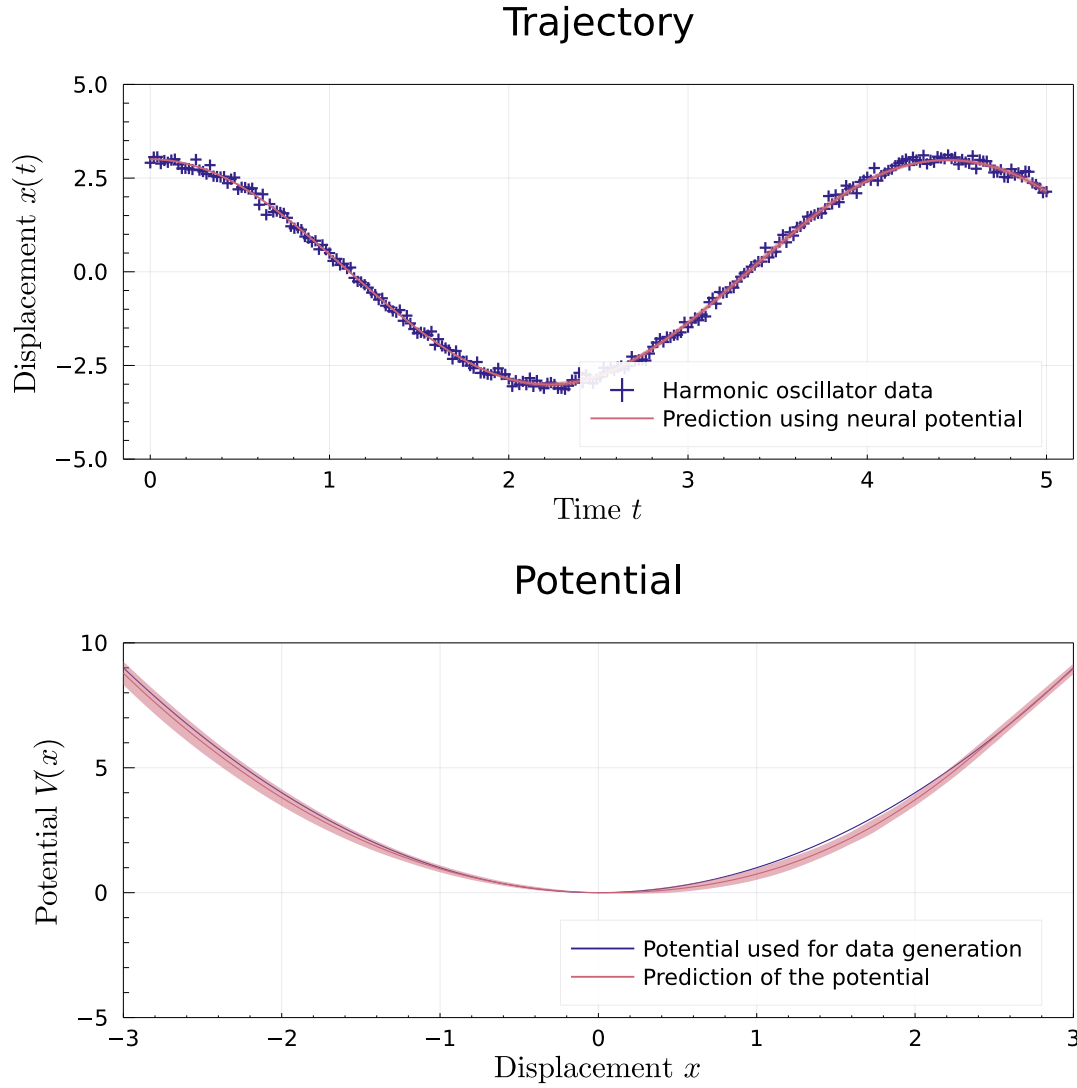
## Trajectory



## Potential



Figure 2.3: Using 1024 bootstrap repetitions where the model learns the potential on the same data, allows to construct $2\sigma$-confidence intervals for the initial conditions and the potential. The confidence intervals for the predictions of the trajectories are very small, which means the model is very confident in its predictions. It was able to reproduce the oscillations with high accuracy. The potential was also well approximated with high confidence. Additionally, the initial conditions were predicted to be $x(0) = 3.004 \pm 0.012$ and $\dot{x}(0) = -0.037 \pm 0.012$.
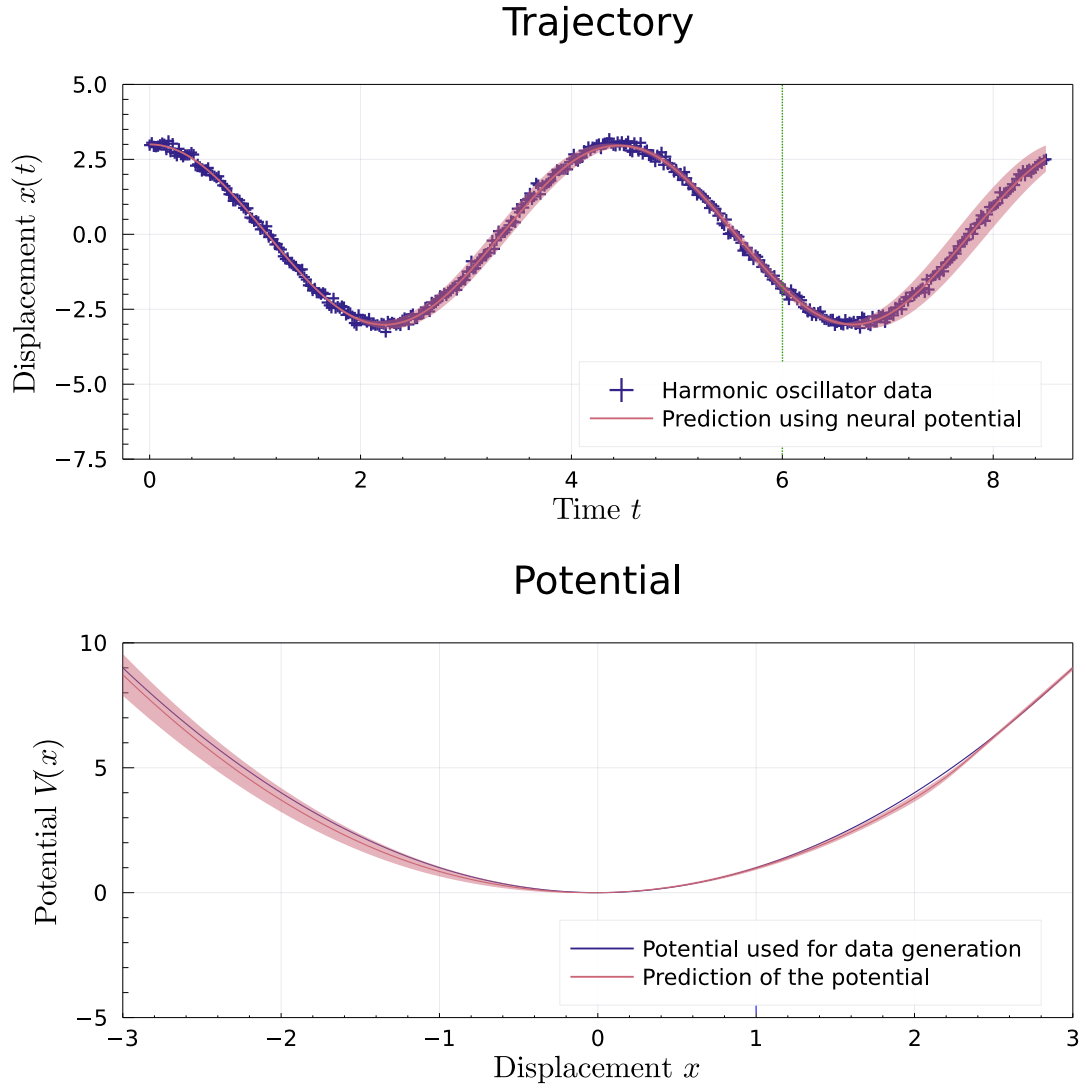
## Trajectory



## Potential



Figure 2.4: The model was trained on 256 observations in the interval $[0.0, 6.0]$ and 1024 bootstrap steps were performed. The training procedure yielded initial conditions of $x(0) = 2.999 \pm 0.008$ and $\dot{x}(0) = 0.029 \pm 0.001$. At the end of the training, the model was evaluated on the entire interval $[0.0, 8.5]$ for every bootstrap repetition. It is easy to see that the uncertainty in the prediction grows on the interval $(6.0, 8.5]$, because the model has not seen this data before. However, the model was able to extrapolate the harmonic oscillation with high accuracy.
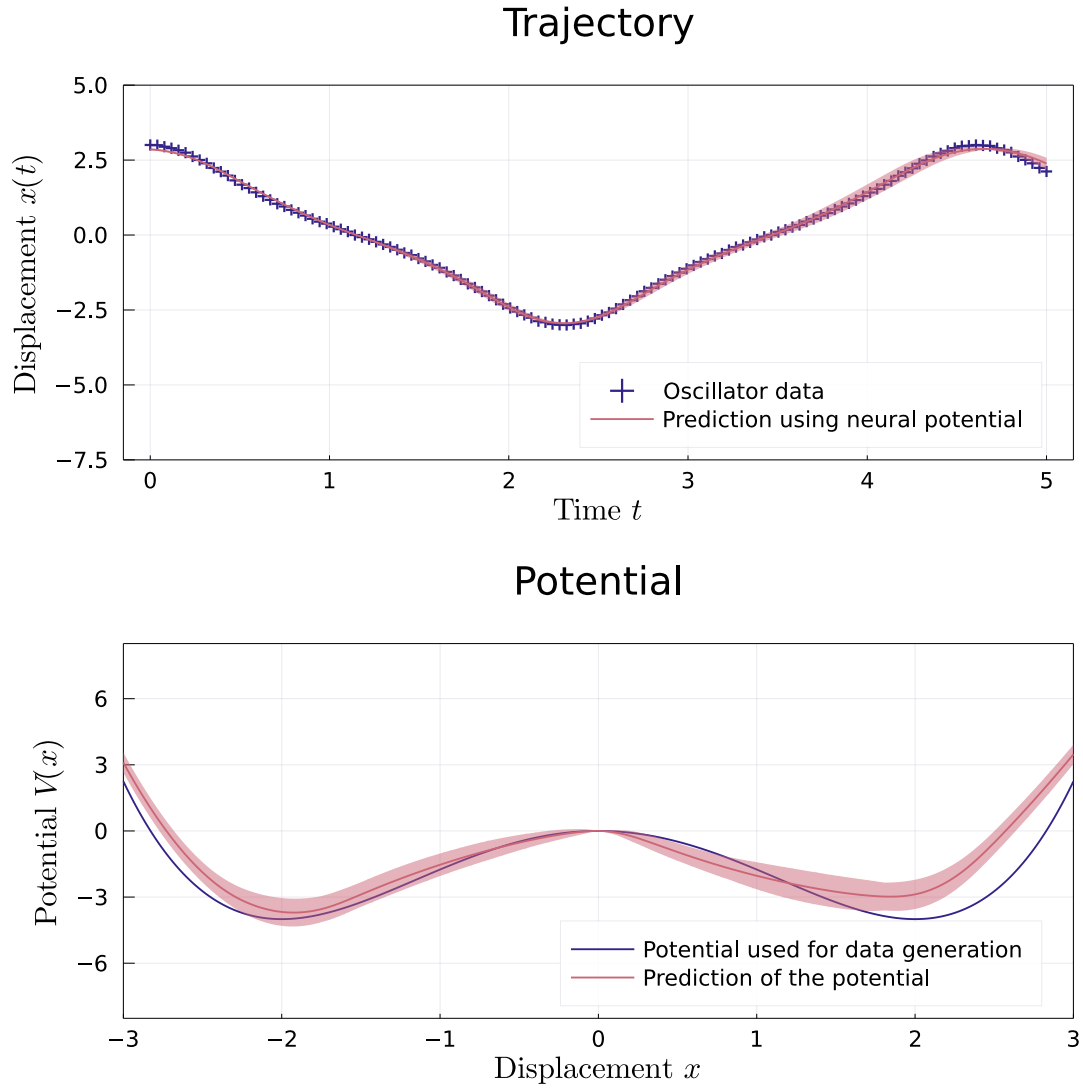
Figure 2.5: The figure displays the potential that the neural network has learned from the data generated with equation (2.3). The $2\sigma$-confidence limits are generated with 512 bootstrap iterations and show that the model was able to capture the important features of the potential, that is the W-shape with a "bump" in the middle. The resulting trajectory fits the data very well, although the precision decreases for increasing time-steps. The prediction of the left valley of the potential is closer to the original potential than the prediction of the right valley. The initial conditions are predicted to be $x(0) = 2.8571 \pm 0.016$ and $\dot{x}(0) = -0.300 \pm 0.038$.

# 3 Real-World Data - Celestial Mechanics

One of the biggest discoveries of the last decade was the verification of Sagittarius $A^*$ as the Milkyway's central supermassive blackhole. This was done by observing the trajectories of stars close to the center of our galaxy orbiting a luminous radio source. Using infrared interferometry techniques, it is possible to directly observe these trajectories [GRAVITY-Collaboration, 2017].

Infrared light has to be used because the interstellar dust absorbs almost all of the visible light coming from the central parsec of the Milkyway. Measuring the position of the stars in this region over the course of a couple of years allows to observe their elliptical orbits. This is the first real-world dataset that we will analyze using the technique of neural potentials. Our goal is to infer Newton's law and possible relativistic corrections from the observations of the star $S2$ which orbits Sagittarius $A^*$ with a period of around 16 years. For the dataset and additional information, see Gillessen et al. [2017] and the respective AdS repository.

## 3.1 Testing the Model on Synthetic Data

### 3.1.1 Defining the Model

Before we start working with the real-world data, we will evaluate the performance of the model on a synthetic dataset. With the formulas presented in section 1.3, it is possible to construct a simple model that can be fitted to the data:

$$
\begin{aligned}
\frac{\mathrm{d}u(\phi)}{\mathrm{d}\phi} &= v(\phi) \\
\frac{\mathrm{d}v(\phi)}{\mathrm{d}\phi} &= -u(\phi) - \mathrm{NN}(u, \theta)
\end{aligned}
\tag{3.1}
$$

This model is quite similar to the model used for the harmonic oscillator in (2.2). Note however that the trajectories are parameterized through the angle $\phi$ and that the quantity $u(\phi)$ is equal to the inverse radius of the trajectory. The term $-u(\phi)$ is a manifestation of the centrifugal forces when transforming to polar coordinates and is included explicitly. This is another example of the implementation of a physical prior. By choosing the neural network NN to depend only on $u$, we already make strong assumptions about the system like the isotropy of the potential which results in angular momentum conservation. However, the ODE alone is not sufficient to construct a model that is able to fit the data. We also have to take into account that the observed trajectory can be rotated against the observational plane which is orthogonal to the line of sight.

## 3.1.2 Rotating the Trajectory

So how do we solve this problem? Assume we have a trajectory that lies fully in the observational plane, i.e. the $xy$-plane. The three angles inclination $i$, longitude of ascension $\Omega$ and argument of periapsis $\omega$ can be used to rotate this trajectory such that it has a component in the direction of the line of sight, i.e. the $z$-direction. By projecting the rotated orbit back onto the observational plane, we get the effective trajectory $(r, \phi)$ that we are able to observe. The transformation is realized by two successive rotations and a projection. First, the observations are rotated around the axis

$$\mathbf{n} = \begin{pmatrix} \cos(\Omega) \\ \sin(\Omega) \\ 0 \end{pmatrix} \tag{3.2}$$

by an angle $i$. Then, the observations are rotated around the axis

$$\mathbf{m} = \begin{pmatrix} \sin(i)\sin(\Omega) \\ \sin(i)\cos(\Omega) \\ \cos(i) \end{pmatrix} \tag{3.3}$$

by an angle of $\Omega + \omega$. Finally, the observations are projected onto the observational plane by neglecting the $z$-component. This is also the way the synthetic data is generated.

Because the ODE solution and the observations have to be compared at the same $\phi$-values to be able to compute a meaningful cost function and because we can only solve the ODE in the observational plane, we have to invert the rotation and take back the trajectory of our observed celestial body into the observational plane. This can be done by reverting the projection through constructing a $z$-component with

$$z = r\sin(\phi - \Omega)\tan(i) \tag{3.4}$$

because the plane of the trajectory in known. After that, the two rotations are performed in reverse order and opposite direction (i.e. first around $-\mathbf{m}$ then $-\mathbf{n}$) to recover the orbit in the observational plane. Then the polar angles of the trajectory are used as steps for solving the model. After that, the solution is transformed back according to the process described above.

Finally, the prediction can be compared to the data using a $\chi^2$-statistic or MSE cost function. Although this is a complicated transformation, Julia ensures that it is differentiable at every step. Therefore, the three rotation angles are new learnable parameters together with the neural network and the initial conditions of the ODE.

### 3.1.3 Pitfalls of Deep Learning

The learning procedure is summarized in the following way:

1. Make a guess for the rotation angles.

2. Transform the observed trajectory into the observational plane using the angles.

3. Use the polar angles $\phi$ as steps to solve the ODE and get a prediction of our model.

4. Construct the prediction by transforming the solution back again.

5. Compare the prediction to the observed trajectory.

6. Calculate gradients for all parameters.

7. Adjust all parameters using an optimizer.

We see that the rotations in steps 2. and 4. are in a sense inverse to each other. Thus, it is to expect that their gradients will almost cancel out and leave the parameters unchanged. But when reverting the projection, we also need to use $i$ and $\Omega$. This additional step between the two transformations counteracts this effect and makes these parameters easily learnable. Yet, the argument of periapsis $\omega$ is not included when reverting the projection and thus suffers from a small gradient of order $10^{-10}$ to $10^{-12}$.

