Department of Physics and Astronomy
University of Heidelberg

Bachelor Thesis in Physics
submitted by

Maximilian Maria Richter

born in Tübingen (Germany)

2020

Machine Learning Quintessence Dark Energy Potentials via Symbolic Regression and Bayesian Inference

This Bachelor Thesis has been carried out by Maximilian Maria Richter at the
Institute for Theoretical Astrophysics in Heidelberg
under the supervision of
Prof. Dr. Björn-Malte Schäfer

ABSTRACT

Quintessence is a hypothesized scalar field forming a popular model for dark energy, the non-luminous energy density in the Universe responsible for the accelerated expansion of space. This field can change in time and henceforth solve difficulties arising in the ΛCDM standard model. The dynamics of the field are governed by its self-interaction potential. To infer the form of this potential by observational data, we construct a suitable potential using machine learning and derive the observable quantities then numerically. Symbolic regression is a form of supervised learning, with the aim of finding the most likely symbolic representations of mathematical models for a given data set.

We demonstrate how to build a symbolic regression machine learning pipeline, searching for self-interaction potentials of the quintessence scalar field Lagrangian. To construct this pipeline, we first implement a symbolic regression algorithm searching for classical Lagrangian potentials and discuss its ability to solve the problem for artificial data. For upgrading this algorithm to search for quintessence potentials, we assembly an automatic process of evaluating the individual potentials and comparing the resulting models to data of supernovae type Ia with Bayesian statistics. We discuss the difficulties of automatically assigning Bayesian measures due to the high complexity of the fitness process as well as limited computing strength. We conclude, that our symbolic regression pipeline is capable of finding suitable quintessence potentials under consideration of sufficiently diverse function populations.

ZUSAMMENFASSUNG

Quintessenz ist ein hypothetisches skalares Feld und ein beliebtes Modell für dunkle Energie, die unsichtbare Energiedichte im Universum, verantwortlich für die beschleunigte Expansion des Weltraums. Dieses Feld ist zeitveränderlich und kann somit viele der Schwierigkeiten lösen, die im ΛCDM Standardmodell entstehen. Die Dynamik des Feldes ist bestimmt durch sein Selbstwechselwirkungspotential. Um die Form dieses Potentials von Beobachtungsdaten abzuleiten, konstruieren wir mithilfe von maschinellem Lernen ein geeignetes Potential und bestimmen die beobachtbaren Größen numerisch. Symbolic Regression ist eine Form von betreutem Lernen mit dem Ziel die wahrscheinlichsten symbolischen Darstellungen mathematischer Modelle eines gegebenen Datensatzen zu finden.

Wir demonstrieren wie eine Symbolic Regression Machine Learning Pipeline gebaut wird, welche nach selbstwechselwirkungs-Potentialen der skalaren Quintessence Lagrangedichte sucht. Um diese Pipeline zu konstruieren, implementieren wir zuerst einen Symbolic Regression Algorithmus, welcher nach klassischen Lagrange'schen Potentialen sucht, und diskutieren seine Fähigkeit das Problem für ein künstlichen Datensatz zu lösen. Der Algorithmus wird erweitert um nach Quintessenzpotentialen zu suchen. Wir erstellen daher einen automatischen Prozess, welcher die individuellen Potentiale auswertet und die sich ergebenden Modelle mit den Daten von Supernovae Type Ia mittels Bayes'scher Statistik vergleicht. Wir diskutieren die Schwierigkeiten der automatischen Zuordnung Bayes'scher Maße durch die hohe Komplexität des Fitness-Prozesses sowie beschränkter Rechenstärke. Wir schlussfolgern, dass unsere Symbolic Regression Pipeline, unter Anbetracht hinreichend diverser Funktionspopulationen, in der Lage dazu ist geeignete Quintessenzpotentiale zu finden.

TABLE OF CONTENTS

Chapter One

THEORETICAL FOUNDATION

1.1 Introduction

Understanding the world we live in and the motion of the celestial bodies has been in the interest of mankind, one way or another, since the beginning of the first civilizations. The study of the Universe as a whole is named cosmology (from the Greek *κόσμολογία*, kosmos ”world” and -logia ”study of”). Even though in the ancient world some remarkable work was done on the systematic study, mathematical modeling and prediction of the "heavenly bodies", late advancements in technology, as well as in physics and mathematics, have facilitated the constuction of revolutionary new models of cosmologies and ways to test them with unprecedented precision.

Today’s standard model of physical cosmology is derived from Albert Einstein’s theory of general relativity. Under the use of universal symmetries, Alexander Friedmann was able to reformulate Einstein’s field equations into differential equations describing the dynamics of the Universe. Unfortunately, many physicists at that time, including Einstein, were still convinced of the static and stationary nature of our Universe. Nevertheless, the eventuality of a dynamic cosmos lead Edwin Hubble to the breakthrough discovery of the proportionality of the velocity and distance of near galaxies, and hence to a renaissance of physical cosmology. After accumulating observational data of high distance supernovae Type Ia (SNe Ia) by the year 1998, Riess and Perlmutter *et. al.* reported that the Universe is accelerating [1, 2]. The source of this accelerated expansion is called "Dark Energy". This form of energy can be clearly distinguished from other ordinary and dark matter in the Universe, but yet its precise nature could not be revealed to full satisfaction. The main lack of confidence may not

essentially be found in the prediction of observational data, but in the emergence of further obstacles, discussed in more detail later.

Already over a decade earlier, Ratra, Peebles and Wetterich suggested a self interacting scalar field causing the inflationary expansion in very early times of[3, 4]. But only with the discovery of the Higgs boson in 2012, the existence of scalar fields in nature could be proven. This gave rise to the natural conclusion of a self interacting scalar field, called quintessence, and its energy density governing the accelerated expansion of our Universe today.

In order to quantitatively analyze the massive amounts of data present in astronomy without human bias, machine learning has become one of the most promising technologies in physics and other sciences in general. As young as the modern improvements in machine based learning may be, the underlying statistics have come a long way and were revolutionized in the 18th century by the British mathematician and philosopher Thomas Bayes and his well-known theorem. Further development of probability theory and the growth in computing power has recently entered a new era of advanced statistical optimization, commonly referred to as ”artificial intelligence”.

Besides the popular neural networks employed for deep learning, genetic programming offers another very promising framework for machine learning algorithms. Symbolic regression is an interesting example of genetic programming, designed to search the space of symbolic and, in particular, human readable mathematical expressions. Schmidt and Lipson even demonstrated that symbolic regression is capable of finding free-form natural laws from experimental data even if the systems observables exhibit chaotic behaviour [5].

The aim of this thesis is to demonstrate and discuss a self-developed symbolic regression machine learning pipeline, capable of finding suitable dark energy models, i.e. symbolic interaction potentials of the quintessence action by means of Bayesian statistics and constraints of the SNe Ia data.

1.2 Cosmology

Physical cosmology rests on the two very fundamental assumptions, namely homogeneity and isotropy of our Universe. The first one based on the cosmological principle and reflects the Copernican revolution, that our position in the Universe is by no means preferred over any other. The second one states the independence of the direction of observation. The discovery of the 2.7K cosmic microwave background (CMB) radiation [6] and the Baryo-acoustic oscillations (BAO) [7] brought evidence supporting those assumptions.
In the following we will adapt natural units, i.e. $c = \hbar = 1$, unless stated otherwise.

1.2.1 Spacetime Geometry

In Einstein's general theory of relativity, space and time are solely degrees of freedom of a four dimensional pseudo-Riemannian manifold [8]. The position of an object in this space-time is described by the contravariant four vector

$$x^\mu = (t, x^i)^T \qquad x^i = (x^1, x^2, x^3)^T \tag{1.1}$$

The 0th component is the time component of the four vector, the other three are the spatial coordinates. We employ Greek indices running from 0 to 3 as notation for the whole four-vector in a given coordinate basis and Latin indices (e.g. x^i) for the spatial dimensions running from 1 to 3.
The separation of two points in a curved spacetime is quantified by the metric tensor $g_{\mu\nu}$. The infinitesimal distance is called the line element $\mathrm{d}s$, which, on the other hand, defines the metric tensor

$$\mathrm{d}s^2 = \mathrm{d}\mathbf{x} \cdot \mathrm{d}\mathbf{x} = \mathrm{d}x_\mu \mathrm{d}x^\mu = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu \tag{1.2}$$

where we used the Einstein convention, summing over identical lowered and raised indices. The dynamics of the metric tensor is governed by Einstein's field equations

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{R}{2} g_{\mu\nu} = 8\pi G T_{\mu\nu} + \Lambda g_{\mu\nu} \tag{1.3}$$

where $G_{\mu\nu}$ is called the Einstein tensor and G is the gravitational constant. The left-hand side of Eq. (1.3) describes the curvature of space-time. The right-hand side specifies the energy content within the energy-momentum tensor. Curvature is described by means of the so called Christoffel symbols Γ and their derivatives in the Ricci curvature tensor

$$R_{\mu\nu} = \partial_\rho \Gamma^\rho_{\nu\mu} - \partial_\nu \Gamma^\rho_{\rho\mu} + \Gamma^\rho_{\rho\lambda}\Gamma^\lambda_{\nu\mu} - \Gamma^\rho_{\nu\lambda}\Gamma^\lambda_{\rho\mu} \tag{1.4}$$

If we chose the Levi-Civita connection (metric compatible and torsion-free, i.e. $\nabla g = 0$ and $\Gamma^\alpha_{\mu\nu} = \Gamma^\alpha_{\nu\mu}$), the Christoffel symbols can be derived easily from the metric as

$$\Gamma^\alpha_{\mu\nu} = \frac{1}{2} g^{\alpha\beta}\left(\frac{\partial g_{\nu\alpha}}{\partial x^\mu} + \frac{\partial g_{\mu\beta}}{\partial x^\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\beta}\right) \tag{1.5}$$

The Ricci scalar R is simply the contraction or the trace of the Ricci curvature tensor with the metric tensor

$$R = g^{\mu\nu} R_{\mu\nu} = R^\mu_\mu \,. \tag{1.6}$$

Λ is called the cosmological constant and a possible candidate for dark energy. It describes a degree of freedom in the Einstein field equations and is discussed in more detail later.