The problem is solved by boosting the resulting gradient by a factor of around $10^{12}$. However, the gradients will then start to oscillate very fast. For strongly oscillating gradients on one parameter, it is common to use an optimizer with Nesterov momentum and per-parameter learning rates to smooth them out. One realization of this requirement is the NADAM optimizer, which is used for all following experiments [Dozat, 2016]. Optimizers with Nesterov momentum often suffer from building up high momentum and then overshooting the minimum of the cost function. To solve this, the learning rate of the optimizer has to be scheduled explicitly.

### 3.1.4 Newton and Relativistic Corrections

Before training the model on the $S2$ data, we use it to examine synthetic data of a hypothetical star orbiting a supermassive black hole (SMBH). First, we test the setup by replacing the neural network term NN with a predefined potential:

$$\frac{\mathrm{d}V}{\mathrm{d}u} = GM\left(\frac{1}{(r_0 v_0)^2} + 3\frac{u^2}{c^2}\right). \tag{3.5}$$

This potential is also used to generate the synthetic data. The first term in equation (3.5) was simplified using conservation of angular momentum and $L_z = \mu r_0 v_0$. Our model has two free parameters apart from the two initial conditions and the three rotation angles, namely $M$ and $\frac{1}{r_0 v_0}$. All parameters are initialized by sampling from a uniform distribution defined on a respective meaningful intervals.

| Parameter | Synthetic Data | Prediction |
|---|---|---|
| Initial inverse position $u(\phi = 0)$ | 2.0 | $1.25 \pm 0.54$ |
| Initial derivative $\left.\dfrac{\mathrm{d}u}{\mathrm{d}\phi}\right|_{\phi=0}$ | $-0.12$ | 0.31 |
| Inclination $i[°]$ | 50.0 | $49.92 \pm 0.13$ |
| Longitude of ascending node $\Omega$ [°] | 100.0 | $100.04 \pm 0.046$ |
| Argument of periapsis $\omega$ [°] | 65.0 | $70.05 \pm 86.55$ |
| Black hole mass $M$ [$10^6 M_\odot$] | 4.35 | $4.19 \pm 0.28$ |
| Parameter $\dfrac{1}{r_0 v_0}$ [yr mpc$^{-2}$] | 0.27 | $0.27 \pm 0.01$ |

Table 3.1: Results for the 144 synthetically generated observations using the fixed potential in equation (3.5) as fiducial model. The quantity $v_0$ describes the velocity of the star at the periapsis. To generate the 68%-confidence limits, 512 bootstrap repetitions were used. The model is particularly successful at reconstructing the parameters of the potential and the angles $i$ and $\Omega$. The initial conditions and the angle $\omega$ do not agree well with the true values and have a high uncertainty.

For example $M$ is sampled from the interval $[0, 8]$, while the inclination is sampled from an interval $[0, \pi]$. The model is trained using a NADAM optimizer with a learning rate of 0.005 for 20000 epochs. For the synthetic data, which has no added noise, we use a MSE cost function. In order to make the model converge faster and numerically stable, the natural constants are readjusted such that distances are measured in milliparsec, time-intervals in years and masses in units of $10^6 M_\odot$. This gives $G = 4.49 \cdot 10^{-6}$ mpc$^3$yr$^{-2}M_\odot^{-1}$ and $c = 306.4$ mpc yr$^{-1}$.

The results of the training procedure are displayed in table 3.1.4 and the figure 3.1. As the figure shows, the model was able to reconstruct the trajectory very well with high confidence. This is also what we would expect since we used the same potential for both data generation and model fitting and have no additional noise. By looking at the table, we see that the model infers the parameters of the potential with high accuracy as well as the angles $i$ and $\Omega$.

However, it struggles with the initial conditions and the angle $\omega$, as there is a degeneracy between them. When generating the trajectory, we assume that the periapsis lies at $\phi = 0$, such that $\frac{\mathrm{d}u}{\mathrm{d}\phi} = 0.0$ and the distance $r$ is the smallest, i.e. $u$ is the largest. By reducing $u(\phi = 0)$ and making $\frac{\mathrm{d}u}{\mathrm{d}\phi}$ different from zero, the periapsis is no longer at $\phi = 0$, but at some other polar angle. Rotating the trajectory accordingly can compensate this, but as a result we have to adjust the argument of periapsis.
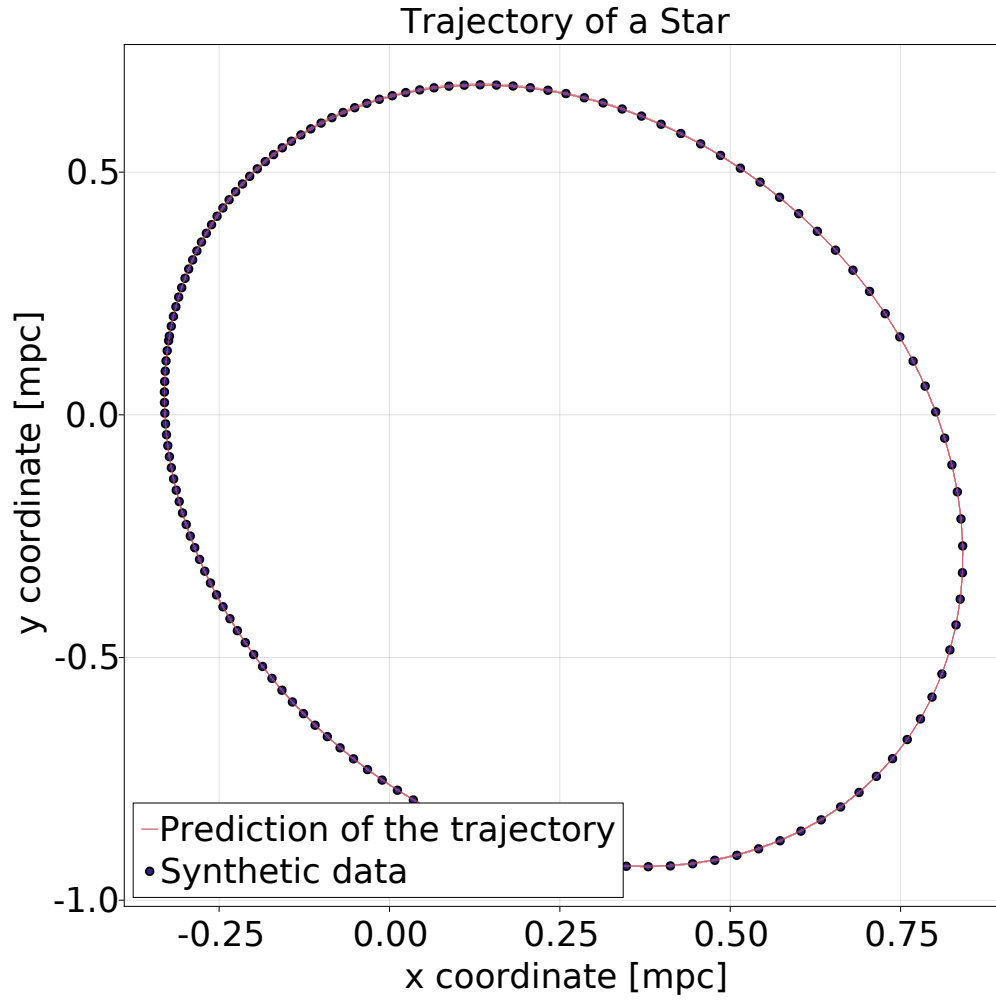
Figure 3.1: Results for the synthetical data using a fixed potential. The the $2\sigma$-confidence intervals were generated using 512 bootstrap repetitions. Our model has predicted the trajectory with a very high confidence as the confidence regions around the trajectory are barely visible.

| Parameter | Bootstrap | Periapsis Shift |
|---|---|---|
| Initial inverse position $u(\phi = 0)$ | $1.16 \pm 0.45$ | $0.86$ |
| Initial derivative $\left.\dfrac{\mathrm{d}u}{\mathrm{d}\phi}\right|_{\phi=0}$ | $-0.04 \pm 0.30$ | $-0.30$ |
| Inclination $i[°]$ | $46.49° \pm 4.44°$ | $49.64$ |
| Longitude of ascending node $\Omega[°]$ | $103.50° \pm 5.66°$ | $100.13$ |
| Argument of periapsis $\omega$ $[°]$ | $64.00° \pm 75.61°$ | $208.74$ |

Table 3.2: Results for the 144 synthetically generated observations using the potential in equation (3.5). The first column shows the results for a single revelation with the same synthetic data as before, while the second column shows the results for two successive revelations. To generate the 68%-confidence limits for the single revelation, 512 bootstrap repetitions were used. Again, we see that the model struggles with the initial conditions and the angle $\omega$.

## 3.1.5 Learning the Gravitational Potential

Next, we return to the neural potential approach and train a neural network using the same synthetic data. The neural network consists of two layers with 8 nodes respectively and is trained for 7500 epochs using a NADAM optimizer with a learning rate of 0.01. Although *ReLU* activation functions would be a reasonable choice, the model converges much faster by using a layer with a *CeLU* activation function and a layer with a radial basis function

$$\mathrm{rbf}(x) = \sqrt{1 + \frac{1}{4}x^2}. \tag{3.6}$$

This approach is inspired by a special machine learning technique called *Radial Basis-function Networks* [Lowe and Broomhead, 1988]. The weights and biases of the neural network are initialized using a Gaussian normal distribution. Training the neural network, initial conditions and rotation angles yields the results presented in figure 3.2 and table 3.1.5. We see that the model is able to reconstruct the potential with acceptable confidence. For a large $u$, i.e. distances close to the SMBH, the uncertainty becomes larger. The reason for this are the relativistic effects which make it harder for the model to learn the correct potential. The trajectory is predicted to satisfactory precision, but there certainly is room for improvement.

We used only a single revelation for the test, but it does also work for multiple revelations, see figure 3.3 and table 3.1.5. For this case, the periapsis shift is correctly captured by the model even though it is not very prominent. The reason for the increase in performance is probably the use of twice as many observations. Again, the model struggles with the degeneracy between the initial conditions and the argument of periapsis.
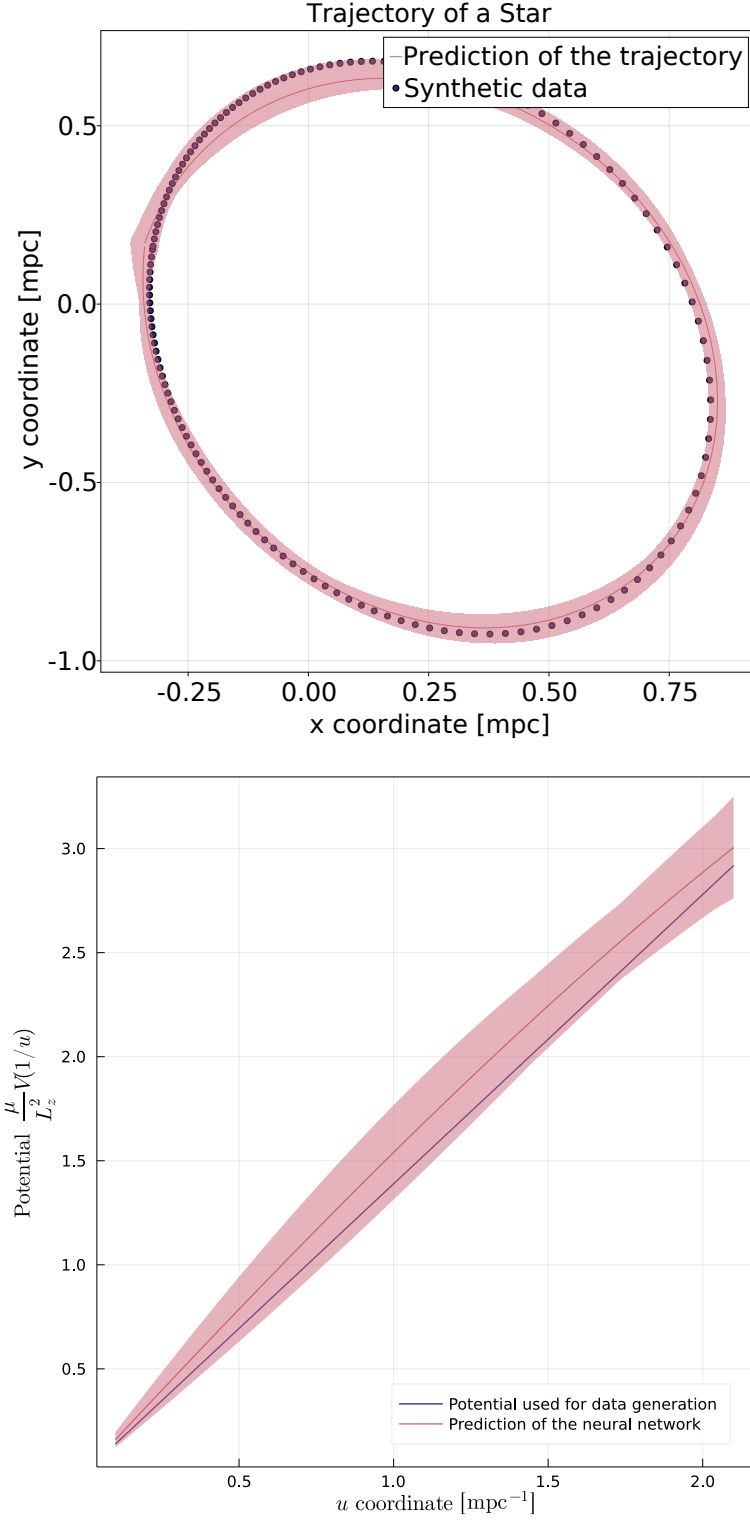
Figure 3.2: The figure shows the inferred potential for a system where a hypothetical star orbits a SMBH of $4.35 \cdot 10^6 M_\odot$. At the periapsis, the star has a distance of 0.5 mpc to the black hole while moving at a velocity of 7.5 mpc yr$^{-1}$. The $2\sigma$-confidence regions were created using 512 bootstrap repetitions.
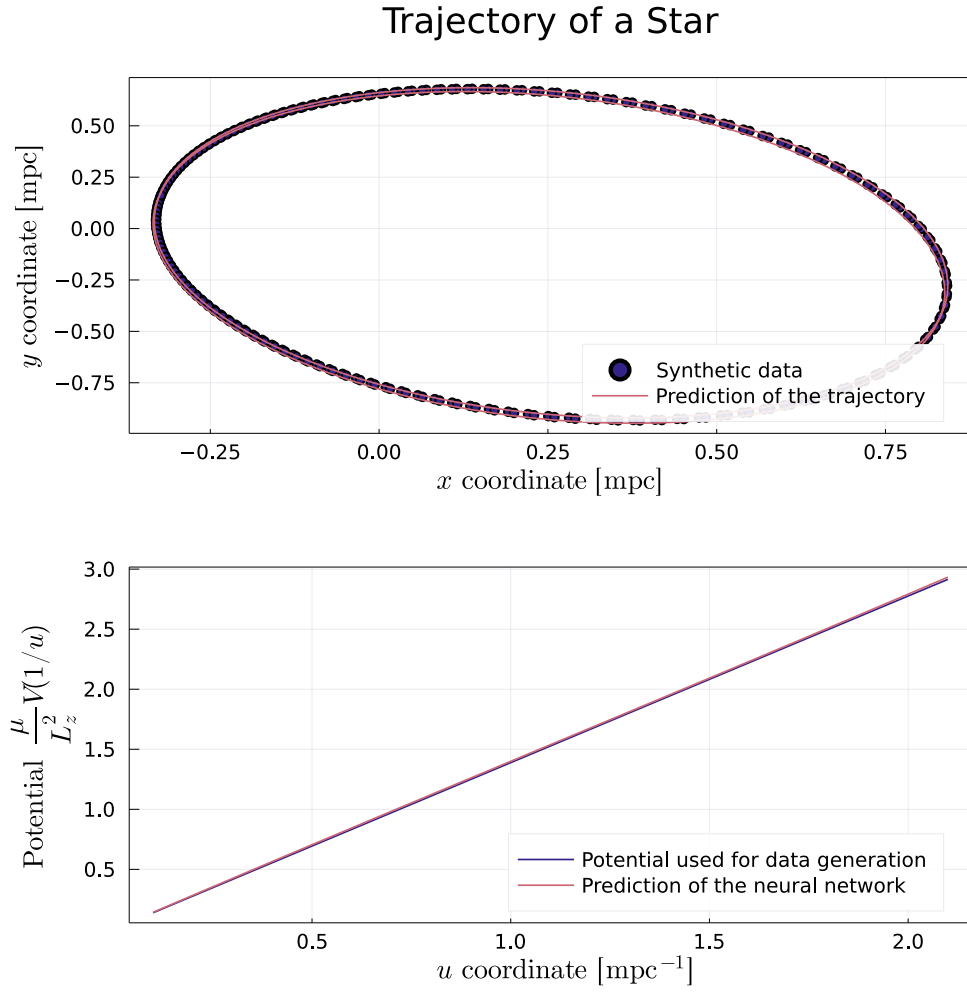
Figure 3.3: Result of the prediction using a neural potential. The data was generated with the same initial conditions as before, but with 288 observations for two revelations, such that a slight periapsis shift becomes observable. Although the periapsis shift is not very prominent, the neural network is able to capture it and the inferred potential agrees extraordinarily well with the underlying model.

## 3.2 Tackling the Galaxy's Central Black Hole

### 3.2.1 Dealing with Equatorial Coordinates

With a working model that is able to capture the periapsis shift as well as the rotation against the observational plane, we can move on to the observational data of the star $S2$. The 144 observed positions of the star $S2$ are given in relative equatorial coordinates, i.e. though right ascension $\alpha$ and declination $\delta$ relative to the position of Sagittarius $A^*$ on the celestial sphere. All angles are given in microarcseconds (mas). To be able to use the neural potential approach, we have to convert the angular distances into physical distances. If we know the distance of Sagittarius $A^*$, which we call $R_0$, then this is a matter of simple geometry. By converting the angles $\alpha$ and $\delta$ into radians and using

$$x = R_0 \tan(\alpha) \text{ and } y = R_0 \tan(\delta), \tag{3.7}$$

we obtain physical distances for the orbit of the star $S2$. In a similar manner, we can also reconstruct the uncertainties in $x$- and $y$-direction. The distance to the galaxy's central supermassive black hole $R_0$ is another free parameter of the model, but it will be fixed to $R_0 = 8178$ kpc for our purposes [GRAVITY-Collaboration, 2019]. Even though this distance is also obtained using observations of the star $S2$ and therefore not statistically independent, it is still a very good estimate with a low uncertainty.

### 3.2.2 Learning the Gravitational Potential from the Orbit of S2

With the positional data computed, we start analyzing the observations by using the Newtonian model with its relativistic correction (3.5). The model is fitted with the same initializations as for the synthetic model and is trained again with a NADAM optimizer with a learning rate of 0.001 for 25000 epochs using a $\chi^2$-statistic as cost functions. Figure 3.4 and table 3.2.2 show the results for this model.

It is especially promising that the model's predicts competitive results for the inclination angle $i$ and the parameters of the potential which are well in agreement with values inferred by other authors [Gillessen et al., 2017, GRAVITY-Collaboration, 2017]. This is not surprising since we essentially fitted the same physical model to the data by using two different methods. Contrary to the synthetic data, the longitude of the ascending node is not in agreement with the literature this time. The reason for this is probably a different choice of the reference direction. Note that a different choice of this reference direction also changes the argument of periapsis, so this introduces an additional systematic error for our prediction.

With these findings in mind, we move on to analyzing the data with the neural potential to see if we can learn Newton's gravitational potential and its relativistic corrections. The model is trained using the same neural network and initializations as for the synthetic data. Using a NADAM optimizer with a learning rate of 0.01 and training for 5000 epochs with a $\chi^2$-squared statistic as cost function gives the results displayed in figure 3.5 and table 3.2.2. We immediately see that the uncertainties for the angles are much larger for the neural potential model and do not agree well with predictions by other authors.

The reason for this is probably the added flexibility, that creates a degeneracy between the angles and the potential. Still, the best fit of the trajectories can explain the data with reasonable accuracy. It is especially successful for the observations at large radii, as the data there has much less uncertainty. On the other hand, the predictions close to the star on the other hand have high uncertainties and thus the model struggles with predictions there.

Additionally, as we have seen from the analysis of the synthetic data, the model struggles with small radii when relativistic corrections become relevant, because they are harder to approximate. Note that the lack of precision in the data close to the star is simply due to the fact that $S2$ takes 16 years complete an orbit. The last time it passed the periapsis was approximately 14 years ago, then measurement technology was not advanced enough to yield better observations.

This is also reflected by the potential, where we can see that for small radii, i.e. large $u$, the uncertainties grow very much. Despite that, the model still manages to capture the data and even predicts a small periapsis shift. However, with such low confidence for small radii, the predictions of the model cannot provide statistically meaningful evidence. This may be fixed by incorporating the data from the next periapsis passage, such that we obtain much more reliant results.

Nonetheless, it is impressive to see that the model has in principle the capabilities to learn the physics of a system with only sparse and noisy data.

| Parameter | Gillessen et al. | Fixed Potential | Neural Potential |
|---|---|---|---|
| Initial inverse position $u(\phi)$ | - | $0.47 \pm 0.21$ | $0.14 \pm 0.03$ |
| Initial derivative $\dfrac{\mathrm{d}u}{\mathrm{d}\phi}$ | - | $0.47 \pm 0.05$ | $-0.13 \pm 0.08$ |
| Inclination $i$ [°] | $135.25 \pm 0.47$ | $135.22 \pm 0.22$ | $121.01 \pm 26.37$ |
| Longitude of ascending node $\Omega$ [°] | $225.39 \pm 0.84$ | $64.36 \pm 1.94$ | $88.01 \pm 20.28$ |
| Argument of periapsis $\omega$ [°] | $63.56 \pm 0.84$ | $114.55 \pm 26.52$ | $243.60 \pm 178.02$ |
| Black hole mass $M$ [$10^6 M_\odot$] | $4.31 \pm 0.38$ | $4.23 \pm 0.48$ | - |
| Parameter $\dfrac{1}{r_0 v_0}$ [yr mpc$^{-2}$] | $0.23 \pm 0.01$ | $0.18 \pm 0.10$ | - |

Table 3.3: Results for the observed positions of the star $S2$. Note that now the initial condition is not at $\phi = 0$ but at some different angle, since the first measurement is taken as the initial point. The first column shows the results for a the fixed potential, while the second column shows the results for the neural potential. Both bootstraps were performed using 512 repetitions to create the 68% confidence limits. For the fixed case, the predictions of the parameters of the potential as well as the inclination angle are well in agreement with the predictions of [Gillessen et al., 2017]. In both cases, our model struggles with the degeneracy between initial conditions and argument of periapsis. Additionally, the angle $\Omega$ is predicted with high accuracy, but does not agree with findings from other authors. The results for the neural potential have much higher uncertainties and do not agree well with the findings of other authors.
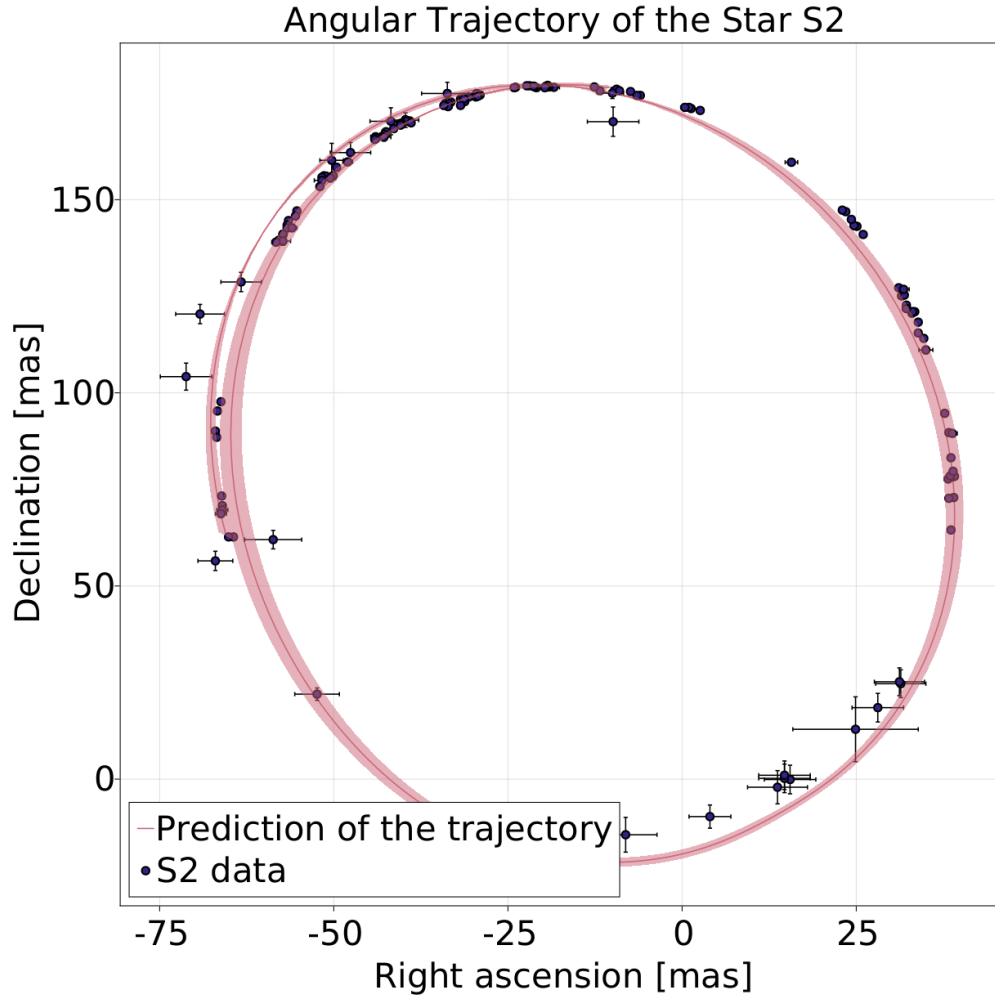
Figure 3.4: This figure displays the prediction of the trajectory and the $2\sigma$-confidence regions for the fixed potential in equation (3.5). The confidence intervals were created with a bootstrap of 512 repetitions. With the fixed potential, the model predicts a slight periapsis shift, that is most apparent on the left part of the orbit. The uncertainty is particularly low for the predictions at the upper part of the trajectory. This is mainly because there are many observations with a low uncertainty and thus the $\chi^2$-statistic gives them more weight.
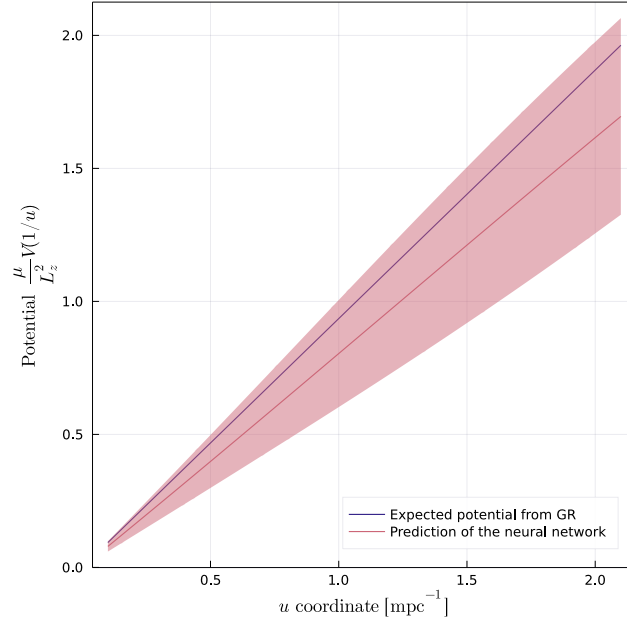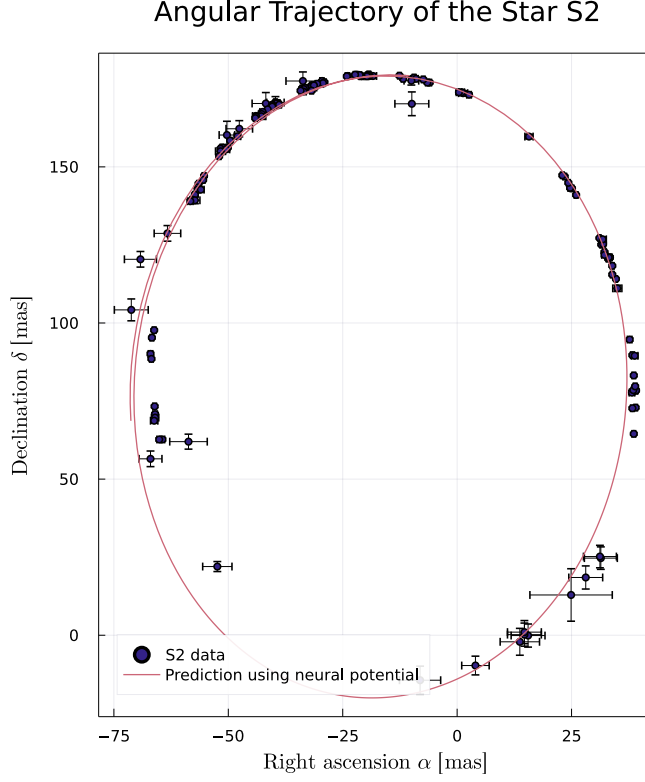
Figure 3.5: This figure shows the best fit trajectory and the predicted neural potential with its $2\sigma$-confidence limits. Due to technical issues, the confidence limits of the trajectory are not shown here. By looking at the potential, we see that the model predicts a potential very close to the one from general relativity. To generate the potential, we used the data from Gillessen et al. as presented in table 3.2.2. For decreasing radii, i.e. increasing $u$, the uncertainties become larger and the neural potential starts to deviate stronger from the expected potential.

### 3.2.3 Room for Improvement

As we have seen, the method of neural potentials can also handle real-world data and already creates promising results with only a few observations. Still, there is a lot of room for improvement. Apart from increasing the amount of data that we feed to our model, we can take several other measures to improve our method. First of all, the model does not take into account several important effects which are used by other, better performing methods [Gillessen et al., 2017, Zucker et al., 2006]:

- Use $D^*$ as a free parameter

- Relativistic time-dilation effects

- Use of radial velocities of the $S$-stars

- Motion of Sagittarius $A^*$ over time

- Geometric retardation effects due to the finite speed of light, also called Rømer effect

- Gravitational redshift

However, these disadvantages might be compensated for by the neural network, which provides the flexibility to include these effects. If we want to learn solely the potential from the data, these effects can contaminate our result and increase the uncertainty of the prediction. But as we have seen from the analysis of the data using the fixed potential, where we also neglected these effects, the impact on the predictions will not be as large as the model yielded reasonable results.

Secondly, to be able to better compare our results to other authors, it would be helpful to clarify which reference direction is used for the determination of the rotation angles and also incorporate the radial velocities of the stars to break the degeneracy.

Finally, we should make use of the vast amount of different $S$-stars observed in the central parsec of the Milkyway to train the neural network on multiple trajectories at different distances. This way, the gravitational potential is predicted on a large range of different distances.

# 4 Investigating the Dark Energy Problem

The nature of dark energy is one of the most interesting and fundamental problems in cosmology, because its is not understood very well and still decides the fate of our Universe. There are many competing dark energy theories which make a lot of prior assumptions about its nature, for example by enforcing a certain equation of state or quintessence potential. Using neural potentials, we can turn this around and learn from the data what the best equation of state or quintessence potential would be by modeling it through a neural network. The model is trained on supernova Ia redshift-luminosity data because it is essentially a time-series and the thus the implementation is straightforward and similar to the other presented examples.

## 4.1 Dark Energy and Dynamic Equations of State

The $\Lambda$CDM model assumes that dark energy is modeled by a fluid with a constant equation of state $w = -1$. We can relax this condition by allowing the equation of state to be a free parameter or even to be dependent on the redshift $z$. This dynamic equation of state model is purely phenomenological in the sense that it is able to describe dark energy as a fluid with a varying equation of state, but it does not provide an explanation in terms of fundamental physics. Still, it makes sense to study it as many dark energy theories allow us to interpret the dark energy component as a fluid.

Therefore, the dynamic equation of state model provides a baseline to investigate more sophisticated theories. To analyze it, we take an approach similar to the neural potential, but this time we replace the equation of state of the dark energy fluid by a neural network. Additionally, we assume that there is a non-vanishing matter component present with an equation of state $w = 0$. The radiation component is neglected as the data ranges only up to redshifts $z \approx 7$ which is far away from the radiation-dominated era.

### 4.1.1 Differential equations for Dynamic Equations of State

In order to be able to treat the the equation of state of the dark energy fluid as a neural network, it is necessary to first derive the governing differential equations. First of all, we realize that the distance moduli are parametrized through the redshift. Therefore, the equations are reparametrized through a change of variables form the cosmic time $t$ to the redshift $z$. To do this, we use the relation between redshift and scale factor $a = \frac{1}{1+z}$ and derive it with respect to the cosmic time such that we obtain

$$\dot{z} = -\frac{\dot{a}}{a^2} = -(1 + z)H(z). \tag{4.1}$$

We can use this rule to rewrite derivatives assuming that the functions are sufficiently smooth:

$$\frac{df}{dt} = \dot{z}\frac{df}{dz} = -(1+z)H\frac{dH}{dz} \tag{4.2}$$

Using this relation on the continuity equation and acceleration equation gives two coupled ODEs for the matter density parameter $\Omega_m$ and expansion rate $H$ (for a full derivation see appendix section A.1.1):

$$\frac{d\Omega_m}{dz} = \left(\frac{3}{1+z} - \frac{2}{H}\frac{dH}{dz}\right)\Omega_m \tag{4.3}$$

$$\frac{dH}{dz} = \frac{3}{2}\frac{H}{1+z}\left(\Omega_m + (1+w(z))\Omega_{\mathrm{DE}}\right) \tag{4.4}$$

Together with the assumption of a flat universe $\Omega_{\mathrm{DE}} = 1 - \Omega_m$, solving this system of ODEs gives us $H(z)$, which can be use to calculate the luminosity distance $d_L$, which in turn is required for the calculation of the distance moduli. But for a flat universe, the luminosity distance is equal to the comoving distance $\chi$ [Weinberg, 2008]. The integral expression (1.36) can be converted into a differential equation

$$\frac{d\chi}{dz} = \frac{c}{H}. \tag{4.5}$$

Combining those three equations gives us an ODE system that we can use to train a model to learn the equation of state for the redshift-luminosity data. For this, the equation of state $w(z)$ is replaced with a neural network $\mathrm{NN}(z, \theta)$ with parameters $\theta$. Due to numerical reasons, the expansion rate is normalized through $E = \frac{H}{H_0}$ and the dimensionless comoving distance $\Delta = \frac{H_0\chi}{c}$ is introduced. The entire ODE system then reads:

$$\begin{aligned}
\frac{d\Omega_m}{dz} &= \left(\frac{3}{1+z} - \frac{2}{E}\frac{dE}{dz}\right)\Omega_m \\
\Omega_{\mathrm{DE}} &= 1 - \Omega_m \\
\frac{dE}{dz} &= \frac{3}{2}\frac{E}{1+z}\left(\Omega_m + (1 + \mathrm{NN}(z,\theta))\,\Omega_{\mathrm{DE}}\right) \\
\frac{d\Delta}{dz} &= \frac{1}{E}
\end{aligned} \tag{4.6}$$

Note that this system of ODEs only holds true for a flat universe, i.e. $\Omega_k = 0$.