Assuming homogeneity and isotropy, we can construct the most general metric fulfilling these symmetries, the Friedmann-Lamaître-Robertson-Walker (FLRW) metric, with the corresponding line element

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\left(\frac{\mathrm{d}r^2}{1 - Kr^2} + r^2 \mathrm{d}\vartheta^2 + r^2 \sin^2(\vartheta)\mathrm{d}\varphi^2\right) \tag{1.7}$$

where a is the cosmic scale factor, characterizing the spatial distance of all points in space, and K is a constant, parameterizing the curvature of space.

Evaluating the Christoffel symbols (1.5) in the FLRW-metric (1.7) (See [9]), we can rewrite Einstein's Eqs. (1.3) into two differential equations of the scale factor a

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{K}{a^2} + \frac{\Lambda}{3} \tag{1.8}$$

$$3H^2 + 2\dot{H} = -8\pi G P - \frac{K}{a^2} + \frac{\Lambda}{3} \tag{1.9}$$

These central results are the so-called first and second Friedmann equations. The first Friedmann equation quantifies the expansion rate of the cosmos and is defined as

$$H \equiv \frac{\dot{a}}{a} \tag{1.10}$$

where H is called the Hubble parameter.

The second Friedmann equation is also called the acceleration equation and will be of importance later, as we will study the accelerated expansion of the cosmos.

We further consider the energy-momentum tensor of an ideal fluid with an energy density ρ and pressure P as our cosmic fluid

$$T^{\mu}_{\nu} = (\rho + P)u^{\mu}u_{\nu} + P\delta^{\mu}_{\nu}. \tag{1.11}$$

A third equation can be derived from energy-momentum conservation

$$\nabla_{\mu}T^{\mu\nu} = 0 \quad \Rightarrow \quad \frac{d\rho}{dt} + 3\frac{\dot{a}}{a}(\rho + P) = 0 \tag{1.12}$$

From statistical physics we know that the equation of state for an ideal fluid reads

$$w \equiv \frac{P}{\rho} \tag{1.13}$$

We can solve the differential equation (1.8, 1.9) by separation of variables, giving us an implicit time dependence of ρ in terms of $a(t)$. The power of the time-scaling is determined by the equation of state parameter w as

$$\rho \propto a^{-3(1+w)} \tag{1.14}$$

Since our Universe is not filled with one single type of matter, we have to consider different energy species scaling separately in time, depending on their w. We can expand the energy density ρ in the Friedmann equations as a sum over all possible powers of a and their contribution to the full energy density

$$\rho = \sum_{n=-\infty}^{\infty} \rho_n a^n \tag{1.15}$$

For cold matter, i.e. dark and baryonic matter, we know from statistical physics that the pressure of those gases is negligible, i.e. $P = 0$. Hence, the energy density of cold matter scales with

$$\rho \propto a^{-3} \tag{1.16}$$

For hot, relativistic gases, such as electromagnetic radiation or neutrinos, the relation between pressure and energy density is given by $P = \rho/3$, thus its energy density scales with

$$\rho \propto a^{-4} \tag{1.17}$$

We can define a critical energy density, at which the curvature of the spatial hypersurfaces K vanishes. This is the case if the entire energy density of the cosmos matches the critical value

$$\rho_{\text{crit}}(t) = \frac{3H^2(t)}{8\pi G} \tag{1.18}$$

Recent work in astronomy has accumulated evidence, that the curvature of the Universe is negligible today [9]. We therefore assume $K = 0$ in the following and thus, the entire energy density ρ of the Universe to be equal to the critical value ρ_{crit}. Having this natural scale at hand, we can now define the simple dimensionless density parameters, commonly used in cosmological calculations, each having their respective index

$$\Omega_i(t) := \frac{\rho_i(t)}{\rho_{\text{crit}}(t)} \qquad \Omega_i = \frac{8\pi G}{3H^2}\rho_i \tag{1.19}$$

Additionally, we recognise the curvature K and the cosmological constant Λ as energy densities too

$$\Omega_K := \frac{-K}{(aH)^2} \qquad \Omega_\Lambda := \frac{\Lambda}{3H^2} \tag{1.20}$$

The sum over all density parameters has to equal unity by definition, i.e.

$$\sum_i \Omega_i = 1 \tag{1.21}$$

We further denote the values of the parameters today with a null-index

$$\rho_{\text{crit}}(t_0) = \frac{3H^2(t_0)}{8\pi G} := \rho^0_{\text{crit}}, \qquad \Omega_i(t_0) := \Omega^0_i \tag{1.22}$$

With measurements of the parameters [10], we can bring the Friedmann equation into a much simpler form, the Hubble function

$$H(a) = H_0 \sqrt{\sum_i \frac{\Omega_i^0}{a^{3(1+w_i)}}} \tag{1.23}$$

$$= H_0 E(a) \tag{1.24}$$

where we defined the expansion function

$$E(a) := \frac{H(a)}{H_0} \tag{1.25}$$

with $H_0 \equiv H(t_0)$ being today's rate of expansion.

1.2.2 Hubble's Law

The frequencies of waves emitted by a moving source are shifted depending on the velocity of the source relative to an observer. This effect is called Doppler shift and is widely used in astronomy to measure the relative speed of distant stars and galaxies by comparing the spectral lines of astronomical light sources with well studied atomic and molecular transition spectra. Edwin Hubble discovered that most of the very distant galaxies appear to have a red-shifted spectrum in proportion to their estimated distance. He concluded that, since our position in the Universe is by no means preferred over any other, that the Universe is expanding and all objects in space drift apart from each other.

Mathematically speaking, the expansion of the Universe is expressed by the time-dependence of the scale factor a. The relative amount of the wavelength shift λ compared to the expected wavelength λ_0 and its relation to the scale factor a is defined as the redshift z

$$z \equiv \frac{\lambda - \lambda_0}{\lambda_0} = \frac{a_0}{a} - 1 \quad \Rightarrow \quad a = \frac{1}{1+z} \tag{1.26}$$

where, by definition, the present value of the scale factor is $a_0 = 1$, and hence $z = 0$ today. As long as the recessional velocity v of an object is much smaller than the speed of light c, we

approximately obtain a linear proportionality between redshift and velocity,

$$\lambda_0 \approx \left(1 + \frac{\upsilon}{c}\right)\lambda \quad \Leftrightarrow \quad z \approx \upsilon/c \tag{1.27}$$

The physical distance $\mathbf{r}$ and its derivative with respect to time in an expanding Universe is given by

$$\mathbf{r} = a(t)\mathbf{x} \quad \Rightarrow \quad \dot{\mathbf{r}} = H\mathbf{r} + a\dot{\mathbf{x}} \tag{1.28}$$

where a is the scale factor and $\mathbf{x}$ denotes the comoving distance. The speed of an object along the direction between an observer and the object is given by

$$\upsilon \equiv \dot{\mathbf{r}} \cdot \frac{\mathbf{r}}{r} = Hr + \boldsymbol{\upsilon}_p \cdot \frac{\mathbf{r}}{r} \tag{1.29}$$

where $r = |\mathbf{r}|$ and $\boldsymbol{\upsilon}_p$ is the so-called peculiar velocity.

Most galaxies do not exceed the speed of 10^6 m/s in the local universe [9], so the peculiar velocity υ_p can be neglected. This gives us Hubble's Law

$$\upsilon \approx H_0 r \tag{1.30}$$

where we have replaced H with its current value H_0, justifiable for a small redshift region ($z \ll 1$).

It is common to set the value of the Hubble constant to

$$H_0 = 100h\frac{\text{km}}{\text{s Mpc}} = 2.1332h \times 10^{-42}\text{GeV} \tag{1.31}$$

where we have used

$$1 \text{ Mpc} = 3.08568 \times 10^{24}\text{cm} = 3.26156 \times 10^6 \text{ light years} \tag{1.32}$$

Here, h is the uncertainty on the value for the Hubble constant. Measurements [9] have constrained this uncertainty to

$$h = 0.72 \pm 0.08 \tag{1.33}$$

For reasons of simplicity, we choose h the be equal to 0.7 in the following.

1.2.3 Distance measures

In order to quantitatively test cosmological models, we need to be able to measure distances of light sources far away in space and time. Light traveling towards us is always moving on a geodesic of the space-time manifold in the direction χ, i.e. it satisfies

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + a^2(t)\mathrm{d}\chi^2 = 0 \quad \Leftrightarrow \quad \mathrm{d}\chi = \frac{c}{a(t)}\mathrm{d}t \tag{1.34}$$

where we recovered c for clarity.