### 4.1.2 Dynamic Equations of State

With the system of ODEs at hand, it is possible to analyze the data. Before we start with fitting a neural equation of state, we analyze the $w$CDM model and the Chevallier-Polarski-Linder(CPL) parametrization of the equation of state [Chevallier and Polarski, 2001, Linder, 2003].

Both will serve as a test of our gradient-based method and whether it yields reasonable results that are comparable to other state-of-the-art methods like MCMC simulations. They will also serve as a baseline for the neural equation of state model. Both models are very commonly used in literature to analyze supernova Ia data. The $w$CDM model simply treats the dark energy equation of state as a free parameter and the CPL parametrization of the equation of state is probably one of the most used dynamic equation of state for dark energy:

$$w_{\text{CPL}}(z) = w_0 + w_a \frac{z}{1 + z} \tag{4.7}$$

Note that for $w_0 = -1$ and $w_a = 0$ we recover the $\Lambda$CDM model. Typically, supernova Ia data is analyzed using a $\chi^2$-statistic. Thus, it will also serve as a cost function here. Using a NADAM optimizer with a learning rate of 0.01 and training for 5000 epochs, the parameters $w_0$, $w_a$ and $\Omega_{m,0}$ are fitted to the data. The experiment is carried out twice, with random density parameter initializations and with fixed $\Omega_{m,0}$ to see how prior knowledge impacts the accuracy of our model. When initializing the density parameter randomly, it is sampled from a uniform distribution in the interval $[0, 1]$ which corresponds to the possible values for a flat universe.

Using a $\Omega_{m,0}$ derived from the analysis of cosmic microwave background radiation and baryon acoustic oscillations gives us an statistically independent prior. For the random initialization, the parameters of the equation of state are sampled from a uniform distribution in the interval $[-\frac{1}{3}, -2]$. The bound $-\frac{1}{3}$ is due to the fact that we want late time acceleration and the bound $-2$ is set to exclude models with a strong phantom dark energy component to avoid a big rip scenario.

The results are presented in table 4.1.2 and compared to the results of Abbott et al. [2019] and Suzuki et al. [2012]. Both papers use independent samples of supernova Ia data and include different sources of uncertainty in their analyses such that their inferred values sometimes notably disagree. It turns out that our inferred parameters are in good agreement with values from other publications, indicating that we have designed a viable model that yields meaningful results. Figures B.2 and B.3 in the appendix show both of the CPL fits with their respective $2\sigma$-confidence intervals.

The predictions improve greatly for the cases where the matter density parameter $\Omega_m$ is fixed. For the $w$CDM model, the uncertainty reduced by more than a factor of 4. Similarly, the uncertainty of the parameter $w_0$ of the CPL parametrization is reduced by more than 50% while there is almost no change in the confidence limits of the parameter $w_a$. It is already very apparent that the predictions are less accurate and reliable compared to other methods.

We have identified three possible sources of uncertainty for our model:

1. *Ill-defined confidence intervals.* As described in section 1.1.3, the $2\sigma$-confidence intervals are obtained by simply creating an ensemble of experiments and taking the 0.025- and 0.975-quantiles as confidence limits. This is a very crude approximation of a confidence interval that is statistically ill-defined and can only serve as a heuristic.

2. *Proper handling of priors.* The only way to implement priors into the model is by fixing the parameter to a certain value, like we did it for the matter density parameter $\Omega_{m,0}$. This however has the drawback of totally neglecting the uncertainty of the prior. As a result we are not able to propagate the confidence limits of the prior and include it in the prediction of the result like for example for the Fisher-matrix formalism. This leads to an underestimation of the confidence limits for fixed parameters. One possibility to include the priors would be to sample from a Gaussian distribution centered around the prior value with a variance equal to the $1\sigma$-confidence limits. The problem is that we do not know the local shape of the parameter hypersurface and thus the optimizer will perform more or less random steps in the beginning which will distort the initial parameter distribution for an ensemble of experiments. Also, it would be hard to prove that the uncertainties are propagated in a consistent way.

3. *Degeneracy between $\Omega_{m,0}$ and w.* The matter density parameter and the dark energy equation of state are not statistically independent. For larger $\Omega_{m,0}$, the expansion deceleration through gravitational attraction becomes larger, while for smaller $w$ the expansion acceleration through dark energy becomes larger. Thus, we can compensate a larger matter density parameter by decreasing the equation of state parameter. This leads to degeneracies and subsequently to larger uncertainties that our model cannot break as opposed to other methods [Zhang et al., 2019a].

### 4.1.3 Learning the Equation of State

Knowing that our approach gives good results for the CPL parametrization of the dark energy equation of state, we continue by replacing the equation of state $w_{\mathrm{CPL}}(z)$ with a neural network $\mathrm{NN}(z, \theta)$. It has one input and one output node and two fully connected hidden layers with 8 nodes each.

While in most of the other experiments, *ReLU* activation functions were used for the hidden layers, it is impractical to use them here. The reason for this is simple. First of all, the equation of state should satisfy $w(z) < -\frac{1}{3}$ most of the time to ensure late-time cosmic acceleration and it should always stay roughly around $w(z) = -1$ to avoid effects of phantom dark energy, i.e. to have a big rip scenario. As our equation of state depends only on the redshift $z$ and *ReLU* is an positively unbounded function, it is likely that our equation of state will have much larger absolute values for higher redshifts. This can lead to violations of the criteria given above.

| SNIa Sample + Prior (Flat $w$CDM) | $\Omega_{m,0}$ | $w$ | |
|---|---|---|---|
| fixed $\Omega_m$ | 0.311 | $-0.811 \pm 0.067$ | |
| free $\Omega_m$ | $0.330 \pm 0.112$ | $-0.937 \pm 0.239$ | |
| Abbott et al. 2019 | $0.321 \pm 0.018$ | $-0.978 \pm 0.059$ | |
| SCP 2012 | $0.272 \pm 0.013$ | $-1.008 \pm 0.054$ | |

| SNIa Sample + Prior (Flat CPL) | $\Omega_{m,0}$ | $w_0$ | $w_a$ |
|---|---|---|---|
| fixed $\Omega_m$ | 0.311 | $-0.728 \pm 0.095$ | $-0.729 \pm 0.268$ |
| free $\Omega_m$ | $0.309 \pm 0.117$ | $-0.780 \pm 0.246$ | $-0.768 \pm 0.290$ |
| Abbott et al. 2019 | $0.316 \pm 0.011$ | $-0.885 \pm 0.114$ | $-0.387 \pm 0.430$ |
| SCP 2012 | $0.271 \pm 0.013$ | $-1.021 \pm 0.123$ | $-0.07 \pm 0.60$ |

Table 4.1: Results for the $w$CDM and CPL models with 68% confidence limits which were created using 1024 bootstrap repetitions. The first two lines of each table describe our results while the last two lines are results from Abbott et al. [2019] and Suzuki et al. [2012]. We see that the predicted values are in good agreement with the literature, but the uncertainties are bigger, especially for the randomly initialized matter density parameter.

The CPL parametrization (4.7) solves this as it becomes bounded for large redshifts. As a matter of fact, many proposed dynamic equations of state satisfy this property [Davis and Parkinson, 2016]. Therefore, we propose a new activation function called *boundReLU* with the following shape:

$$\text{boundReLU}(x) = \min(\max(x, 0), 1) \tag{4.8}$$

The function is zero for $x < 0$, grows linearly in the interval $[0, 1]$ and equals 1 for all $x > 1$. With this new activation function at hand, the neural equation of state model is fitted to the redshift-luminosity data. Again, we execute the experiments for a fixed and a free density parameter to get an idea of how the uncertainties behave when including prior knowledge. In the case of a free matter density parameter is sampled from the interval $[0, 1]$ while the parameters of the neural network are sampled form a normal distribution when the model is initialized.

Training the neural equation of state model with a NADAM optimizer with a learning rate of 0.01 and a $\chi^2$-statistic as cost function for 5000 epochs give the results presented in figures 4.1 and 4.2. The model with the free matter density parameter predicts

$$\Omega_{m,0} = 0.317 \pm 0.084, \tag{4.9}$$

which agrees well with the earlier results. Also, the uncertainty is much lower compared to the $w$CDM model and CPL parametrization, which is also what we would expect as a model with more parameters is able to capture the data much better. By looking at figures 4.1 and 4.2, we see that using a fixed $\Omega_{m,0}$ again notably reduced the uncertainty of the prediction.

In general, it seems that the neural equation of state model favors an equation of state that is approximately constant at $w(z) \approx -1.5$ for redshifts $z \in [1, 7]$ and then starts to increase towards $w(z) \approx -0.5$. This is a similar behavior to the CPL parametrization and shows that CPL is a valid choice as a dynamic dark energy equation of state. For both cases, the model predicts that the equation of state is always smaller than $-0.5$ with a 95% confidence which leads to late-time acceleration. Also, the equation of state does not show excessive growth for high redshifts such that large contributions from phantom dark energy are in fact avoided.

Even though both models predict a dynamic equation of state that is smaller than $-1$ for low redshifts $z \approx 0.1$, we have to be careful with the result as the uncertainties are not very reliable due to the ill-defined confidence intervals. It is however interesting to see that the model does not prefer a constant equation of state, which it could have provided as well. By further refining the results, the model might be able to make a case for favoring a dynamic equation of state over the $\Lambda$CDM model.
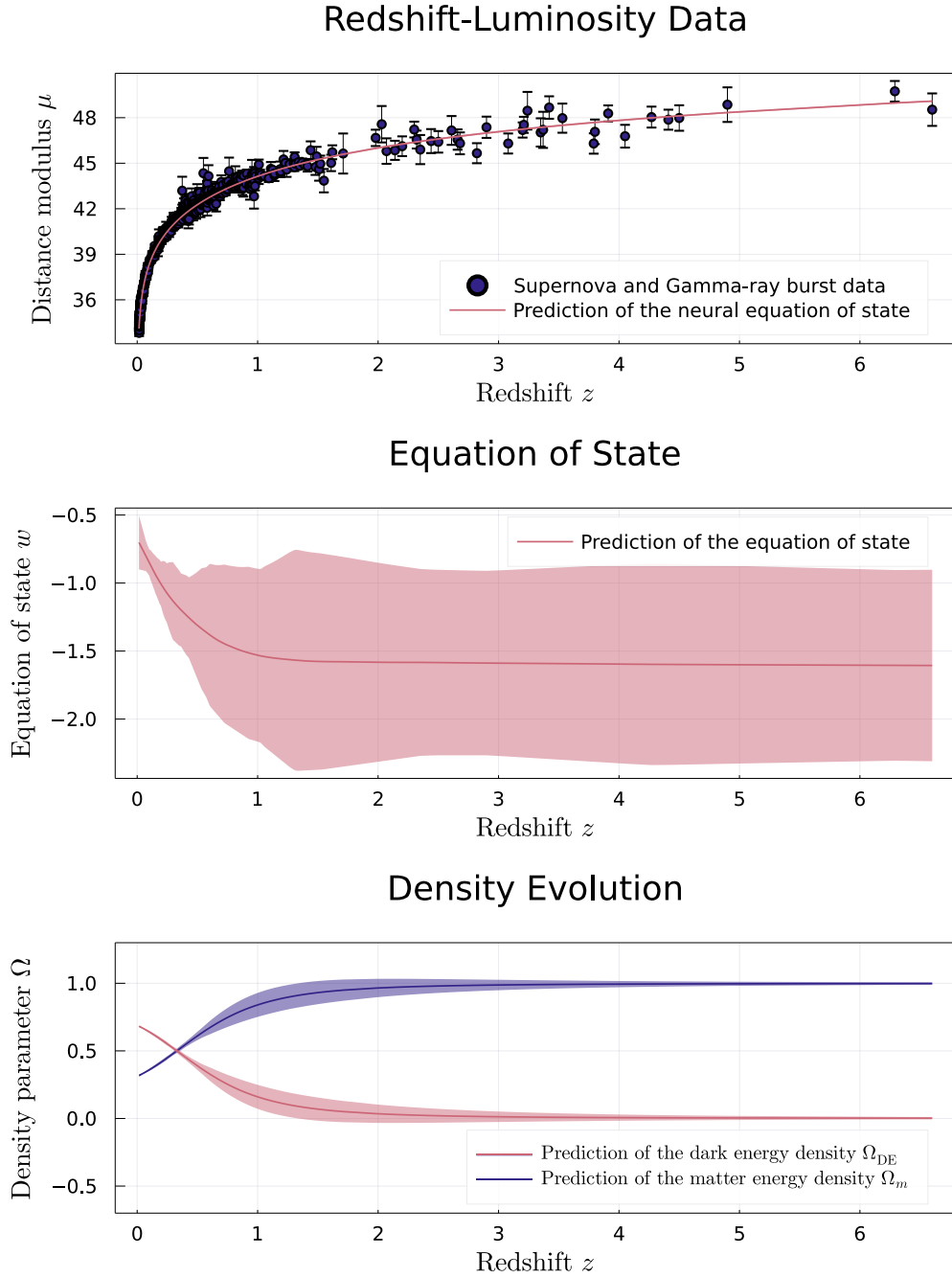
Figure 4.1: Predictions of the neural equation of state model for a fixed matter density parameter of $\Omega_{m,0} = 0.3111$. The model was evaluated for 1024 bootstrap repetitions to generate the $2\sigma$-confidence intervals around the predictions. The uncertainty in the dark energy equation of state decreases significantly in the interval $z \in [0, 0.5]$.
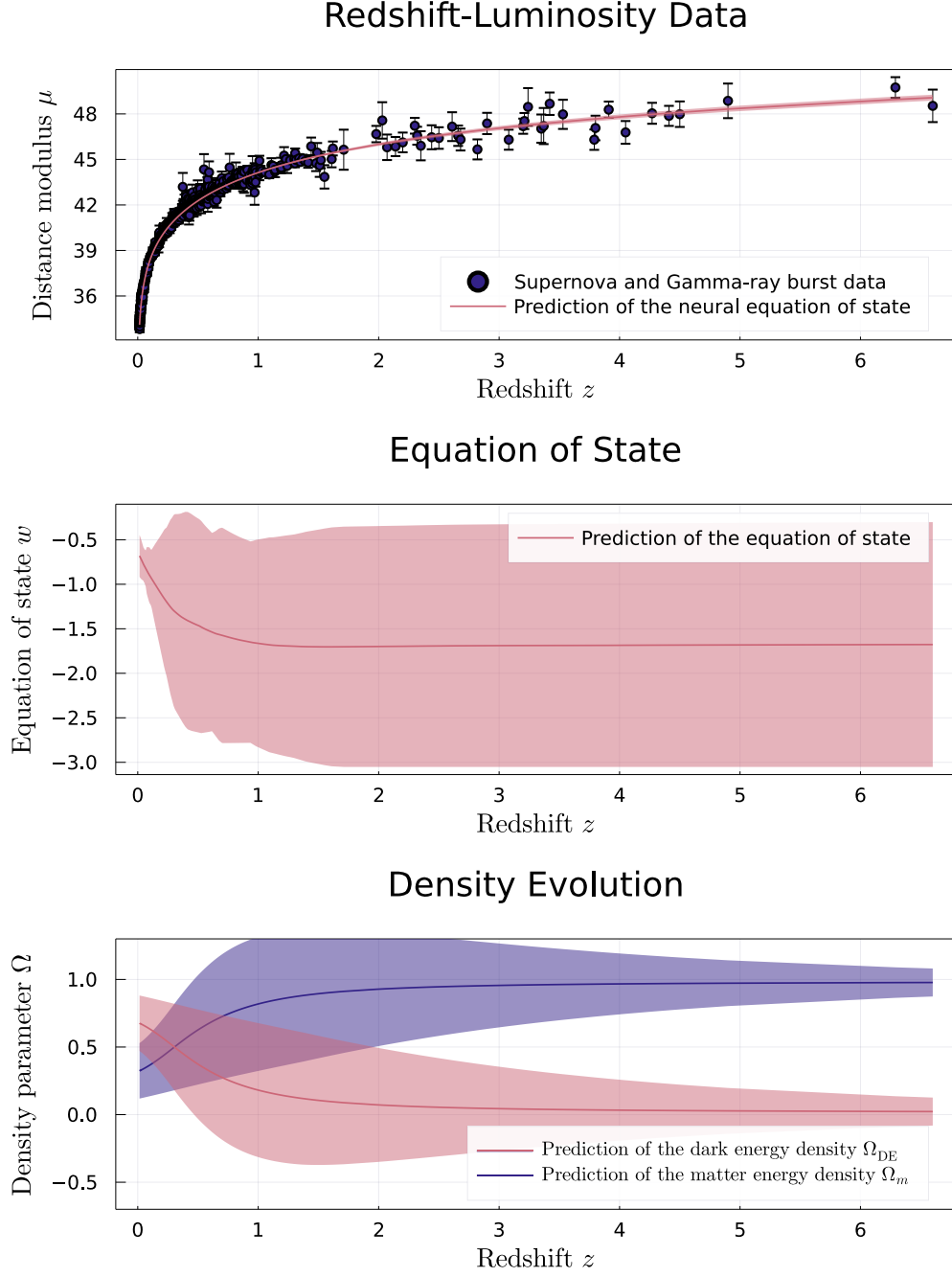
Figure 4.2: Predictions of the neural equation of state model for a free, learnable matter density parameter. The model was evaluated for 1024 bootstrap repetitions to generate the $2\sigma$-confidence intervals around the predictions. It is interesting to observe that the uncertainty of the density parameters grows the largest in the interval $z \in [0.5, 2.5]$. As the confidence intervals are created using quantiles, there had to exist a trajectory that passed through every point within the covered area, i.e. there exist trajectories where the matter density parameter occasionally became less than 0.