Light emitted at time $t = t_1$ with $\chi = \chi_1$ and a redshift of z will reach an observer at $t = t_0$ with $\chi = 0$ and respective $z = 0$. The comoving distance between source and observer is then given by the integration of Eq. (1.34) back in time

$$d_c \equiv \int_0^{\chi_1} \mathrm{d}\chi = -\int_{t_0}^{t_1} \frac{c}{a(t)}\mathrm{d}t \tag{1.35}$$

Since we measure the time "distances" of the observed light source by its redshift, it is convenient to change our integration variable with Eq. (1.26)

$$\mathrm{d}t = -\frac{\mathrm{d}z}{H(z)(1+z)} \tag{1.36}$$

Expressing our integral (1.35) through equation (1.36) yields

$$d_c = \frac{c}{a_0 H_0}\int_0^z \frac{\mathrm{d}z'}{E(z')} \tag{1.37}$$

which can be shown to satisfy Hubble's law if expanded around $z = 0$ (See Ref. [9]).

The flux $\mathcal{F} = L_0/S$ of a point source with observed luminosity L_0 in a flat Universe ($K = 0$) is expected to scale proportionally to the inverse of the surface of a sphere $S = 4\pi(a_0 d_c)^2$:

$$\mathcal{F} = \frac{L_0}{4\pi(a_0 d_c)^2} \tag{1.38}$$

Due to the large distances in cosmology, relativistic effects gain weight when estimating the distance of objects from their apparent luminosity. This means the luminosity L_0 observed at

$\chi = 0$ and $z = 0$ is different from the absolute luminosity L_S of the source emitted at comoving distance χ with redshift z. We therefore define the commonly used luminosity distance

$$d_L^2 \equiv \frac{L_S}{4\pi\mathcal{F}} \tag{1.39}$$

Combining Eqs. (1.38) and (1.39) yields

$$d_L^2 = (a_0 d_c)^2 \frac{L_S}{L_0} \tag{1.40}$$

The luminosity of a light source is defined by the energy emitted in a given time interval $L = \Delta E/\Delta t$. The energy of photons is inversely proportional to their wavelength λ, hence $\Delta E_0/\Delta E_S = \lambda_S/\lambda_0$ with ΔE_0 the energy of the observed photon at Δt_0 and ΔE_S the energy of the photon at emission at Δt_S. Together with Eq. (1.26) this gives $\lambda_S/\lambda_0 = (1 + z)$. The constancy $c = \lambda/\Delta t$ implies that $\Delta t_0/\Delta t_S = \lambda_0/\lambda_S = (1 + z)$. Comparing the absolute and observed luminosities L_S/L_0, we obtain

$$\frac{L_s}{L_0} = \frac{\Delta E_S}{\Delta E_0}\frac{\Delta t_0}{\Delta t_S} = (1 + z)^2 \tag{1.41}$$

This reduces Eq. (1.39) to

$$d_L = a_0 d_c (1 + z) = \frac{c(1 + z)}{H_0}\int_0^z \frac{\mathrm{d}z'}{E(z')} \tag{1.42}$$

Based on the astronomical magnitude system, we define the distance modulus μ as the difference between the apparent magnitude m and absolute magnitude M. The modulus is related to our luminosity distance measured in parsec by

$$\mu = m - M = \log_{10}(d_L) - 5 \tag{1.43}$$

1.2.4 Supernovae Type Ia

The most frequent form of supernovae throughout the Universe are those of type Ia. The common picture is that the progenitor system of such a supernova type Ia consists of at least one massive ($\approx$ 1 solar mass $M_\odot$) carbon-oxygen white dwarf [11]. White dwarfs are stabilized by

the Fermi pressure of a degenerate electronic gas. This pressure can hold gravity only up the Chandrasekhar limit of about $1.4M_\odot$ [12]. If this limit gets exceeded (by mass transfer from a red giant or by two binary white dwarfs merging) a thermonuclear runaway is triggered, which eventually causes the white dwarf to explode and is responsible for the extraordinary brightness of type Ia supernovae.

To infer the luminosity distance of supernovae it is necessary to assign absolute luminosities. Fortunately, supernovae of type Ia are "standard candles", i.e. there exists an empirical relation between the width of the observable light curve and the peak luminosity. This allows calibration to an absolute luminosity. Measuring the redshift of their spectra, it is possible to reconstruct the luminosity distance as a function of redshift. For a matter dominated Universe ($\Omega_M^0 = 1, \Omega_K^0 = 0, \Omega_\Lambda = 0$, also called Einstein-de Sitter limit [13]), the expansion function (1.25) can be written as

$$E(z) = \sqrt{\Omega_M^0 (1+z)^3} \tag{1.44}$$

The luminosity distance (1.42) for this expansion function can be integrated analytically

$$\Rightarrow d_L(z) = \frac{c(1+z)}{H_0} \int_0^z \frac{dz'}{\sqrt{\Omega_M^0 (1+z')^3}} = \frac{c(1+z)}{H_0} \left[2 - \frac{2}{\sqrt{\Omega_M^0}\sqrt{1+z}} \right] \tag{1.45}$$

Riess and Perlmutter *et al.* gathered data of over 500 type Ia supernovae, provided by the Supernova cosmology project [14]. Surprisingly they do not match a Universe filled with only baryonic and dark matter (See Fig. 1.1). Together with measurements of the CMB and the BAO, it is now a well-established fact that the change of the Hubble parameter H at low redshift is caused by acceleration of the expansion rate of our cosmos [9]. This acceleration is undoubtedly unsuspected and ordinary matter is obviously not capable of explaining it. This implies, that our Universe is filled with an, as of yet, unknown, non-luminous energy density, causing the late time accelerated expansion of the cosmos, hereinafter referred to as "dark energy"

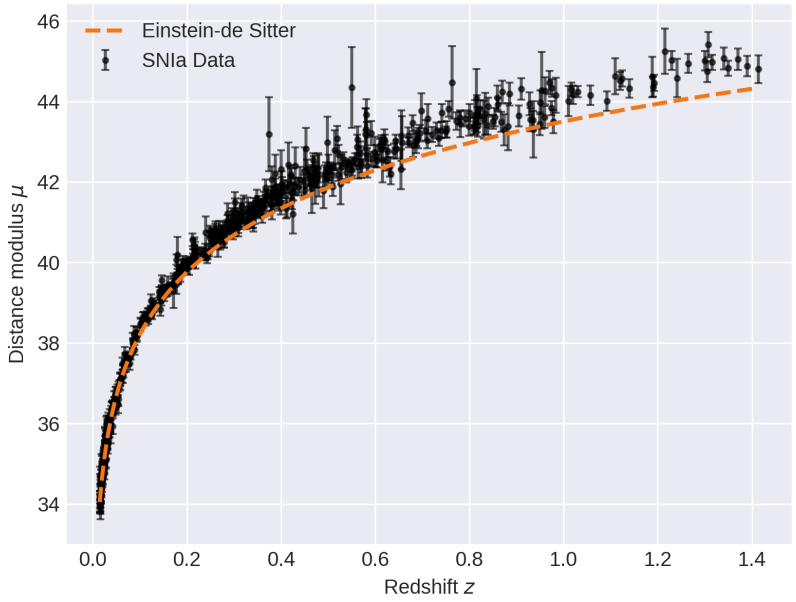

Figure 1.1 Distance modulus against redshift for SN Ia data and their corresponding 1σ-error. On top, the Einstein-de Sitter limit solution to the Friedmann equations for a Universe filled with only dark and baryonic matter $\Omega_M = 1, \Omega_K = 0$ and $\Omega_\Lambda = 0$.

1.3 Dark Energy

The data of the supernovae type Ia suggests, that the speed of expansion is growing, i.e. there has to be a force in the Universe, counteracting the attractive gravitation and hence causing this accelerated expansion on large scales. Einstein's general theory of relativity provides a mechanism for which gravity can become repulsive, namely if the pressure of a fluid is sufficiently negative

$$P < -\frac{\rho}{3} \tag{1.46}$$

This condition is equivalent to an equation of state parameter of

$$w < -\frac{1}{3} \tag{1.47}$$

Since is not yet clear if the equation of state parameter w_{DE} of dark energy is constant in time or dynamical, we have to, again, consider the adiabatic equation

$$\dot{\rho}_{DE} + 3H(\rho_{DE} + P_{DE}) = 0 \tag{1.48}$$

Now assuming a time dependence of w_{DE}, integrating equation (1.48) with respect to the redshift z takes the following form

$$\rho_{DE} = \rho_{DE}^{0} \exp\left[\int_0^z \frac{3\,(1 + w_{DE}(z'))}{1 + z'} \mathrm{d}z'\right] \tag{1.49}$$

where we used the transformation Eq. (1.36) again.

1.3.1 ΛCDM

The most simple candidate for dark energy is Einstein's cosmological constant Λ. It would be sufficient for most of the astronomical cases if Λ is equal to zero, but, as Lovelock's theorem [8] proofs, any tensor T, which depends on the metric g and only its first and second derivatives, must be a linear combination of the Einstein and metric tensor. Thus, the Einstein tensor must be of the form

$$G = \kappa T - \Lambda g \tag{1.50}$$

with $\kappa = 8\pi G$ in the Newtonian limit and the energy-momentum tensor T.