## 4.2 Dark Energy and Quintessence

Even though it is interesting to discuss the dynamic equation of state model, the results tell us little about the nature of dark energy itself. To learn more about it, we have to choose among the many competing frameworks that have been used to explain dark energy. One of the most popular and simple explanations is called quintessence, where an additional scalar field is introduced into the action. The physics of this scalar field is defined through its potential, which is also the quantity that is learned from the supernova data.

### 4.2.1 Differential Equations for Quintessence Potentials

To be able to apply the neural potential formalism to the quintessence problem, we again have to reparameterize through the redshift and apply the Friedmann equations as for the neural equation of state formalism. The full calculations can be found in the appendix section A.1.2, which gives

$$\frac{\mathrm{d}^2\phi}{\mathrm{d}z^2} = \left(\frac{2}{1+z} - \frac{1}{H}\frac{\mathrm{d}H}{\mathrm{d}z}\right)\frac{\mathrm{d}\phi}{\mathrm{d}z} - \frac{c^2}{H^2(1+z)^2}\frac{\mathrm{d}V}{\mathrm{d}\phi}, \tag{4.10}$$

$$\frac{\mathrm{d}H}{\mathrm{d}z} = \frac{3H}{2(1+z)}\left(\frac{8\pi G(1+z)^2}{3c^2}\left(\frac{\mathrm{d}\phi}{\mathrm{d}z}\right)^2 + \Omega_m\right). \tag{4.11}$$

Together with $\Omega_m = 1 - \Omega_\phi$, solving this system of ODEs gives functions $H(z)$ and $\phi(z)$ which are used to construct an equation of state and calculate the luminosity distance $d_L$ by integration according to equation (1.36). Last but not least, there are two other quantities we have to take into account. Only the gradient of the potential will be modeled as neural network, but we still require the values of the potential $V$ to be able to determine $\Omega_m$. We can get those through integration over the field $\phi$ which can be transformed into an integration over the redshift $z$ by change of variables according to

$$\int_{\phi(0)}^{\phi(z)} \frac{\mathrm{d}V}{\mathrm{d}\phi}\mathrm{d}\phi = \int_0^z \frac{\mathrm{d}V}{\mathrm{d}\phi}\bigg|_{\phi(z)} \frac{\mathrm{d}\phi}{\mathrm{d}z}\mathrm{d}z. \tag{4.12}$$

This can be rewritten as a differential equation which is added to our system of ODEs:

$$\frac{\mathrm{d}V}{\mathrm{d}z} = \frac{\mathrm{d}V}{\mathrm{d}\phi}\bigg|_{\phi(z)} \frac{\mathrm{d}\phi}{\mathrm{d}z} \tag{4.13}$$

As for the dynamic equation of state, it is necessary to calculate the comoving distance again using the differential equation in (4.5). However, the entire system of ODEs, i.e. (4.5), (4.10), (4.11) and (4.13), has to be solved numerically and therefore requires some minor modifications. First, it should be free of natural constants like $c$, $G$ and $H_0$ as they can cause the numerical values of $H$ and $\phi$ to grow large enough such that numerical truncation errors due to finite machine precision may become relevant.

The other reason is the fact that we are looking at an effect on cosmic scales (the expansion of the Universe) while simultaneously solving an equation of motion of a particle at quantum scales. Making calculations with values from both regimes can also lead numerical inconsistencies. For example, the division of a large quantity by a very small number can lead to an almost infinite answer, breaking the numerical calculation. Thus, we redefine the relevant quantities in the following way:

$$Q = \frac{\sqrt{G}}{c}\phi, \quad U = \frac{G}{H_0^2}V,$$
$$E = \frac{H}{H_0}, \qquad \Delta = \frac{H_0\chi}{c}. \tag{4.14}$$

Again, $E$ is simply the dimensionless Hubble expansion rate and $\Delta$ is the unitless comoving distance. The quantity $U$ will later be replaced by a neural network $\mathrm{NN}(Q, \theta)$. Using the redefinition of the relevant quantities, the resulting system of ODEs looks like

$$\frac{\mathrm{d}^2 Q}{\mathrm{d}z^2} = \left(\frac{2}{1+z} - \frac{1}{E}\frac{\mathrm{d}E}{\mathrm{d}z}\right)\frac{\mathrm{d}Q}{\mathrm{d}z} - \frac{1}{E^2(1+z)^2}\frac{\mathrm{d}U}{\mathrm{d}Q}$$
$$\frac{\mathrm{d}U}{\mathrm{d}z} = \frac{\mathrm{d}U}{\mathrm{d}Q}\bigg|_{Q(z)}\frac{\mathrm{d}Q}{\mathrm{d}z}$$
$$\Omega_m = 1 - \Omega_\phi \tag{4.15}$$
$$\frac{\mathrm{d}E}{\mathrm{d}z} = \frac{3E}{2(1+z)}\left(\frac{8\pi}{3}(1+z)^2\left(\frac{\mathrm{d}Q}{\mathrm{d}z}\right)^2 + \Omega_m\right)$$
$$\frac{\mathrm{d}\Delta}{\mathrm{d}z} = \frac{1}{E}$$

where we have the quintessence density parameter $\Omega_\phi$ which now looks like

$$\Omega_\phi = \frac{8\pi}{3}\left(\frac{(1+z)^2}{2}\left(\frac{\mathrm{d}Q}{\mathrm{d}z}\right)^2 + \frac{U}{E^2}\right). \tag{4.16}$$

Moving from the cosmic time to the redshift parametrization and redefining the systems physical quantities also changes the formulas for the quintessence equation of state:

$$w_\phi(z) = \frac{E^2(1+z)^2\left(\frac{\mathrm{d}Q}{\mathrm{d}z}\right)^2 - U}{E^2(1+z)^2\left(\frac{\mathrm{d}Q}{\mathrm{d}z}\right)^2 + U}, \tag{4.17}$$

## 4.2.2  Freezing and Thawing Models

Before we apply the neural potential approach to the redshift-luminosity data, we want to analyze two popular models for quintessence potentials:

$$V(\phi) = M^4 \left( 1 + \cos\left(\frac{\phi}{f}\right) \right) \tag{4.18}$$

$$V(\phi) = M^4 \phi^{-2} \tag{4.19}$$

The first potential (4.18) arises in a class of models based on the Pseudo-Nambu-Goldstone boson [Amendola and Tsujikawa, 2010] and is an example for a thawing quintessence model, i.e. a model where the field was frozen until recently and has just now started to evolve towards a minimum. It has the free parameters $f$ and $M$ which represent the spontaneous symmetry breaking (SSB) scale and the mass scale respectively.

The second potential (4.19) arises in certain fermion condensate and supergravity models and is an example for a freezing quintessence model, i.e. a model where the field rolls down the potential indefinitely [Binétruy, 1999]. It has the free parameter $M$ which represents the mass scale of the problem.

As we have assumed a flat universe, it is not possible to do experiments where we fix $\Omega_\phi$ to a certain value as this would require a fine-tuning of the initial conditions depending on the potential. Therefore, we can not use prior knowledge from other experiments like *WMAP* or *Planck*.

We train both models with an ADAM optimizer with a learning rate of 0.01 for 10000 epochs using a $\chi^2$-statistic as cost function. The results for both fits are displayed in table 4.2.2 and the figures 4.3 and 4.4 respectively. In the case of the thawing model, we see that the model predicts a very small mass scale $M$ for the particle as well as a very small change in the field amplitude $\frac{d\phi}{dz}$ today. When looking at shapes of the potential and the field evolution, we see that they are not very steep implying that the field is evolving slowly. Still, these results are not very reliable as the uncertainties in the predictions are much larger than the changes in the respective quantities. This also produces big uncertainties in the prediction of the equation of state for $z > 2$.

For smaller redshifts, the model predicts that $\frac{d\phi}{dz} \approx 0$ with a high confidence, such that for $0 < z < 1$, we have $w_\phi \approx -1$ as by equation (1.41). By looking at the shape of the potential and the evolution of the field, we may deduce that the field has only recently begun to evolve. This however would require for the field to have $w_\phi \approx -1$ for all larger redshifts. The uncertainties in this region are so high that we cannot make any clear statements about the equation of state in this region, so that we cannot be sure if our thawing model is even really "thawing".

Although the quintessence-related uncertainties are very high, the model predicts a density evolution very similar to the $\Lambda$CDM model, but with a smaller dark energy component of $\Omega_{\phi,0} = 0.618 \pm 0.049$ with an acceptable confidence. It also captures the the redshift-luminosity data very well.

This shows us that the supernova Ia data is not good choice for in-depth studies of dark energy so that for more accurate predictions we would have to revert to other observational data. Still, the predicted parameters of the model in (4.18) are of the right order as discussed in [Weller and Albrecht, 2001]. We can calculate the mass of the quintessence particle by using the ratio of the two parameters [Amendola and Tsujikawa, 2010]:

$$m_\phi^2 = \frac{M^4}{f^2} \tag{4.20}$$

With this formula and the typical rules for error propagation, we find that the mass is

$$m_\phi = (0.78 \pm 2.30) \cdot 10^{-33} \text{eV}. \tag{4.21}$$

A mass of the order $10^{-33}$eV agrees well with what has been found by other authors [Amendola and Tsujikawa, 2010, Binétruy, 2006, Chung et al., 2003].

In contrast to this, we have the simple freezing model in equation (4.19). The model predicts a mass scale that is approximately twice as high as for the thawing model, but with a much higher uncertainty. As for the thawing model, the neural potential approach predicts initial conditions with a field amplitude of about $10^{-3}$m$_P$ and a vanishing derivative. The models also agree on the initial matter and dark energy densities.

However, the uncertainties in the the prediction of the density parameters are much larger for large redshifts. The predicted equation of state evolves from $w_\phi = 1$ to $w_\phi = -1$. The confidence in this prediction is larger than compared to the thawing model, such that we can reasonably assume that the model actually inferred an equation of state that evolves towards $w_\phi = -1$ as we would expect for a freezing model.

This disagrees with Gerke and Efstathiou [2002], who also used supernova Ia data to fit a potential $V(\phi) \propto \phi^{-2}$ with a different method. They predicted an equation of state of $w_\phi \approx -0.6$ for late times. Considering the large uncertainties of our model, we do not have statistical evidence to disprove these findings. The predictions of the potential and the field amplitude suffer from the same uncertainties as for the thawing model, such that we cannot infer much more about the potential. Thus, it becomes already apparent that the model's prediction of the potential are not very reliable.

### 4.2.3 Learning the Quintessence Potential

Finally, we can set out to learn the quintessence model from the supernova data. By employing a two-layer neural network again with 8 nodes per layer and *tanh* activation functions, the model is trained to give a prediction of the quintessence potential. It is trained using a $\chi^2$-statistic and an ADAM optimizer with a learning rate of 0.01 for 5000 epochs. Figure 4.5 contains the results of the training procedure. The model predicted the following initial conditions, density parameters and $1\sigma$-confidence intervals for the quintessence field:

$$
\begin{aligned}
\phi(z=0) &= (2.73 \pm 1.92) \cdot 10^{-3} \text{m}_P \quad \Omega_{m,0} = 0.450 \pm 0.057 \\
\frac{\mathrm{d}\phi}{\mathrm{d}z}\Big|_{z=0} &= (-0.02 \pm 0.30) \cdot 10^{-3} \text{m}_P \quad \Omega_{\phi,0} = 0.550 \pm 0.057
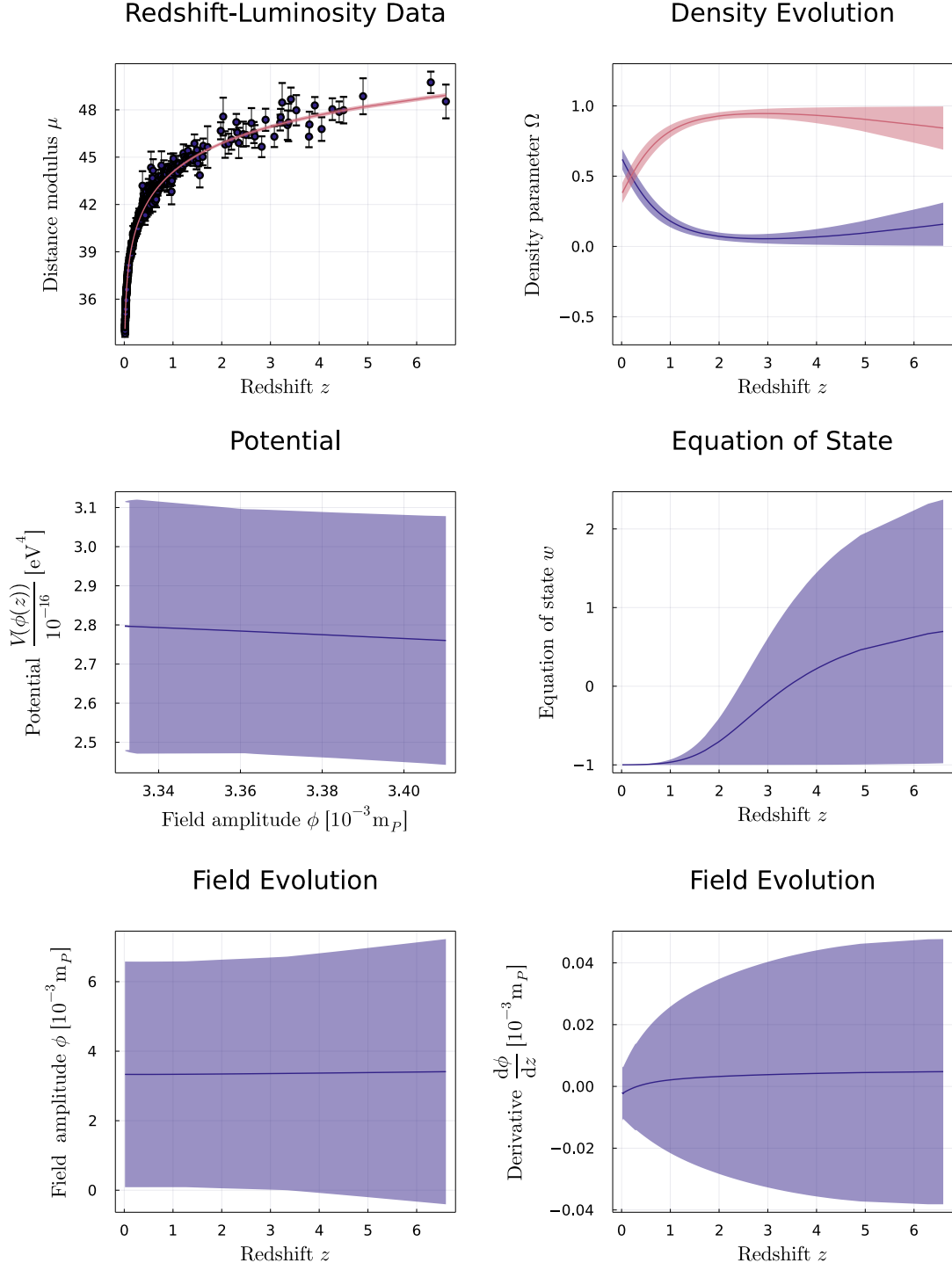\end{aligned} \tag{4.22}
$$

Figure 4.3: This figure shows the results of fitting the thawing model (4.18) to the supernova Ia data. The 95% confidence limits were created by using 1024 bootstrap repetitions. Our model was able to fit the data successfully, but with a high uncertainty. It is however interesting to see, that the it predicts an equation of state $w_\phi \approx -1$ for the late-time evolution $0 < z < 1$.
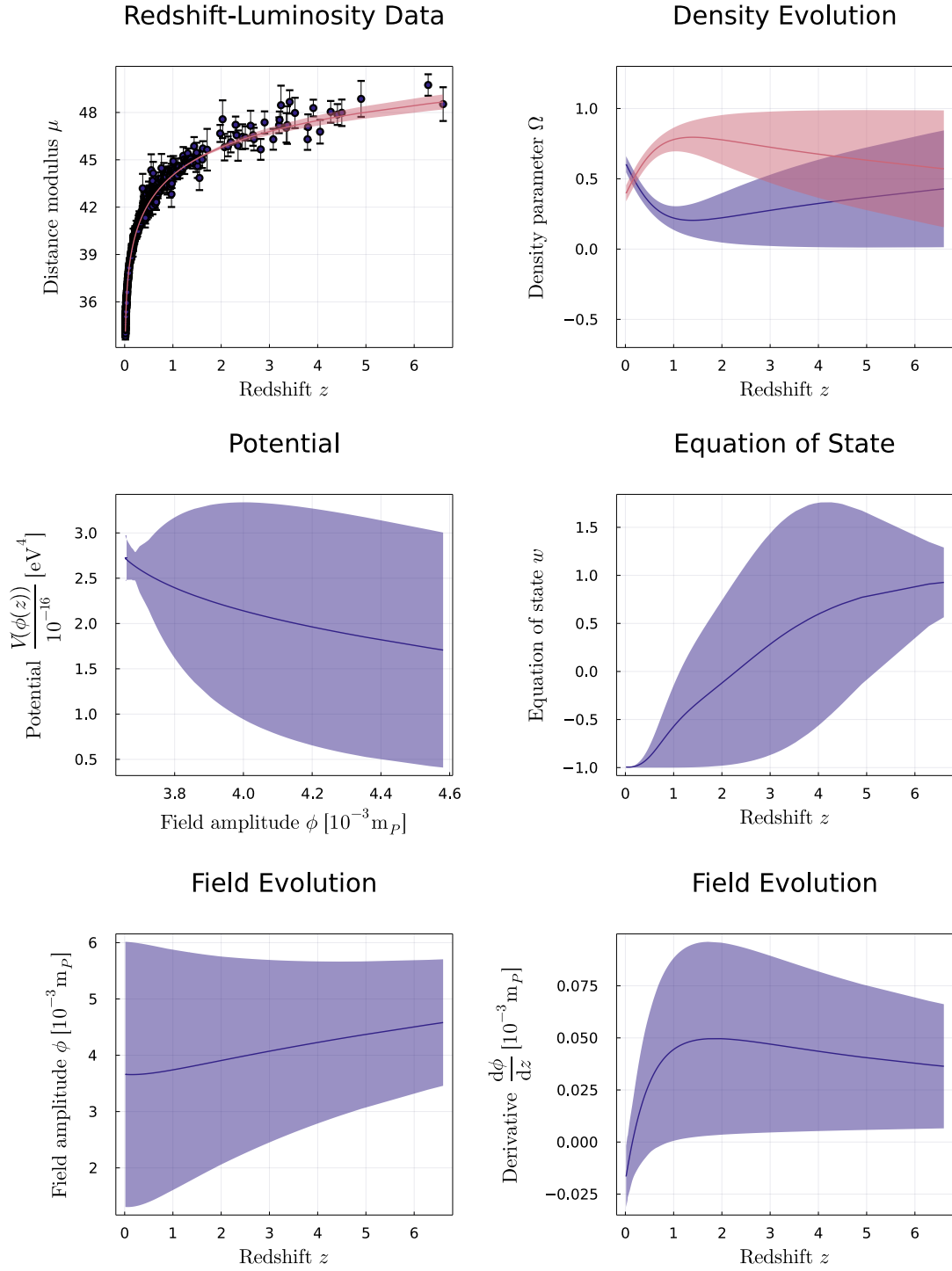
Figure 4.4: In this figure, we see the predictions for a freezing model (4.19). The 95% confidence limits were created by using 1024 bootstrap repetitions. It is particularly interesting to observe the large uncertainty in the density parameters at large redshifts. The model also predicts that the equation of state has only reached $w_\phi \approx 1$ recently.

| Parameter | Thawing Model | Freezing Model |
|---|---|---|
| Initial field amplitude $\phi(z = 0)$ [$10^{-3}$m$_P$] | $0.97 \pm 0.64$ | $1.06 \pm 0.41$ |
| Initial derivative $\left.\dfrac{\mathrm{d}\phi}{\mathrm{d}z}\right\|_{z=0}$ [$10^{-3}$m$_P$] | $-0.002 \pm 0.006$ | $-0.01 \pm 0.01$ |
| Mass scale $M^4$ [$10^{-16}$eV$^4$] | $1.61 \pm 0.35$ | $3.43 \pm 2.23$ |
| Scaling parameter $f^{-1}$ [$10^3$m$_P^{-1}$] | $1.76 \pm 0.95$ | - |
| Matter energy density today $\Omega_{m,0}$ | $0.382 \pm 0.049$ | $0.397 \pm 0.048$ |
| Dark energy density today $\Omega_{\phi,0}$ | $0.618 \pm 0.049$ | $0.603 \pm 0.048$ |

Table 4.2: Predictions of the model for the fixed thawing and freezing models. The 68%-confidence limits were constructed using bootstraps with 1024 repetitions. Both models predict similar initial conditions, that is an initial field amplitude of approximately unity and an almost vanishing derivative. Also both models predict a ratio of 60:40 for the dark energy and dark matter densities. It is however interesting to observe that the mass scales are predicted to be different for both models. However, the uncertainty of the mass scale for the freezing model is high enough to include the mass scales of the thawing model. Note that m$_P$ is the Planck mass.

Even though these values differ from the generally accepted $\Lambda$CDM model predictions, the model is very confident in the prediction of the density parameters. If we look at the potential and the field amplitude of the quintessence field, we are confronted with high uncertainties again. Only for field values of $\phi \approx 2.725 \cdot 10^{-3}$m$_P$ the potential is predicted with a reasonable accuracy. The equation of state again has high uncertainties for large redshifts and evolves down towards $w_\phi = -1$. In contrast to the freezing and thawing models discussed before, we see that the equation of state starts to depart from $-1$ again for very small redshifts.

This behavior can be understood by looking at the evolution of $\frac{\mathrm{d}\phi}{\mathrm{d}z}$, where we see that for the region about $z = 1$ it approaches zero with a very high confidence. By virtue of equation (1.41), this implies $w_\phi \rightarrow -1$. On approaching $z = 0$, $\frac{\mathrm{d}\phi}{\mathrm{d}z}$ seems to depart from zero, but due to the large uncertainties we cannot make any definitive statement. As we have seen before, it is not possible to make any statistically well-founded statements about the potential. With the given accuracy, we cannot even make a clear statement about whether we have a freezing or a thawing model. This significantly limits the power of our approach for this application.
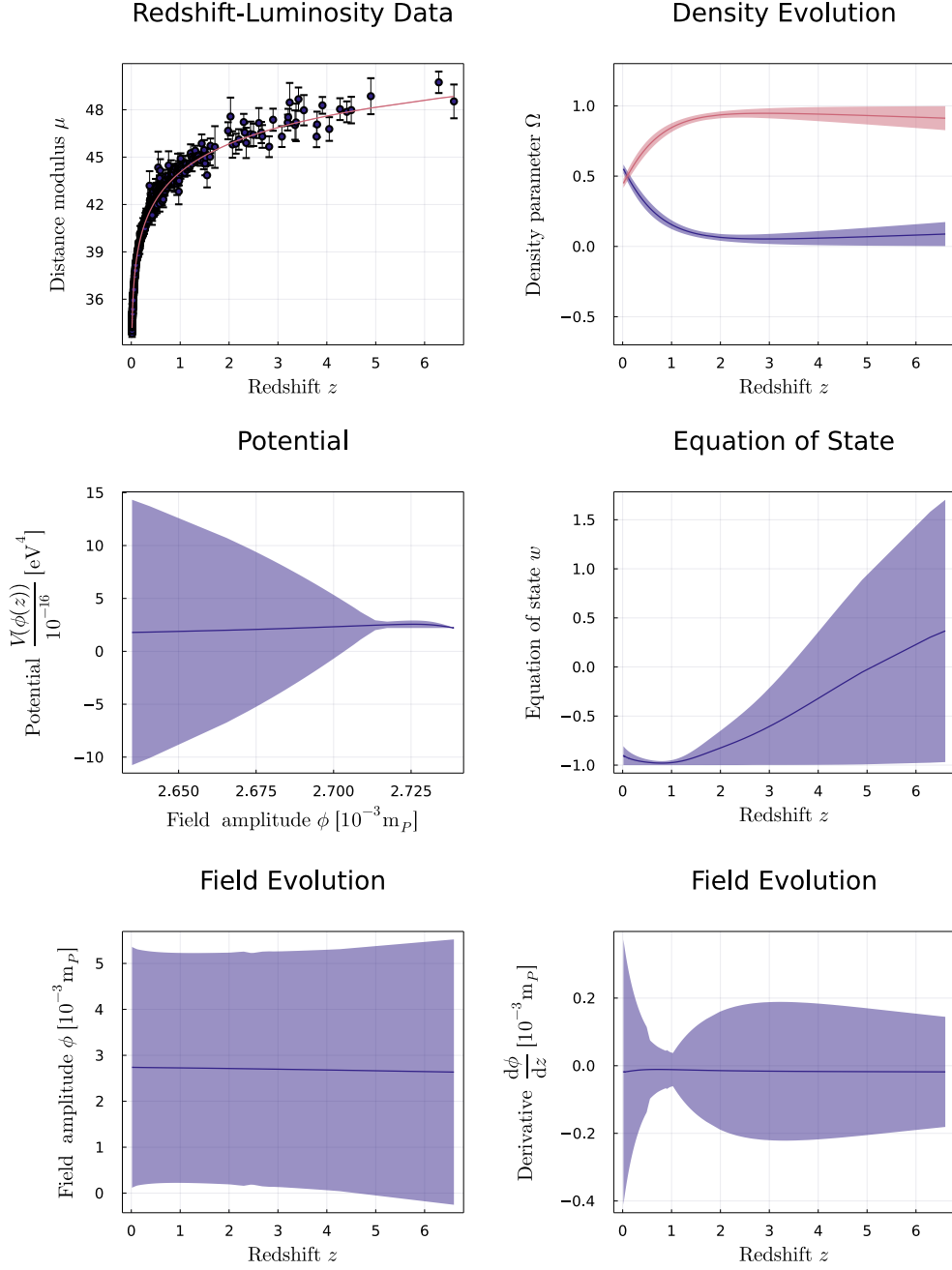
Figure 4.5: The figure shows the predictions of the neural potential for the quintessence model. To create the $2\sigma$-confidence limits, we used 1024 bootstrap repetitions. We see that the density parameters are predicted with high confidence and that the supernova Ia data is captured very well. However, the uncertainty in the prediction of the potential grows very fast for a decreasing field amplitude. It is especially small in the region around $\phi = 2.725 \, [10^{-3} \mathrm{m}_P]$. The equation of state has a low confidence for redshifts $z > 2$. For small $z < 1$, the model predicts a slight derivation from $w_\phi = -1$. The derivative of the field $\phi$ has a particularly low uncertainty in the region around $z = 1$.

### 4.2.4 Dark Energy remains in the Shadows

As we have seen, both the neural equation of state and the neural potential approach yield reasonable results for the equation of state in the sense that at late times, it satisfies approximately $w \approx -1$. Both models even favor a small decrease of the equation of state for $z \to 0$, which is in both cases not statistically significant.

Their predictions differ a lot in when considering the early shape of the equation of state as well as the matter density $\Omega_m$ today. While the neural equation of state predicts an equation of state with $w < -\frac{1}{3}$ at all times, the neural quintessence potential favors a model where te equation of state even becomes positive for early times.

Furthermore, the neural equation of state favors a value of $\Omega_{m,0} = 0.317 \pm 0.084$, while the neural quintessence potential predicts $\Omega_{m,0} = 0.450 \pm 0.057$. We see that the two predictions significantly disagree, but due to the large uncertainties we cannot favor one model over the other one. Thus, the analysis is inconclusive and we have not gained much new knowledge.

However, there is one unique feature that sets the neural quintessence potential method apart from other methods. It predicted the values of a hypothetical quintessence field today to $\phi(z = 0) = (2.73 \pm 1.92) \cdot 10^{-3} \mathrm{m}_P$ and $\frac{\mathrm{d}\phi}{\mathrm{d}z}|_{z=0} = (-0.02 \pm 0.30) \cdot 10^{-3} \mathrm{m}_P$. Although the field amplitude has a very high uncertainty and thus is not very constraining, the prediction of the change of the field amplitude with time is interesting as we find that it is very small, which indicates that we have a field that is rolling slowly today.

It is important to mention, that the large uncertainties are not solely the fault of the model but are also a result of the fact that supernova data is not very sensitive in constraining dark energy. Thus, for future analyses, it is probably more efficient to move to more sensitive methods like weak lensing shear. The problem with data of this type is that the related mathematical models are hard to differentiate and one has to define very complicated adjoints, such that the neural differential approach may runs in other problems.

# 5 Conclusion and Outlook

The presented results look very promising, indicating that the method of neural potentials can reproduce existing knowledge and even give new insights on open problems. The analysis of the harmonical oscillator and the W-shaped potential has shown that the model is able to capture intricate features of dynamical motion with high confidence even for a small dataset. It is also able to correctly extract the underlying physical laws and initial conditions and is thus able to reproduce similar trajectories.

In particular, the model makes very accurate predictions even for large time intervals and might be even superior to RNNs under certain circumstances. This suggests that the method of neural potentials could be applicable to control problems coming from dynamical systems research and robotics. Especially the ability to implement physical priors into the machine learning model is a big advantage of this method. As neural potentials are able to deal with time-series data and ordinary differential equations in a well-defined way, it might also be interesting for other research areas where these topics are relevant, for example weather forecasting and finance.

However, our model only truly unfolds its full potential in what one would describe as *data-driven theoretical physics research*. The idea is to take a dataset, construct a model for it by using already known physical constraints and equations and replace the unknown parts by a neural network. Then the model is trained on this data and provides a prediction for a possible theory. Thus, the data dictates the physics and we let the neural network fill out the blanks. This was exactly the idea we had in mind when we were discussing the Kepler problem and investigated dark energy theories.

In our analysis of the Kepler problem, we saw that we reconstructed the potential from a synthetic dataset as well as from real world data. By looking at the orbit of the star $S2$ around Sagittarius $A^*$, we not only inferred the Newtonian potential but also its relativistic correction. In a similar manner, the analysis of the supernova data gave us an idea of how a possible dynamic equation of state could look like. By assuming that dark energy is the result of a new scalar field, the model even predicted the potential of this hypothetical quintessence particle.

On the downside, all these predictions have very large uncertainties. In most cases, the learned potential does not offer any statistically significant evidence for the inferred results. This becomes particularly apparent when one looks at the results of the neural equation of state and the neural quintessence potential. Although both models predict a slight derivation from $w = -1$ for late times, both have a very high uncertainty and are therefore not very reliable. In relation to this stands the fact that for this method there is no known way to quantify uncertainties in a mathematically well-defined way, let alone describe how priors from other experiments affect the confidence of the prediction like the

Fisher matrix does. Furthermore, the models capabilities are limited to ordinary differential equations only such that it is not able to tackle many problems in modern physics or only with great simplifications. Even if we have a prediction of some unknown quantity, we have it in form of a neural network and not as a closed form expression. This offers only low flexibility for other applications and thus does not generalize very well. Still, the use of symbolic regression may circumvent these issues.

In addition to these conceptual problems, there are various technical problems that sometimes significantly reduce the models performance. Fist of all, for our usecase Julia is not the optimal choice. Although it is blazing fast in computing the derivatives and adjoints, it lacks the ability to differentiate through derivatives of functions. As a result, we cannot replace the potential through a neural network but instead have to approximate the gradient of the potential. The potential was then reconstructed by integrating the network, which limits the approach to one-dimensional problems.

A solution to this may be the use of the *JAX* framework developed by Google [Bradbury et al., 2018]. Secondly, everything has to be differentiable. This especially problematic when differentiating through integral expressions or array mutations because then the derivatives are often not well-defined.

To summarize everything, we have shown that neural potentials have the ability to reconstruct existing knowledge, but at its current state the predictions are not competetive to other machine learning methods. Furthermore, it remains to be seen if the model can provide statistically significant discoveries. This will require an in-depth analysis of its statistical properties.

# Acknowledgments

# Other Resources

The Github repository that contains all the related Julia code is provided under:
`https://github.com/jamiegrieser/NeuralPotentials`

# A Derivations

## A.1 Redshift-parameterized ODE System

This section gives a full derivation of the equations used in chapter 4. The central aspect for this derivation is the equation

$$\dot{z} = -(1+z)H(z) \tag{A.1}$$

and the assumption that the functions are smooth enough such that a change of coordinates is possible everywhere.

### A.1.1 Dynamic Equations of State

**Conservation Equation**

The definition of the matter density parameter with the critical density $\rho_c(z) = \frac{3H^2(z)}{8\pi G}$, reads

$$\Omega_m(z) = \frac{\rho_m(z)}{\rho_c(z)} = \frac{8\pi G}{3H^2(z)}\rho_m(z). \tag{A.2}$$

We can derive the ODE for the matter density parameter by deriving with respect to $z$:

$$\frac{d\Omega_m}{dz} = \frac{8\pi G}{3}\left(\frac{1}{H^2}\frac{d\rho_m}{dz} - \frac{2\rho_m}{H^3}\frac{dH}{dz}\right) \tag{A.3}$$

The conservation equation (1.33) is expressed in terms of redshift using equation (A.1):

$$\frac{d\rho_m}{dz} = \frac{3(1+w)}{1+z}\rho_m \tag{A.4}$$

This allows us to rewrite the equation for the matter density parameter into

$$\begin{aligned}
\frac{d\Omega_m}{dz} &= \frac{8\pi G}{3H^2}\left(\frac{3}{1+z}\rho_m - \frac{2\rho_m}{H}\frac{dH}{dz}\right) \\
&= \left(\frac{3}{1+z} - \frac{2}{H}\frac{dH}{dz}\right)\Omega_m.
\end{aligned} \tag{A.5}$$

where we have set $w = 0$ as the equation of state for the matter component.

**Acceleration Equation**

We plug the Friedman equation (1.30) into the acceleration equation (1.31). This yields

$$\dot{H} = -4\pi G \left( \rho_m + \rho_{\rm DE} + \frac{p_{\rm DE}}{c^2} \right). \tag{A.6}$$

Moving from cosmic time $t$ to redshift $z$ and using a dynamic equation of state for the dark energy component $p_{\rm DE} = w_{\rm DE}(z)\rho_{\rm DE}c^2$ gives

$$\begin{aligned}
\frac{{\rm d}H}{{\rm d}z} &= \frac{4\pi G}{(1+z)H}(\rho_m + (1+w(z))\rho_{\rm DE}) \\
&= \frac{3}{2}\frac{H}{1+z}(\Omega_m + (1+w(z))\Omega_{\rm DE})
\end{aligned} \tag{A.7}$$

where $\Omega_{\rm DE}$ is the energy density of the dark energy fluid.