Under the assumption that Λ's energy density remains constant over the expansion of the background, the corresponding equation of state parameter for Λ simply reads

$$w = -1 \tag{1.51}$$

so the cosmological constant fulfills the condition (1.47), as its w results in a sufficiently negative pressure and hence an accelerated expansion of space at late times.

To match experimental observation, like the SN Ia data [14], the CMB or the BAO [10], the cosmological constant is estimated to have a value of

$$\Lambda \approx H_0^2 = (2.1332h \times 10^{-42}\mathrm{GeV})^2 \tag{1.52}$$

Its energy density parameter today is estimated to be approximately $\Omega_\Lambda^0 \approx 0.7$, i.e. 70% of the energy content of the Universe. The missing 30% are divided into cold baryonic and dark matter, denoted as Ω_M. "Cold" means that the particles momenta are negligible, compared to

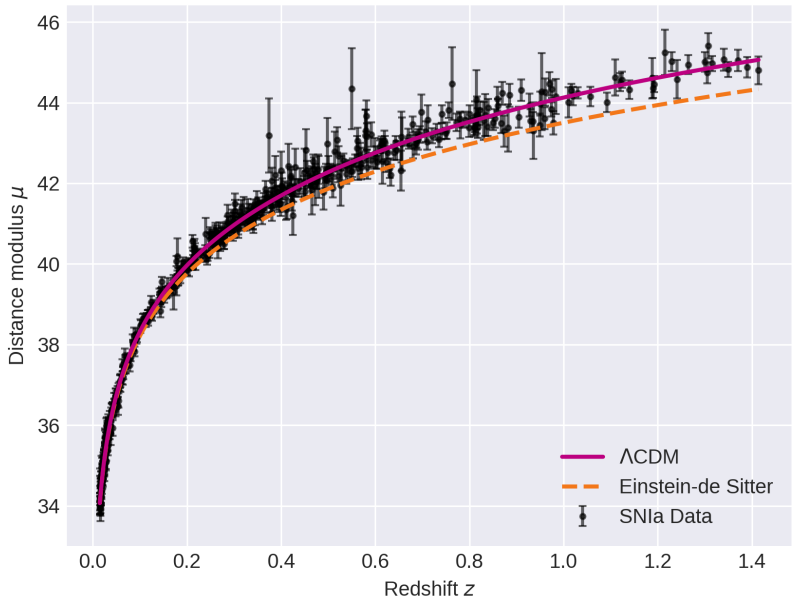

Figure 1.2 SN Ia data with ΛCDM-Universe in blue and a Universe filled with only cold dark matter in dashed orange

their rest mass, so the expectation value for the statistical pressure of such gases is zero. This yields $w = 1/3$ for both, ordinary and dark matter.

Assuming further the curvature and the electromagnetic radiation radiation density of the Universe to vanish ($\Omega_K^0 \approx 0$, $\Omega_\gamma^0 \approx 10^{-5}$, as shown observationally [10]), the Hubble function (1.24) can be written as

$$H^2(z) = H_0^2 \left[\Omega_M^0 (1+z)^3 + \Omega_\Lambda^0\right] \tag{1.53}$$

Integrating Eq. (1.42) for this $H(z)$ numerically yields the luminosity distance and redshift relationship for this type of Universe

$$d_L(z) = \frac{c(1+z)}{H_0} \int_0^z \frac{\mathrm{d}z'}{\sqrt{\Omega_M^0 (1+z')^3 + \Omega_\Lambda^0}} \tag{1.54}$$

This approach is called the "Lambda Cold Dark Matter" (ΛCDM) model and is referred to as the standard model of physical cosmology.

Despite its success, Λ itself is rather questionable, since it requires extended fine tuning to

explain the accelerated expansion of our Universe today and only arguments based on the anthropic principle support its present value [15]. This issue is called the fine tuning problem of the cosmological constant. The true nature of Λ remains unclear. Attempts to explain the energy density of the cosmological constant in terms of quantum field vacuum fluctuations results in the awkwardly big discrepancy of $\sim 10^{120}$ (See Ref. [9]).

Precise fine tuning is also necessary to explain, that Ω_M^0 and Ω_Λ^0 seem to be of the same order of magnitude for no apparent reason. Another way to formulate this is the question of why today? This is the so-called coincidence problem.

The inability to explain its nature and its present value disfavors the cosmological constant Λ as an explanation for dark energy. If Λ needs to vanish, we should find an alternative model.

1.3.2 Quintessence

A canonical scalar field ϕ with a self interacting potential $V(\phi)$ responsible for the late-time cosmic acceleration is called "Quintessence". Many scalar fields are present in particle physics, as well as in types of string theory and supergravity. The action for such a quintessence field ϕ reads

$$S[\phi] = \int \mathrm{d}^4x \sqrt{-g}\left[\frac{1}{2\kappa^2}R + \mathcal{L}_\phi\right] + S_M \tag{1.55}$$

where $\kappa^2 = 8\pi G$, R is the Ricci scalar, S_M is the matter action and g denotes the determinant of the metric $g \equiv \det g_{\mu\nu}$ [9]. The Lagrange density of the quintessence field is given by

$$\mathcal{L}_\phi = \frac{1}{2}g^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi - V(\phi) \tag{1.56}$$

with a kinetic term on the left hand side and the potential term on the right hand side. Its energy-momentum tensor can be derived from the variational principle and is given by

$$T^{(\phi)}_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta(\sqrt{-g}\mathcal{L}_\phi)}{\delta g^{\mu\nu}} = \nabla_\mu\phi\nabla_\nu\phi - g_{\mu\nu}\mathcal{L}_\phi = (\rho_\phi + P_\phi)u_\mu u_\nu - P_\phi g_{\mu\nu} \tag{1.57}$$

Since $\vec{\nabla}\phi = 0$, due to homogeneity, the energy-momentum tensor (1.57) can be evaluated in comoving coordinates. We obtain the following simple expression for the energy-density ρ_ϕ

and pressure P_ϕ of the field

$$\rho_\phi = -T_0^{0(\phi)} = \frac{1}{2}\dot{\phi}^2 + V(\phi) \tag{1.58}$$

$$P_\phi = \frac{1}{3}T_i^{i(\phi)} = \frac{1}{2}\dot{\phi}^2 - V(\phi) \tag{1.59}$$

Henceforth, its equation of state parameter is given by

$$w_\phi \equiv \frac{P_\phi}{\rho_\phi} = \frac{\frac{1}{2}\dot{\phi}^2 - V(\phi)}{\frac{1}{2}\dot{\phi}^2 + V(\phi)} \tag{1.60}$$

The quintessence equation of state parameter is able to become negative if the kinetic energy of the field $\dot{\phi}^2/2$ is negligible compared to its potential value $V(\phi)$.

$$\frac{1}{2}\dot{\phi}^2 \ll V(\phi) \tag{1.61}$$

This also means that our quintessence equation of state parameter w_ϕ eventually changes over time to get sufficiently negative, triggering the late cosmic acceleration we observe today. This dynamic property is beneficial for a reliable model of dark energy.

The Hubble function and its time derivative for a Universe with quintessence and cold dark matter is given by

$$H^2 = \frac{8\pi G}{3}\left(\frac{1}{2}\dot{\phi}^2 + V(\phi) + \rho_M\right) \tag{1.62}$$

$$\dot{H} = -4\pi G(\dot{\phi}^2 + \rho_M + P_M) \tag{1.63}$$

Further, in the FLRW-geometry (1.7), Eq. (1.12) for energy-conservation can be brought to the form of a Klein-Gordon equation

$$\ddot{\phi} + 3H\dot{\phi} + \frac{\mathrm{d}V(\phi)}{\mathrm{d}\phi} = 0 \tag{1.64}$$

This is the equation of motion of our scalar field. The fate of the Universe is governed by the form of the quintessence interaction potential $V(\phi)$. As the field "rolls down" the potential slowly, it will continue to gain kinetic energy and thus entering a region in which its energy

density starts to dominate the expansion of the cosmos. To achieve this desired accelerated expansion only at late times, it is convenient to define the so-called slow roll parameters

$$\epsilon_S \equiv \frac{1}{2\kappa^2}\left(\frac{V_{,\phi}}{V}\right)^2 \qquad \text{and} \qquad \eta_S \equiv \frac{V_{,\phi\phi}}{\kappa^2 V} \tag{1.65}$$

where $V_{,\phi} \equiv dV/d\phi$. Both, ϵ_S and η_S, must be smaller than one for the potential to be shallow enough to realize the acceleration at late times.

There are several classes of possible potentials possessing the desired properties, roughly divided into "freezing" and "thawing" models [9].

- **Freezing models**: Potentials for which w_ϕ starts non-negative and "freezes" at late times to -1 due to gain in kinetic energy. Examples would be ($M^2 = M_p^2/16\pi$)
  - $V(\phi) = M^{4+n}\phi^{-n} \quad (n > 0)$
  - $V(\phi) = M^{4+n}\phi^{-n}\exp(\alpha\phi^2/m_{pl}^2)$

  Such potentials appear for example in the fermion condensate model as a dynamical supersymmetry breaking or in the framework of supergravity.

- **Thawing models**: Potentials for which w_ϕ is initially at -1 and starts to become less, so the equation of state starts "thawing". Corresponding potentials are
  - $V(\phi) = V_0 + M^{4-n}\phi^n \quad (n > 0)$
  - $V(\phi) = M^4\cos^2(\phi/f)$

  The first potential is similiar to the one of chaotic inflation, the second one appears as a potential of the Pseudo-Nambu-Goldstone Boson.

The "classical" approach of constructing a suitable potential for the quintessence field is to parametrize w_ϕ analytically and to integrate Eqs. (1.8) and (1.63) (See Ref. [9]). This inverse problem forces to assume a form of w_ϕ, so the reconstructed potential is therefore only effective. Our approach is to directly guess the potential by means of machine learning and calculate the equation of state parameter numerically out of the free-form laws.

1.4 Evolutionary Programming

Developing automated problem solvers is one of the central issues of mathematics and computer science. As the title already suggests, evolutionary programming facilitates the natural selection process of evolution to maximize a certain fitness. This idea is not surprising, since the success of evolution in Nature is evident. The real challenge though is to formulate her principles executable for a computer. Alan Turing already suggested evolving computer programs in the 1950s [16] but the first working evolutionary algorithms have been published in the 60s by Bremermann [**?**].

Darwin pointed out in his theory of evolution [17], that it is natural selection that plays a central role in explaining the rich biological diversity and the adaptation of all life to their individual habitat. Given a set of individuals, called population, with a certain growth rate, it is necessary to limit this otherwise exponential growth by selection. Those individuals which compete for a given resource best, i.e. which fit to the environmental conditions best, will be favored by natural selection. This is also known as "survival of the fittest".

1.4.1 Genetic Algorithms

The microscopic view of natural evolution is offered by the discipline of molecular genetics. Its fundamental observation is that an individuals genotype, i.e. its internal structure, encodes its phenotype, i.e. its external "visible" structure. Genes are functional units of inheritance encoding phenotypic characteristics. These genes can be altered by spontaneous mutation or by sexual reproduction. Through these variations, new combination of traits occur and get evaluated. The best ones survive and reproduce and so the evolution progresses.

Every genetic algorithm therefore needs the following attributes:

- **Representation**: In order for the algorithm to evaluate the phenotype of the individuals, their genes get a representation, e.g. as bits or syntax-trees (See Fig. 1.4).

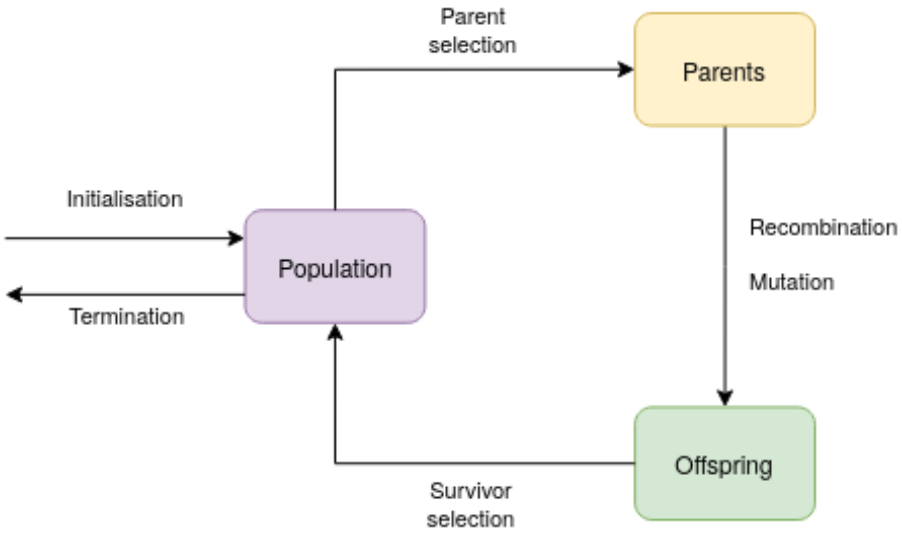

Figure 1.3 General scheme of an evolutionary algorithm as a flowchart. Adapted from [**?**]

- **Fitness function**: Evaluating the genes of individuals and assigns a fitness to solve a given problem, e.g. minimizing or optimizing a certain error-metric.
- **Population**: A set of initially random generated individuals of size n
- **Parent selection mechanism**: An algorithm selecting the best individuals to reproduce from the population. For example reproduction by copy such as bacteria or monogamous crossover of two parents like in mammals.
- **Variation operators, recombination and mutation**: Randomly replacing bits or tree-nodes with a certain probability and thus creating diversity within the population (See Fig. 1.5)
- **Survivor selection mechanism (replacement)**: Creating a new child-population

1.4.2 Symbolic Regression

Finding mathematical models for a given set of data is loosely speaking *the* key challenge in all of science. Traditionally, it is the task of the scientist to decide which function may predict

data the best. But even experts tend to have bias towards data when choosing functions to fit. Unlike traditional linear and non-linear regression methods, the aim of symbolic regression is not the sole parameter regression of predefined functions but also to find the whole symbolic mathematical representation of the underlying model.

Considering the one-dimensional case, the problem of symbolic regression is therefore defined by a number of pairs $(x_i, y_i) \in \mathbb{R} \times \mathbb{R}$, where $i = 1, ..., n$, and the task is to find a function $f : \mathbb{R} \to \mathbb{R}$, such that $f(x_i) = y_i$ for all $i = 1, ..., n$. In order to define a genetic algorithm for this problem we need to find a suitable representation. It is common to express the individuals in symbolic regression as syntax trees [18].

Initial expressions are randomly generated by combining mathematical building blocks like algebraic operators $+, -, \times, \div$ or analytical functions $\exp, \sin, \cos$, etc. We also need at least one variable x and some constants. To be able to evaluate the trees, it is further necessary to define the arity of the operator, e.g. exp is unary, + is binary, which is trivially self-evident.

For the variation of individuals, symbolic regression offers huge variety of algorithms [18]. However, it suffices to stick to standard operations of genetic programming. The generational algorithm used in this thesis is the described in [19]. It involves crossover *and* mutation operations, as well as a $1 : 1$ replacement ratio.

For parent selection, m individuals are randomly chosen to compete against each other in a tournament. Measured by their fitness, the best of those m individuals is allowed to mate with its consecutive tournament winner. The so-called one-point crossover is applied with a certain (usually high) probability P_{mate}: Two random nodes of the same arity in both parents get selected and combined, the other parts of the parental trees get discarded (See Fig. 1.4). After the generation of new individuals through crossover, the offspring population will be mutated with a (usually small) probability P_{mutate}: A randomly selected node is selected and replaced by a randomly generated sub-tree (See Fig. 1.5). The fitness of each individual is usually given as some form of curve fitting error. In most cases a simple χ^2-test is enough to achieve adequate results.

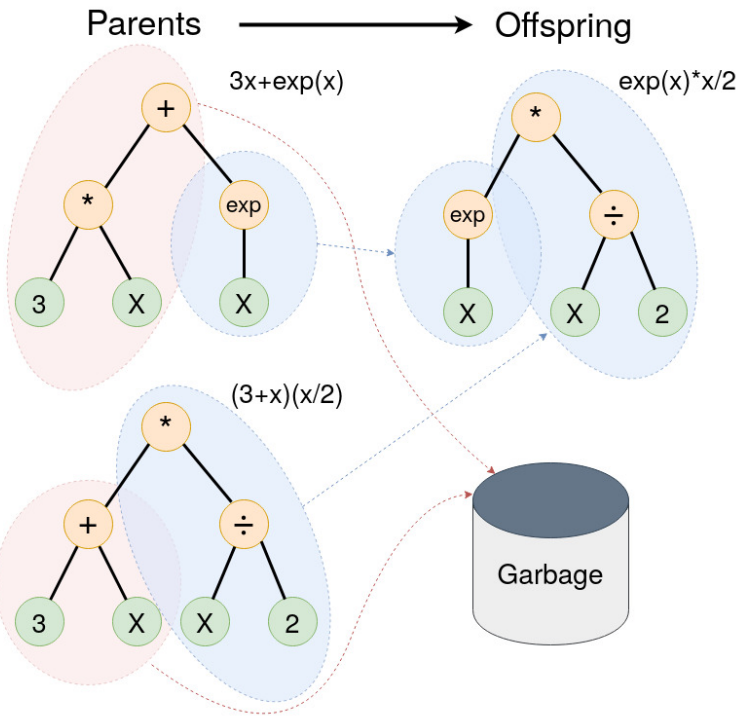

Figure 1.4 Example of subtree crossover. Note that the trees on the left are actually copies of the parents. So, their genetic material can freely be used without altering the original individuals. Adapted from [**?**]

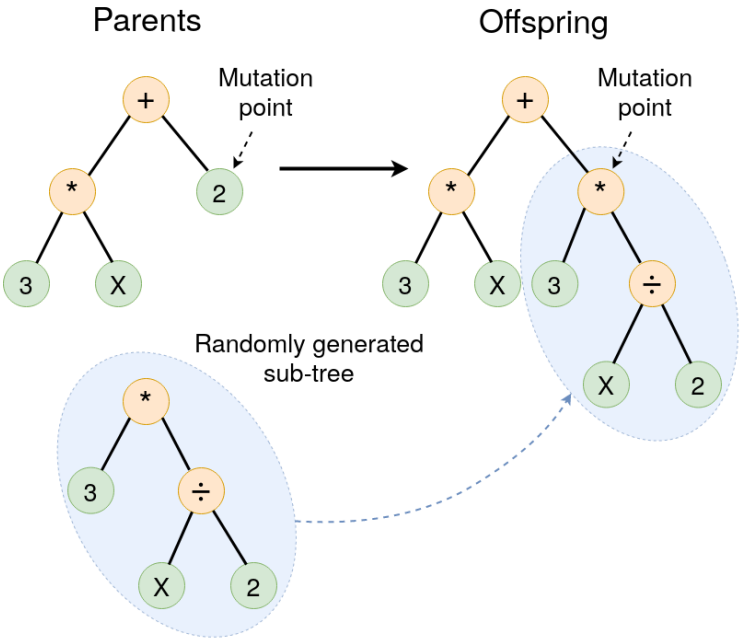

Figure 1.5 Example of one-point sub-tree mutation. A randomly selected tree-node gets replaced by a randomly generated sub-tree. Adapted from [**?**]

1.5 Bayesian Inference

Bayesian methods have become increasingly popular among cosmologists to infer parameters of models and select these models based on their observational evidence [20]. Not only did the cosmological models get more complex, but also the statistics evaluating those models needed to advance. The classical, so-called frequentist, approach of statistics involves many difficulties, which the Bayesian approach tries to avoid.

The Bayesian viewpoint is based on the tenet that "probability is a measure of the degree of belief about a proposition" [20]. This definition of probability applies to any event, also to those occurring only once (e.g. "what is the probability that it will rain tomorrow?"). "Randomness" is therefore only a consequence of our lack of information about the exact conditions of the system. Central for Bayesian inference is Bayes Theorem. The introduction here is only basic and we refer to more sophisticated literature [21, 22].

1.5.1 Bayes' Theorem

Bayes' theorem follows as a simple consequence of the Kolmogorov axioms of probability theory. For two random events A and B, the conditional probability of A, given the occurrence of B, is given by

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \tag{1.66}$$

where $P(A \cap B)$ is the joint probability of both A and B being true. Bayes' theorem basically states the symmetry of $P(A \cap B) = P(B \cap A)$, thus

$$P(A \mid B)P(B) = P(A \cap B) = P(B \cap A) = P(B \mid A)P(A) \tag{1.67}$$

$$\Leftrightarrow P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \tag{1.68}$$

We can expand Bayes' Theorem (1.68), giving

$$P(H \mid D, \mathcal{M}) = \frac{P(D \mid H, \mathcal{M})P(H \mid \mathcal{M})}{P(D \mid \mathcal{M})} \tag{1.69}$$

All probabilities are now further conditioned by a model $\mathcal{M}$ while the hypothesis H is expressed by the values of the parameters that the model $\mathcal{M}$ may have. The left-hand side $P(H \mid D, \mathcal{M})$ is called the posterior probability of the hypothesis taking the data D into account. It is proportional to the sampling distribution of the data $P(D \mid H, \mathcal{M})$ assuming the hypothesis is true, times the prior probability for the hypothesis $P(H \mid \mathcal{M})$. The prior represents our state of knowledge before seeing the data and takes up an important role in Bayesian inference, discussed in more detail in Ref. [20]. As a function of the hypothesis for fixed data, the sampling distribution is called likelihood function, where we will employ the notation

$$\mathcal{L}(H) \equiv p(D \mid H, \mathcal{M}) \tag{1.70}$$

As a function of H, the likelihood is *not* a probability distribution. The normalisation constant on the right-hand-side in the denominator is the marginal likelihood or Bayesian evidence, summing up all possible hypotheses

$$P(D \mid \mathcal{M}) \equiv \sum_H P(D \mid H, \mathcal{M}) P(H \mid \mathcal{M}) \tag{1.71}$$

This is interpreted as the likelihood for the data D under the model M, specified by all possible sets of parameters characterising the model.

1.5.2 Model selection

Assuming that the parameter space of the model M is continuous, i.e. the parameter θ varies within a parameter space $\Omega_{\mathcal{M}}$, the sum (1.71) turns into an integral over the parameter space, hence the evidence turns into

$$P(D \mid \mathcal{M}) \equiv \int_{\Omega_{\mathcal{M}}} P(D \mid \theta, \mathcal{M}) P(\theta \mid \mathcal{M}) \mathrm{d}\theta \tag{1.72}$$

When comparing two models, $\mathcal{M}_0$ versus $\mathcal{M}_1$, one is interested in the posterior odds, given by

$$\frac{P(\mathcal{M}_0 \mid D)}{P(\mathcal{M}_1 \mid D)} = B_{01} \frac{P(\mathcal{M}_0)}{P(\mathcal{M}_1)} \tag{1.73}$$

B_{01}	**Strength of evidence**
1 to 3	Barely worth mentioning
3 to 20	Positive
20 to 150	Strong
> 150	Very strong

Table 1.1 Empirical scale for evaluation the strength of evidence when comparing two models (so-called "Jeffreys' scale").

with the Bayes factor B_{01} being the ratio of the models' evidences:

$$B_{01} \equiv \frac{p(D \mid \mathcal{M}_0)}{p(D \mid \mathcal{M}_1)} \tag{1.74}$$

The Bayes factor is our relevant quantity to choose whether $\mathcal{M}_0$ or $\mathcal{M}_1$ is more strongly supported by the data, independent of the precise value of the model parameters.

Usually Bayes factors are interpreted against the empirically calibrated Jeffreys' scale [20] for the strength of evidence. This does not replace parameter inference, which is performed within each of the models separately.

1.5.3 Parameter Estimation

Let the probability distribution of data points y_i at x_i, when expected $y_\theta(x_i)$ with variance σ_i^2, be Gaussian distributed

$$P(y_i \mid y_\theta(x_i), \sigma_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(y_i - y_\theta(x_i))^2}{2\sigma_i^2}\right) \tag{1.75}$$

The classical approach for parameter estimation is the maximum likelihood method. Those parameters $\theta \in \Omega_\mathcal{M}$ of our model $y_\theta(x_i)$ which maximize the joint likelihood function are ought to be the "true" parameters

$$\theta_{\max} = \arg\max_{\theta \in \Omega_\mathcal{M}} \mathcal{L}(\{y_i\} \mid \theta) \tag{1.76}$$

Assuming statistical independence, we can express the likelihood function of our set of data points $\{y_i\}$ as the product of the individual univariante probability distributions (1.75)

$$\mathcal{L}(\{y_i\}) = \prod_{i=1}^{n} P(y_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(y_i - y_\theta(x_i))^2}{2\sigma_i^2}\right) \tag{1.77}$$

Turning the product over all data points in Eq. (1.77) into a sum in the exponential, we recovered the χ^2 dependency of the likelihood function, i.e. maximizing $\mathcal{L}$ is equivalent to minimizing χ^2

$$\mathcal{L} \propto \exp\left(-\frac{\chi^2}{2}\right), \quad \text{with} \quad \chi^2 = \sum_{i=1}^{n} \frac{(y_i - y_\theta(x_i))^2}{\sigma_i^2} \tag{1.78}$$

1.5.4 Markov Chain Monte Carlo

In general it turns out that evaluating the marginal likelihood integral (1.72) is rather difficult and numerically challenging, due to the multi-dimensional integration over the whole parameter space. The common way to tackle this problem is to sample the posterior using Markov Chain Monte Carlo (MCMC) methods [21].

A Markov chain is a stochastic process with a countable space $\Omega = \{\omega_1, \omega_2, ..., \omega_{n+1}\}$ of results and the Markov property

$$P(X_{n+1} = \omega_{n+1} \mid X_n = \omega_n, ..., X_1 = \omega_i) = P(X_{n+1} = \omega_{n+1} \mid X_n = \omega_n) \tag{1.79}$$

A Markov chain Monte Carlo algorithm defines a sequence, a "chain", of random variables $\{X_1, X_2, ..., X_n\}$, so that the probability of each step of the chain only depends on previous result.

A common example for a scheme achieving this desired property is the Metropolis-Hastings algorithm. Suppose $P(X)$ is the probability distribution we want to sample a parameter vector X from by such a Markov chain. Let further $G(X_{n+1} \mid X_n)$ be the conditional probability distribution for taking the step from X_{n+1} to X_n, the so-called proposal distribution. The Markov property (1.79) is achieved by choosing the transition probabilities

$$P_{ij} = \min\left(1, \frac{P(X_{n+1})}{P(X_n)} \frac{G(X_{n+1} \mid X_n)}{G(X_n \mid X_{n+1})}\right) \tag{1.80}$$

It says that the step from X_n to X_{n+1} is taken with certainty if $P(X_{n+1})$ is larger at the arrival point than it was at the starting point, but with a lower probability otherwise. How wide the steps are taken and how far they can reach is controlled by the proposal distribution, usually taken to be flat between reasonable parameter limits or multivariate Gaussians [21].

The start of the Markov chain will be some randomly chosen position in parameter space. However, the final outcome must not depend any more on this starting point. The chain can be shown to converge to the desired distribution after a certain burn-in phase. This part should be cut off for the resulting chain to be independent of the initial starting point and to hence achieve better results.