## A.1.2  Quintessence Models

**Acceleration Equation for Quintessence**

We start from equation (1.42), and again plug the Friedman equation into the acceleration equation. The result is similar to (A.6):

$$\dot{H} = -4\pi G\left(\rho_m + \rho_\phi + \frac{p_\phi}{c^2}\right) = -4\pi G\left(\frac{\dot{\phi}^2}{c^2} + \rho_m\right) \tag{A.8}$$

Making a change of variables from cosmic time $t$ to redshift $z$ and using the rule in (A.1) allows us to rewrite the equation as

$$(1+z)H\frac{{\rm d}H}{{\rm d}z} = 4\pi G\left(\frac{H^2(1+z)^2}{c^2}\left(\frac{{\rm d}\phi}{{\rm d}z}\right)^2 + \rho_m\right). \tag{A.9}$$

Using the definition of the matter density parameter, this can be rearranged to look like

$$\frac{{\rm d}H}{{\rm d}z} = \frac{3H}{2(1+z)}\left(\frac{8\pi G(1+z)^2}{3c^2}\left(\frac{{\rm d}\phi}{{\rm d}z}\right)^2 + \Omega_m\right). \tag{A.10}$$

**Quintessence Field Equation**

The derivation of the redshift-parameterized quintessence field equation works in a similar way. We have a second-order time derivative which can be converted to a second-order redshift derivative plus some other terms by making use of the chain rule:

$$\begin{aligned}
\ddot{\phi} &= (1+z)H\frac{{\rm d}}{{\rm d}z}\left\{(1+z)H\frac{{\rm d}\phi}{{\rm d}z}\right\} \\
&= (1+z)H^2\frac{{\rm d}\phi}{{\rm d}z} + (1+z)^2H\frac{{\rm d}H}{{\rm d}z}\frac{{\rm d}\phi}{{\rm d}z} + (1+z)^2H^2\frac{{\rm d}^2\phi}{{\rm d}z^2}
\end{aligned} \tag{A.11}$$

Together with $\dot{\phi} = -(1 + z)H\frac{\mathrm{d}\phi}{\mathrm{d}z}$ and dividing by $(1 + z)^2 H^2$, we can then rewrite the quintessence field equation (1.39) as

$$\frac{1}{1 + z}\frac{\mathrm{d}\phi}{\mathrm{d}z} + \frac{1}{H}\frac{dH}{dz}\frac{d\phi}{dz} + \frac{\mathrm{d}^2\phi}{\mathrm{d}z^2} - \frac{3}{1 + z}\frac{\mathrm{d}\phi}{\mathrm{d}z} + \frac{c^2}{(1 + z)^2 H^2}\frac{\mathrm{d}\phi}{\mathrm{d}z} = 0. \tag{A.12}$$

Rearranging for $\frac{\mathrm{d}^2\phi}{\mathrm{d}z^2}$ gives the final equation:

$$\frac{\mathrm{d}^2\phi}{\mathrm{d}z^2} = \left(\frac{2}{1 + z} - \frac{1}{H}\frac{\mathrm{d}H}{\mathrm{d}z}\right)\frac{\mathrm{d}\phi}{\mathrm{d}z} - \frac{c^2}{(1 + z)^2 H^2}\frac{\mathrm{d}V}{\mathrm{d}\phi} \tag{A.13}$$
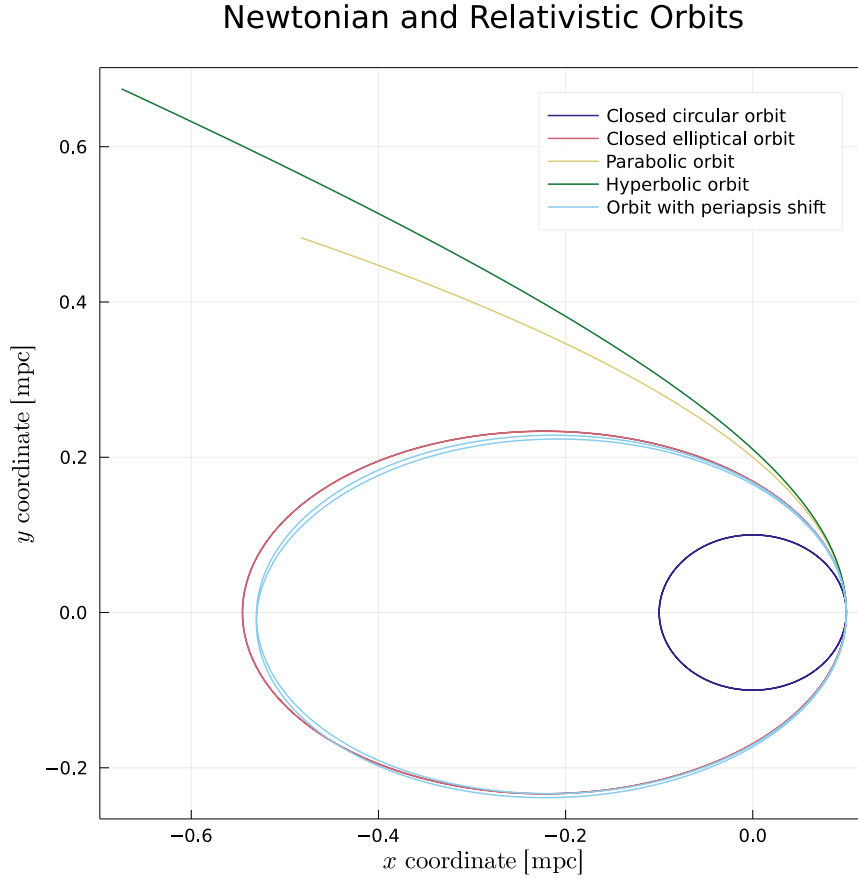
# B Figures



Figure B.1:  This figure gives an overview over the different possible orbits for an object orbiting a central mass. The data was created by assuming that a star orbits a supermassive blackhole of mass $M = 4.35 \cdot 10^6 M_\odot$ with a periapsis distance of 0.1 mpc. The red plot is the elliptical orbit of the star as predicted by Newton's gravitational law, while the light bluish plot shows the same orbit with the relativistic correction for two subsequent revelations.´
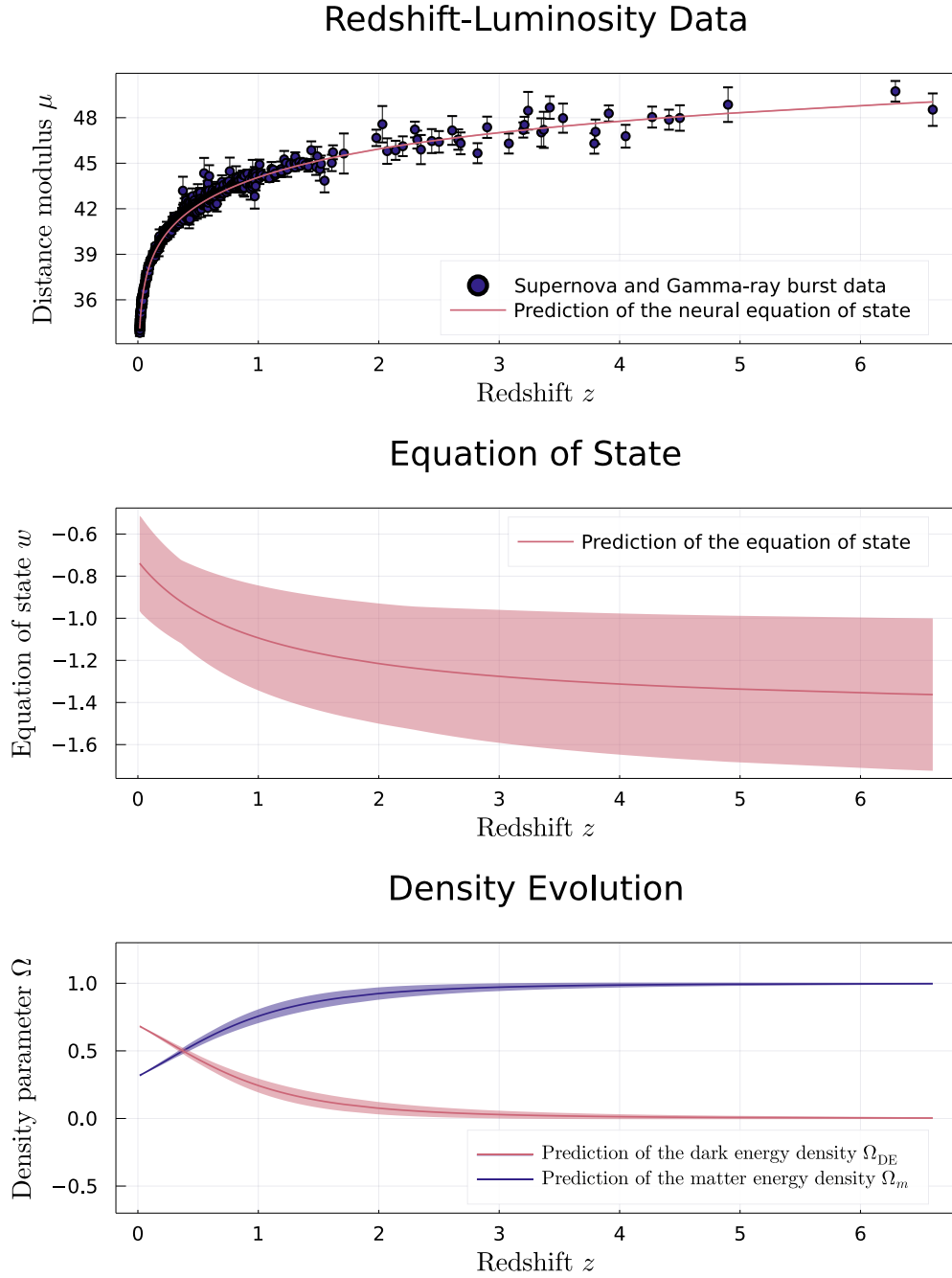
## Redshift-Luminosity Data



## Equation of State



## Density Evolution



Figure B.2: The figure shows the results for the CPL parametrization fit with a fixed matter density parameter $\Omega_{m,0} = 0.3111$. By using 1024 bootstrap repetitions, the $2\sigma$-confidence intervals were created around the prediction. The model puts very tight constraints on the evolution of the matter density parameter and the equation of state.
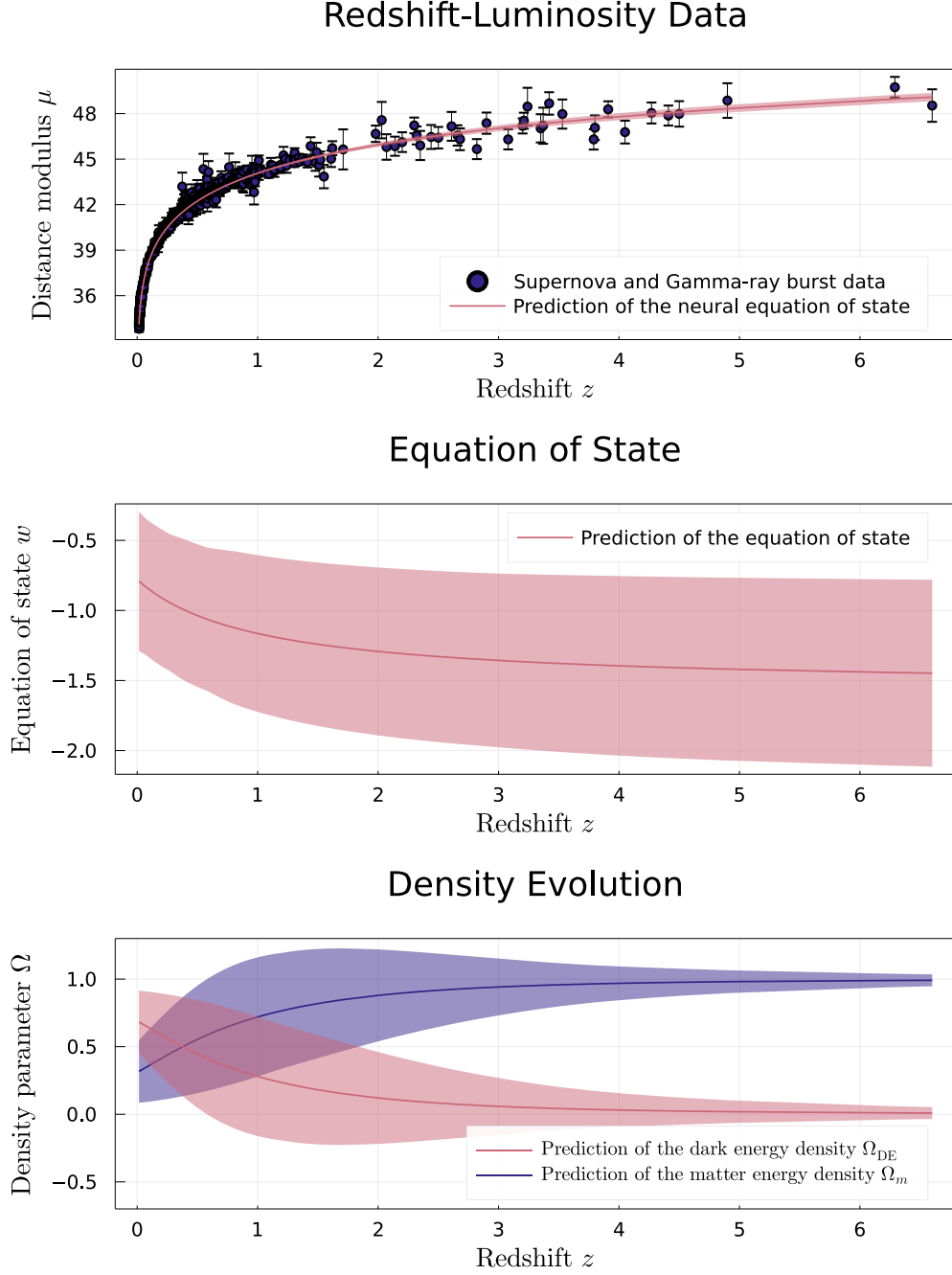
## Redshift-Luminosity Data

Distance modulus $\mu$ vs Redshift $z$

● Supernova and Gamma-ray burst data
— Prediction of the neural equation of state

## Equation of State

Equation of state $w$ vs Redshift $z$

— Prediction of the equation of state

## Density Evolution

Density parameter $\Omega$ vs Redshift $z$

— Prediction of the dark energy density $\Omega_{DE}$
— Prediction of the matter energy density $\Omega_m$

Figure B.3: The figure shows the results for the CPL parametrization fit with a free, learnable matter density parameter. By again using 1024 bootstrap repetitions, the $2\sigma$-confidence intervals were created. Comparing the models prediction to the case where $\Omega_{m,0}$ is fixed, we see that uncertainties have grow, particularly for the evolution of the density parameters. The model predicted a matter density parameter of $\Omega_{m,0} = 0.309 \pm 0.011$.

# C Lists

## C.1 List of Figures

## C.2 List of Tables

# D  Bibliography

Kepler orbital elements. `https://commons.wikimedia.org/wiki/File:Orbit1.svg`. Accessed: 2021-10-18.

Declination and right ascension. `https://skyandtelescope.org/wp-content/uploads/RA-Dec-wiki-Tom-RuenCC-BY-SA-3.0.jpg`. Accessed: 2021-10-28.

T. M. C. Abbott, S. Allam, P. Andersen, C. Angus, J. Asorey, A. Avelino, S. Avila, B. A. Bassett, K. Bechtol, G. M. Bernstein, E. Bertin, D. Brooks, D. Brout, P. Brown, D. L. Burke, J. Calcino, A. Carnero Rosell, D. Carollo, M. Carrasco Kind, J. Carretero, R. Casas, F. J. Castander, R. Cawthon, P. Challis, M. Childress, A. Clocchiatti, C. E. Cunha, C. B. D'Andrea, L. N. da Costa, C. Davis, T. M. Davis, J. De Vicente, D. L. DePoy, S. Desai, H. T. Diehl, P. Doel, A. Drlica-Wagner, T. F. Eifler, A. E. Evrard, E. Fernandez, A. V. Filippenko, D. A. Finley, B. Flaugher, R. J. Foley, P. Fosalba, J. Frieman, L. Galbany, J. García-Bellido, E. Gaztanaga, T. Giannantonio, K. Glazebrook, D. A. Goldstein, S. González-Gaitán, D. Gruen, R. A. Gruendl, J. Gschwend, R. R. Gupta, G. Gutierrez, W. G. Hartley, S. R. Hinton, D. L. Hollowood, K. Honscheid, J. K. Hoormann, B. Hoyle, D. J. James, T. Jeltema, M. W. G. Johnson, M. D. Johnson, E. Kasai, S. Kent, R. Kessler, A. G. Kim, R. P. Kirshner, E. Kovacs, E. Krause, R. Kron, K. Kuehn, S. Kuhlmann, N. Kuropatkin, O. Lahav, J. Lasker, G. F. Lewis, T. S. Li, C. Lidman, M. Lima, H. Lin, E. Macaulay, M. A. G. Maia, K. S. Mandel, M. March, J. Marriner, J. L. Marshall, P. Martini, F. Menanteau, C. J. Miller, R. Miquel, V. Miranda, J. J. Mohr, E. Morganson, D. Muthukrishna, A. Möller, E. Neilsen, R. C. Nichol, B. Nord, P. Nugent, R. L. C. Ogando, A. Palmese, Y.-C. Pan, A. A. Plazas, M. Pursiainen, A. K. Romer, A. Roodman, E. Rozo, E. S. Rykoff, M. Sako, E. Sanchez, V. Scarpine, R. Schindler, M. Schubnell, D. Scolnic, S. Serrano, I. Sevilla-Noarbe, R. Sharp, M. Smith, M. Soares-Santos, F. Sobreira, N. E. Sommer, H. Spinka, E. Suchyta, M. Sullivan, E. Swann, G. Tarle, D. Thomas, R. C. Thomas, M. A. Troxel, B. E. Tucker, S. A. Uddin, A. R. Walker, W. Wester, P. Wiseman, R. C. Wolf, B. Yanny, B. Zhang, and Y. Zhang and. First cosmology results using type Ia supernovae from the dark energy survey: Constraints on cosmological parameters. *ApJ*, 872(2):L30, feb 2019. doi: 10.3847/2041-8213/ab04fa. URL `https://doi.org/10.3847/2041-8213/ab04fa`.