Once a Markov chain has been constructed, obtaining Monte Carlo estimates of expectations for any function of the parameters becomes a trivial task. The integral for the Bayesian evidence Eq. (1.72) can be approximated by a sum over all parameter samples obtained in the Markov chain

$$P(D \mid \mathcal{M}) = \int_{\Omega_{\mathcal{M}}} P(D \mid X', \mathcal{M}) P(X' \mid \mathcal{M}) dX' \approx \frac{1}{N} \sum_{i=1}^{N} P(D \mid X_i, \mathcal{M}) \tag{1.81}$$

This follows because the samples X_i are generated from the posterior by construction. Equality is given in the limit of large sample sizes N.

Chapter Two

IMPLEMENTATION AND RESULTS

The implementation of the evolutionary algorithm took place in two steps. The goal of the first part is to construct a symbolic regression algorithm able to find the right potential terms of classical Lagrangian systems with simulated experimental data. The second part is dedicated to finding suitable quintessence self interaction potentials with the previously implemented symbolic regression algorithm, constrained by the SNIa data and measurements of parameter values from the CMB/BAO.

2.1 Preparations

In preparation it is necessary to develop an intuition for symbolic regression and the parameters governing the evolution of the individual functions. As a light-weight framework for evolutionary algorithms we employ the python library DEAP [23]. The highly instructive documentation of DEAP offers a simple symbolic regression example, which is the starting point of the following constructions. DEAP comes with an excellent toolbox and a variety of simple evolutionary algorithms. They welcome the implementation of self-written methods, schemes and especially fitness functions. This is a great advantage over other EA frameworks. In the scientific community, DEAP is well-known not only for its versatility in evolutionary programming, but also for its remarkable parallelization scheme [24]. Since the large amount of individuals in a given population has to be calculated separately and independently, the calculation time can be divided by the number of available CPU's. This makes evolutionary programming viable on large computing clusters. Nevertheless, the handling of variables and memory in real multi-processing applications is a difficult task and should be done with care.

Parameter	Description	Typical value
n	Initial population size	$10^1 < n < 10^5$
m	Tournament size	$2 < m < 10$
g	Number of generations	$5 < g < 30$
P_{mate}	Crossover probability	$\sim 75\%$
P_{Mutate}	Mutation probability	$\sim 25\%$

Table 2.1 Empirical boundaries for the symbolic regression parameters.

2.2 Symbolic Regression

The ability of symbolic regression to find simple function terms [18], dynamical and chaotic system's differential equations [25], and even physical laws [5] has been demonstrated numerous times. Despite the rich literature for symbolic regression, the evolutionary dynamics of such a genetic algorithm can be counter intuitive. This issue can only be solved by countless rounds of trial and error. In the end, the successful termination of the algorithm depends significantly on the type of problem. Nevertheless, for the sake of orientation, heuristic bounds and fixed values for the evolutionary parameters are given in Table 2.1. The upper limits for the parameters n, m and g are mainly restricted by the computing resources and termination time. This can be very problematic in more challenging situations, since a sufficiently dense population is necessary to reach the needed diversity. The more individuals there are, the higher the chances that we can find the correct expression, which solves the symbolic regression problem.

Concerning the probabilities of mating and mutation, different views are found throughout the literature. Most authors agree, that the crossover probability of mating individuals should be high enough to achieve optimization, but not too high for the set to become overpopulated by a semi-good individual. Mutation, some authors argue, is not relevant for evolution at all [**?**]. In our experiments, however, we found that, in order to prevent the algorithm to get stuck in a

local minimum, it is necessary to allow a small mutation probability.

As a representation for the functions we decided to use the python library Sympy [26]. Sympy is advanced symbolic computer algebra system, able to perform all sorts of algebraic operations on symbolic mathematical expressions. This involves solving equations, simplification as well as symbolic differentiation and integration. If needed, the expression can be returned as an executable python function and integrated numerically.

Another important aspect of symbolic regression is the protection of individuals and the fitness function. The algorithm does not know per se whether an expression makes "sense" or not. Non-sense expressions, e.g. $f(x) = (x - x)/(x - x) = 0/0$ or $f(x) = \log(x - x)$, are mathematically not defined and return errors in Python and thus abort the algorithm. Hence, many exceptions have to be formulated for symbolic regression to not produce faulty individuals. In such a case, the defective individuals get excluded from the population by disfavoring them in the selection process with the worst fitness ($\pm\infty$ depending on the error metric).

2.3 Classical Lagrangian Potentials

As already mentioned before, the primary goal of this section is to construct a working algorithm able to find the right classical Lagrangian potential that optimally fits simulated data. This means that we first need to understand how the equations of motion of physical objects follow from a given potential.

The action S for any given mechanical system is defined by the time integral of the Lagrangian L, which is given by the difference of kinetic energy T and potential energy V

$$S = \int L \mathrm{d}t \quad \text{with} \quad L = T - V \tag{2.1}$$

From Hamilton's principle ($\delta S = 0$), we can derive the Euler-Lagrange equations

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = 0 \tag{2.2}$$

with the generalized coordinates q and their time derivative $\dot{q}$.

A very simple and well-studied Lagrangian mechanical system is the classical pendulum. We place the suspension of the pendulum in the coordinate origin. Without loss of generality, we can assume the pendulum to swing in the x-y-plane, hence the position and velocity vector for its mass m at the end of a mass-less connection of length l are constrained to move on a circle. Therefore, they can be parametrized by the pendulum's angle φ

$$\mathbf{r} = l\,(\sin\varphi, \cos\varphi, 0)^T \qquad \Rightarrow \qquad \dot{\mathbf{r}} = l\dot{\varphi}(\cos\varphi, -\sin\varphi, 0) \tag{2.3}$$

Note that the spatial position of the pendulum is now described by a single parameter, our generalized coordinate. This reduces our system to a one dimensional problem.

The pendulum is placed in a gravitational field $\mathbf{g} = g\mathbf{e}_y$ pointing in the y direction of our reference frame. Its potential is therefore given by

$$V(\mathbf{r}) = mg\mathbf{e}_y \cdot \mathbf{r} = mgl\cos\varphi \tag{2.4}$$

We can now construct the Lagrangian for our pendulum as

$$L = \frac{1}{2}m\dot{\mathbf{r}}^2 - V(\mathbf{r}) = \frac{1}{2}ml^2\dot{\varphi}^2 + mgl\cos\varphi \tag{2.5}$$

Plugging this Lagrangian into the Euler-Lagrange Eq. (2.2) yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial\dot{\varphi}} - \frac{\partial L}{\partial\varphi} = 0 \quad \Rightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t}\left(ml^2\dot{\varphi}\right) + mgl\sin\varphi = 0 \tag{2.6}$$

$$\Rightarrow \ddot{\varphi} = -\frac{g}{l}\sin\varphi \tag{2.7}$$

This is the equation of motion of our classical pendulum, a second-order non-linear differential equation, which implies that it can be expressed by two first-order differential equations. We introduce the angular velocity ω, and so we obtain

$$\dot{\varphi} = \omega \tag{2.8}$$

$$\dot{\omega} = -\frac{g}{l}\sin\varphi \tag{2.9}$$

After choosing suitable initial conditions for φ and ω at time $t = 0$, this system of first-order differential equations can be integrated numerically. We employ the Solve_ivp ODE solver by Scipy [27]. This solver is based on a explicit Runge-Kutta method with 4th order in error accuracy and 5th order step-size.

Having this analytically solved system and its numerical integration at hand, we can generate an artificial data set for the pendulum's angle φ_i^0 and its corresponding velocity ω_i^0 at time steps t_i. To emulate the natural uncertainty of measurements, all data points are salted with additional Gaussian noise with mean $\mu = 0$ and standard deviation σ

$$\varphi_i = \varphi_i^0 + \mathcal{N}(0, \sigma) \tag{2.10}$$

For our symbolic regression algorithm to be able to fit the data with potential terms as individuals populating the evolutionary dynamics, it is also necessary to estimate suitable initial conditions. In this case, the values for φ_0 and ω_0 are treated as ordinary fit parameters, which can be estimated by MCMCs and the maximum likelihood method (see section 1.5)

As soon as the Markov chain exists, assigning a fitness to each individual is done by calculating their Bayesian evidence (1.81). Each potential can thus be assigned its own inherent fitness value deciding which individual ought to survive. Estimating the initial values for differential equations is a time-expensive task, because, for every single step, the whole procedure of numerically solving the ODE can easily take seconds to finish. However, the simple case of a classical pendulum can be treated efficiently and its symbolic potential term is quickly found. The pipeline of the algorithm can be seen in Fig. 2.1 as a flowchart.

We now explore the ability of parameter estimation with more intricate mechanics. We choose a system of two coupled pendula of same length l and following Lagrangian

$$L = T - V = \frac{1}{2}m(\dot{\varphi}_1^2 + \dot{\varphi}_2^2) - \frac{mg}{2l}\varphi_1^2 - \frac{mg}{2l}\varphi_2^2 - \frac{1}{2}k(\varphi_2 - \varphi_1)^2 \tag{2.11}$$

where m_1 and m_2 are the masses of the pendula. k is the coupling-constant between the two pendula. The equations of motion for this system can be derived with the Euler-Lagrange

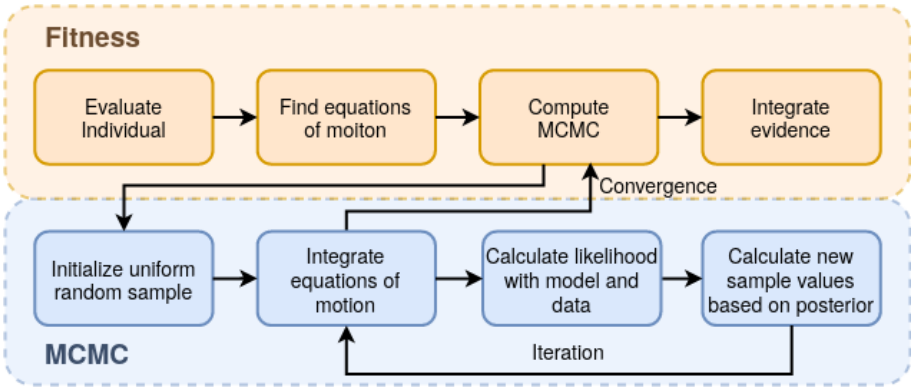

Figure 2.1 Flowchart of the fitness function for classical Lagrangian potentials. On the top is the main fitness function. On the bottom one can see a schematic overview of the MCMC algorithm.

Eq. (2.2) again and are given by

$$m\ddot{\varphi}_1 = -\frac{mg}{l}\varphi_1 + k(\varphi_2 - \varphi_1) \tag{2.12}$$

$$m\ddot{\varphi}_2 = -\frac{mg}{l}\varphi_2 - k(\varphi_2 - \varphi_1) \tag{2.13}$$

We estimate the proper initial conditions for the equations of motion by Bayesian Inference. We therefore treat the initial conditions of the differential equations as fit parameters and estimate them by a MCMC. The samples will be distributed according to the posterior distribution of the model given the dataset. Usually a flat prior (Jeffreys prior) in a given limit for the parameters is the favored choice for the prior distribution [22]. This ensures, a priori, all values in the given limits to be equal probable. The maximum of the posterior distribution will therefore be the best fit-parameter, also maximizing the likelihood.

Finding the right initial conditions works surprisingly well for dynamical systems such as a coupled pendulum (See Fig. 2.2). Even for very noisy and low sampled data points, it is possible to infer the initial conditions for the equations of motion of a simple classical mechanical system. After a certain burn-in phase (See sec. 1.5.4) the Markov chains converges around the posterior maximum (See Fig. 2.3) and the sampled initial conditions most supported by the data can be inferred easily. Exploitation of multiprocessing is unfortunately not possible for MCMCs, due to the Markov property (1.79). The handling of memory for MCMC tasks over

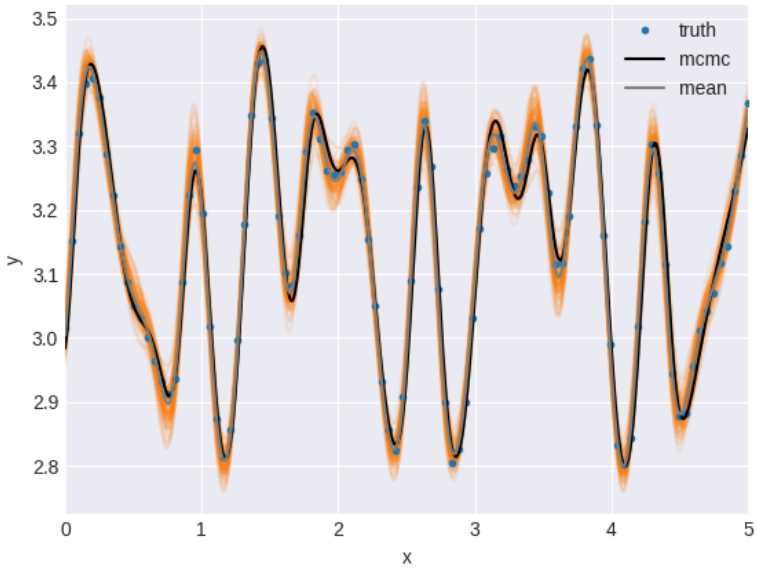

Figure 2.2 Angle φ of a coupled pendulum with $m_1 = m_2 = l = 1$. The truth is given as noisy simulated points on a evenly distributed grid. The black line is the MCMC sample, i.e. initial conditions, with maximum likelihood. In grey is the statistical mean and in orange the 1σ area of uncertainty of the MCMC samples.

parallel processes is in general even slower than conventional serial execution. This limitation tightens with model complexity, which makes initial condition estimation by MCMC in symbolic regression only possible for simple examples, such as the single and coupled pendulum.

2.4 Quintessence Potentials

Finding the correct quintessence interaction potential for a scalar field to match the distance and redshift relationship is undoubtedly a more difficult task than finding simple classical potentials. It involves solving a system of three differential equations, multiple interpolations and integration. We will start by finding equations of motion for individual quintessence potentials.

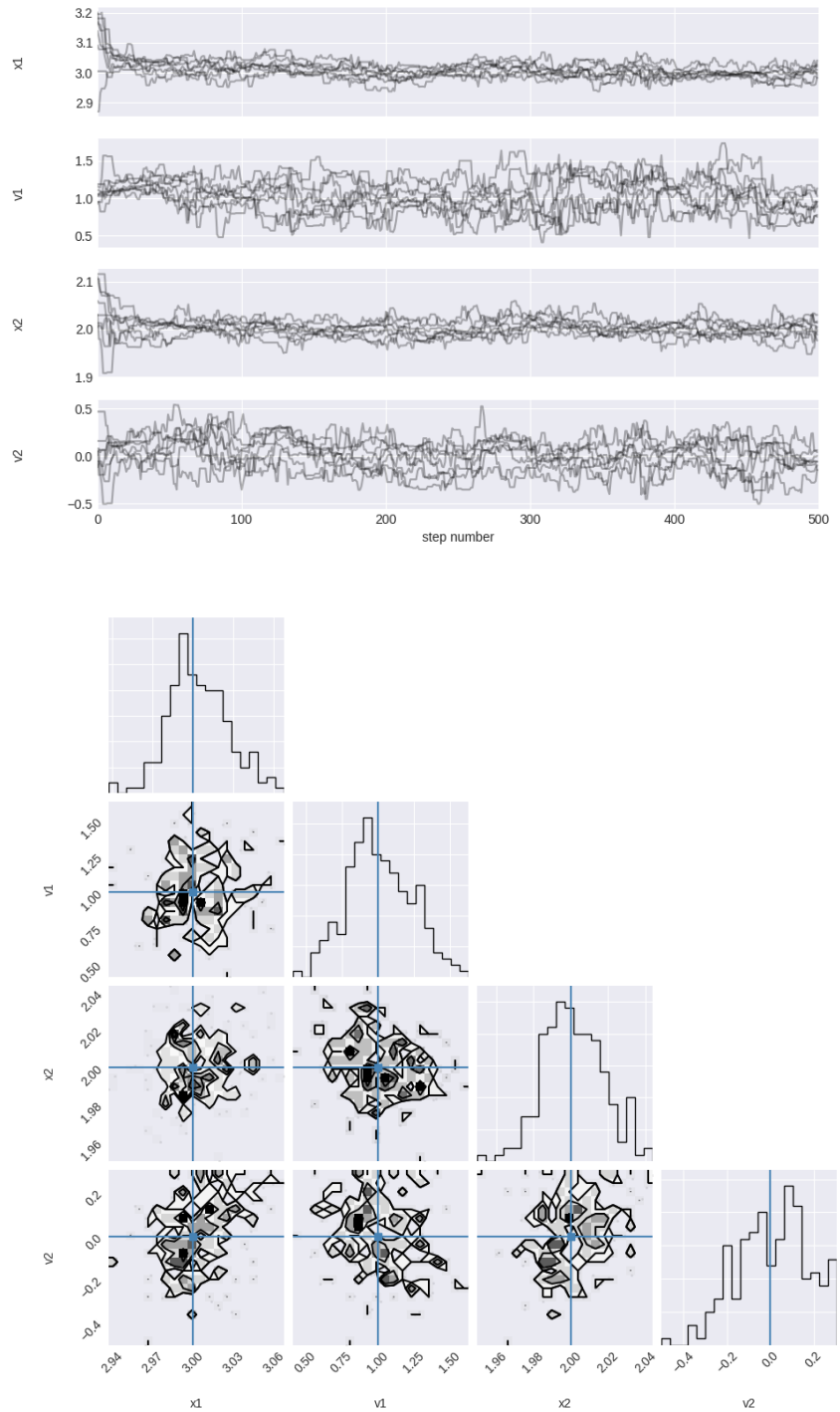

Figure 2.3 Monte Carlo Markov Chain for the coupled pendulum. On the top one can see the values of the samples against the number of steps of the chain. All walkers of the chain converged around a maximum value and thus sample the posterior sufficiently. On the bottom one can see the covariances of the samples as histograms.

2.4.1 Equations of motion

The SNIa data will be the measure for the evidence of our individual models. This prompts us to express our fields equation of motion (1.64) and the Friedmann Eqs. (1.62) and (1.63) in terms of redshift z instead of the time t. We will again make use of the coordinate transformation

$$\mathrm{d}t = -\frac{\mathrm{d}z}{H(z)(1+z)} \tag{2.14}$$

Expressing the time-derivative in the second Friedmann equation for a Universe with Quintessence and cold dark matter gives the following z-depended acceleration equation

$$-H(1+z)\frac{\mathrm{d}H}{\mathrm{d}z} = -4\pi G\left[\left(\frac