Ronald J. Adler, Brendan Casey, and Ovid C. Jacob. Vacuum catastrophe: An elementary exposition of the cosmological constant problem. *American Journal of Physics*, 63(7): 620–626, July 1995. doi: 10.1119/1.17850.

Luca Amendola and Shinji Tsujikawa. *Dark Energy*. Camebridge University Press, Camebridge, MA, 2010. ISBN 978-0-521-51600-6.

Matthias Bartelmann. *Theoretische Physik*. Springer Spektrum, Heidelberg, D, 2015. ISBN 978-3-642-54617-4.

Atılım Güneş Baydin, Barak A. Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind. Automatic differentiation in machine learning: A survey. *J. Mach. Learn. Res.*, 18(1):5595–5637, January 2017. ISSN 1532–4435.

Jeff Bezanson, Stefan Karpinski, Viral B. Shah, and Alan Edelman. Julia: A fast dynamic language for technical computing. *CoRR*, abs/1209.5145, 2012. URL http://arxiv.org/abs/1209.5145.

J.P.M. Binet. Mémoire sur un systeme de formules analytiques, et leur application à des considerations géométriques. *J. de l'École Polytechnique IX*, 16:280 – 302, 1813.

Pierre Binétruy. Models of dynamical supersymmetry breaking and quintessence. *Physical Review D*, 60:063502, 1999.

Pierre Binétruy. Course 9 - particle astrophysics and cosmology. In Dmitri Kazakov, Stéphane Lavignac, and Jean Dalibard, editors, *Particle Physics Beyond the Standard Model*, volume 84 of *Les Houches*, pages 457–536. Elsevier, 2006. doi: https://doi.org/10.1016/S0924-8099(06)80029-6. URL https://www.sciencedirect.com/science/article/pii/S0924809906800296.

James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. JAX: composable transformations of Python+NumPy programs, 2018. URL http://github.com/google/jax.

Robert R. Caldwell and Marc Kamionkowski. The physics of cosmic acceleration. *Annual Review of Nuclear and Particle Science*, 59(1):397–429, 2009. doi: 10.1146/annurev-nucl-010709-151330. URL https://doi.org/10.1146/annurev-nucl-010709-151330.

Robert R. Caldwell, Marc Kamionkowski, and Nevin N. Weinberg. Phantom Energy: Dark Energy with w<-1 Causes a Cosmic Doomsday. *PRL*, 91(7):071301, August 2003. doi: 10.1103/PhysRevLett.91.071301.

Augustin-Louis Cauchy. Méthode générale pour la résulution des systèmes d'équations simultanées. *Compte Rendu a l'Academie des Sciènces*, 1847.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper/2018/file/69386f6bb1dfed68692a24c8686939b9-Paper.pdf.

Michel Chevallier and David Polarski. Accelerating universes with scaling dark matter. *Int. J. Mod. Phys. D*, 10:213–224, 2001. doi: 10.1142/S0218271801000822.

D.J.H. Chung, L.L. Everett, and A. Riotto. Quintessence and the underlying particle physics theory. *Physics Letters B*, 556(1):61–70, 2003. ISSN 0370-2693. doi: https://doi.org/10.1016/S0370-2693(03)00099-6. URL https://www.sciencedirect.com/science/article/pii/S0370269303000996.

William G. Cochran. The $\chi^2$ Test of Goodness of Fit. *The Annals of Mathematical Statistics*, 23(3):315 – 345, 1952. doi: 10.1214/aoms/1177729380. URL https://doi.org/10.1214/aoms/1177729380.

S. A. Colgate. Supernovae as a standard candle for cosmology. *ApJ*, 232:404–408, September 1979. doi: 10.1086/157300.

M. Cranmer, Sam Greydanus, Stephan Hoyer, Peter W. Battaglia, David N. Spergel, and Shirley Ho. Lagrangian neural networks. *ArXiv*, abs/2003.04630, 2020.

Tamara S. Davis and David Parkinson. *Handbook of Supernovae*, chapter Characterizing Dark Energy Through Supernovae, pages 1–23. Springer, Heidelberg, D, 2016.

Scott Dodelson and Fabian Schmidt. *Modern Cosmology (Second Edition)*. Academic Press, second edition edition, 2021. ISBN 978-0-12-815948-4. doi: https://doi.org/10.1016/B978-0-12-815948-4.00006-1.

Timothy Dozat. Incorporating nesterov momentum into adam. In *ICLR 2016 Workshop Proceedings*, 2016. URL https://openreview.net/pdf?id=OM0jvwB8jIp57ZJjtNEZ.

A. Eckart and R. Genzel. Observations of stellar proper motions near the Galactic Centre. *Nature*, 383(6599):415–417, October 1996. doi: 10.1038/383415a0.

B. Efron and R. Tibshirani. Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science*, 1(1):54 – 75, 1986.

Frank Emmert-Streib, Zhen Yang, Han Feng, Shailesh Tripathi, and Matthias Dehmer. An introductory review of deep learning for prediction models with big data. *Frontiers in Artificial Intelligence*, 3:4, 2020. ISSN 2624–8212. doi: 10.3389/frai.2020.00004. URL https://www.frontiersin.org/article/10.3389/frai.2020.00004.

Torsten Fließbach. *Statistische Physik*. Springer Spektrum, 5. auflage edition, 2010. ISBN 978-3-8274-2527-0.

Brian F. Gerke and G. Efstathiou. Probing quintessence: reconstruction and parameter estimation from supernovae. *MNRAS*, 335(1):33–43, September 2002. doi: 10.1046/j.1365-8711.2002.05612.x.

A. M. Ghez, B. L. Klein, M. Morris, and E. E. Becklin. High proper-motion stars in the vicinity of sagittarius a: Evidence for a supermassive black hole at the center of our galaxy. *The Astrophysical Journal*, 509(2):678–686, dec 1998. doi: 10.1086/306528. URL https://doi.org/10.1086/306528.

S. Gillessen, P. M. Plewa, F. Eisenhauer, R. Sari, I. Waisberg, M. Habibi, O. Pfuhl, E. George, J. Dexter, S. von Fellenberg, T. Ott, and R. Genzel. An Update on Monitoring Stellar Orbits in the Galactic Center. *ApJ*, 837(1):30, March 2017. doi: 10.3847/1538-4357/aa5c41.

GRAVITY-Collaboration. First light for GRAVITY: Phase referencing optical interferometry for the Very Large Telescope Interferometer. *AAP*, 602:A94, June 2017. doi: 10.1051/0004-6361/201730838.

GRAVITY-Collaboration. A geometric distance measurement to the galactic center black hole with 0.3% uncertainty. *AAP*, 625:L10, 2019. doi: 10.1051/0004-6361/201935656. URL https://doi.org/10.1051/0004-6361/201935656.

J. Guy, M. Sullivan, A. Conley, N. Regnault, P. Astier, C. Balland, S. Basa, R. G. Carlberg, D. Fouchez, D. Hardin, I. M. Hook, D. A. Howell, R. Pain, N. Palanque-Delabrouille, K. M. Perrett, C. J. Pritchet, J. Rich, V. Ruhlmann-Kleider, D. Balam, S. Baumont, R. S. Ellis, S. Fabbro, H. K. Fakhouri, N. Fourmanoit, S. González-Gaitán, M. L. Graham, E. Hsiao, T. Kronborg, C. Lidman, A. M. Mourao, S. Perlmutter, P. Ripoche, N. Suzuki, and E. S. Walker. The Supernova Legacy Survey 3-year sample: Type Ia supernovae photometric distances and cosmological constraints. *AAP*, 523:A7, November 2010. doi: 10.1051/0004-6361/201014468.

M. P. Hobson, G. P. Efstathiou, and A. N. Lasenby. *General Relativity: An Introduction for Physicists*. Cambridge University Press, 2006. doi: 10.1017/CBO9780511790904.

Edwin Hubble. A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the National Academy of Sciences*, 15(3):168–173, 1929. ISSN 0027-8424. doi: 10.1073/pnas.15.3.168. URL https://www.pnas.org/content/15/3/168.

Michael Innes. Don't unroll adjoint: Differentiating ssa-form programs, 2019.

Saurabh Jha, Adam G. Riess, and Robert P. Kirshner. Improved distances to type ia supernovae with multicolor light-curve shapes: MLCS2k2. *The Astrophysical Journal*, 659(1):122–148, apr 2007. doi: 10.1086/512054. URL https://doi.org/10.1086/512054.

A. Khokhlov, E. Mueller, and P. Hoeflich. Light curves of type IA supernova models with different explosion mechanisms. *AAP*, 270:223–248, March 1993.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning*

*Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. URL http://arxiv.org/abs/1412.6980.

E. Kolb and M. Turner. *The Early Universe*. Addison-Wesley, Redwood City, CA, 1990. ISBN 978-0-201-62674-2.

Vera Kurkova. Kolmogorov's Theorem and Multilayer Neural Networks. *Neural Networks*, 5:501–506, 1992.

Qianli Liao and Tomaso A. Poggio. Bridging the gaps between residual learning, recurrent neural networks and visual cortex. *CoRR*, abs/1604.03640, 2016. URL http://arxiv.org/abs/1604.03640.

Eric V. Linder. Exploring the expansion history of the universe. *Phys. Rev. Lett.*, 90: 091301, Mar 2003. doi: 10.1103/PhysRevLett.90.091301. URL https://link.aps.org/doi/10.1103/PhysRevLett.90.091301.

Seppo Linnainmaa. The representation of the cumulative rounding error of an algorithm as a taylor expansion of the local rounding errors, 1970. URL https://people.idsia.ch//~juergen/linnainmaa1970thesis.pdf.

Seppo Linnainmaa. Taylor expansion of the accumulated rounding error. *BIT Numerical Mathematics*, 16:146–160, 1976.

D. Lowe and B.S. Broomhead. M ultivar ia ble functional interpolation and adaptive n etworks. *Complex Systems*, 1(1):321–355, 1988. URL https://sci2s.ugr.es/keel/pdf/algorithm/articulo/1988-Broomhead-CS.pdf.

Miles Lutter, C. Ritter, and J. Peters. Deep lagrangian networks: Using physics as model prior for deep learning. In *7th International Conference on Learning Representations (ICLR)*, New Orleans, 2019. URL http://tubiblio.ulb.tu-darmstadt.de/122300/.

Charles W. Misner, Kip S. Thorne, and John A. Wheeler. *Gravitation*. W.H. Freeman and Company, USA, 1973. ISBN 0-7167-0334-3.

Hayato Motohashi and Teruaki Suyama. Third order equations of motion and the ostrogradsky instability. *Phys. Rev. D*, 91:085009, Apr 2015. doi: 10.1103/PhysRevD.91.085009. URL https://link.aps.org/doi/10.1103/PhysRevD.91.085009.

S. Perlmutter, S. Gabi, G. Goldhaber, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, R. Pain, C. R. Pennypacker, I. A. Small, R. S. Ellis, R. G. McMahon, B. J. Boyle, P. S. Bunclark, D. Carter, M. J. Irwin, K. Glazebrook, H. J. M. Newberg, A. V. Filippenko, T. Matheson, M. Dopita, and W. J. Couch. Measurements of the Cosmological Parameters $\Omega$ and $\Lambda$ from the First Seven Supernovae at z >= 0.35. *ApJ*, 483(2):565–581, July 1997. doi: 10.1086/304265.

PLANCK-Collaboration. Planck 2018 results - i. overview and the cosmological legacy of planck. *Astronomy and Astrophysics*, 641:A1, 2020. doi: 10.1051/0004-6361/201833880. URL https://doi.org/10.1051/0004-6361/201833880.

Lev S. Pontryagin. *The mathematical theory of optimal processes*. Oldenbourg, 1962.

Adam G. Riess, Alexei V. Filippenko, Peter Challis, Alejandro Clocchiatti, Alan Diercks, Peter M. Garnavich, Ron L. Gilliland, Craig J. Hogan, Saurabh Jha, Robert P. Kirshner, B. Leibundgut, M. M. Phillips, David Reiss, Brian P. Schmidt, Robert A. Schommer, R. Chris Smith, J. Spyromilio, Christopher Stubbs, Nicholas B. Suntzeff, and John Tonry. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal*, 116(3):1009–1038, sep 1998. doi: 10.1086/300499. URL https://doi.org/10.1086/300499.

Horst Rinne. *Taschenbuch der Statistik*. Harri Deutsch, 2008. ISBN 978-3-8171-1827-4.

Jürgen Schmidhuber. Deep learning in neural networks: An overview. *CoRR*, abs/1404.7828, 2014. URL http://arxiv.org/abs/1404.7828.

D. N. Spergel, R. Bean, O. Doré, M. R. Nolta, C. L. Bennett, J. Dunkley, G. Hinshaw, N. Jarosik, E. Komatsu, L. Page, H. V. Peiris, L. Verde, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications for Cosmology. *ApJS*, 170(2):377–408, June 2007. doi: 10.1086/513700.

R. Stompor, M. Abroe, P. Ade, A. Balbi, D. Barbosa, J. Bock, J. Borrill, A. Boscaleri, P. de Bernardis, P. G. Ferreira, S. Hanany, V. Hristov, A. H. Jaffe, A. T. Lee, E. Pascale, B. Rabii, P. L. Richards, G. F. Smoot, C. D. Winant, and J. H. P. Wu. Cosmological Implications of the MAXIMA-1 High-Resolution Cosmic Microwave Background Anisotropy Measurement. *ApJ*, 561(1):L7–L10, November 2001. doi: 10.1086/324438.

N. Suzuki, D. Rubin, C. Lidman, G. Aldering, R. Amanullah, K. Barbary, L. F. Barrientos, J. Botyanszki, M. Brodwin, N. Connolly, K. S. Dawson, A. Dey, M. Doi, M. Donahue, S. Deustua, P. Eisenhardt, E. Ellingson, L. Faccioli, V. Fadeyev, H. K. Fakhouri, A. S. Fruchter, D. G. Gilbank, M. D. Gladders, G. Goldhaber, A. H. Gonzalez, A. Goobar, A. Gude, T. Hattori, H. Hoekstra, E. Hsiao, X. Huang, Y. Ihara, M. J. Jee, D. Johnston, N. Kashikawa, B. Koester, K. Konishi, M. Kowalski, E. V. Linder, L. Lubin, J. Melbourne, J. Meyers, T. Morokuma, F. Munshi, C. Mullis, T. Oda, N. Panagia, S. Perlmutter, M. Postman, T. Pritchard, J. Rhodes, P. Ripoche, P. Rosati, D. J. Schlegel, A. Spadafora, S. A. Stanford, V. Stanishev, D. Stern, M. Strovink, N. Takanashi, K. Tokita, M. Wagner, L. Wang, N. Yasuda, H. K. C. Yee, and The Supernova Cosmology Project. The Hubble Space Telescope Cluster Supernova Survey. V. Improving the Dark-energy Constraints above z > 1 and Building an Early-type-hosted Supernova Sample. *ApJ*, 746(1):85, February 2012. doi: 10.1088/0004-637X/746/1/85.

Matus Telgarsky. benefits of depth in neural networks. In Vitaly Feldman, Alexander Rakhlin, and Ohad Shamir, editors, *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 1517–1539, Columbia University, New York, New York, USA, 23–26 Jun 2016. PMLR.

Ch. Tsitouras. Runge–kutta pairs of order 5(4) satisfying only the first column simplifying assumption. *Computers and Mathematics with Applications*, 62(2):770–775, 2011. ISSN 0898–1221. URL https://www.sciencedirect.com/science/article/pii/S0898122111004706.

Steven Weinberg. *Cosmology*. Oxford Univseristy Press, Oxford, UK, 2008. ISBN 978-0-519-852682-7.

Jochen Weller and Andreas Albrecht. Future supernovae observations as a probe of dark energy. *Physical Review D*, 65, 06 2001. doi: 10.1103/PhysRevD.65.103512.

Jing-Fei Zhang, Li-Yang Gao, Dong-Ze He, and Xin Zhang. Improving cosmological parameter estimation with the future 21 cm observation from ska. *Physics Letters B*, 799:135064, 2019a. ISSN 0370-2693. doi: https://doi.org/10.1016/j.physletb.2019.135064. URL https://www.sciencedirect.com/science/article/pii/S0370269319307865.

Ning Zhang, Shui-Long Shen, Annan Zhou, and Ye-Shuang Xu. Investigation on performance of neural networks using quadratic relative error cost function. *IEEE Access*, 7:106642–106652, 2019b. doi: 10.1109/ACCESS.2019.2930520.

Shay Zucker, Tal Alexander, Stefan Gillessen, Frank Eisenhauer, and Reinhard Genzel. Probing post-newtonian physics near the galactic black hole with stellar redshift measurements. 639(1):L21–L24, feb 2006. doi: 10.1086/501436. URL https://doi.org/10.1086/501436.

Erklärung:

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Weiterhin versichere ich, dass die Arbeit bei keiner anderen Fakultät zur Erlangung eines akademischen Grades eingereicht wurde.

Heidelberg, den 1. November, 2021        . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .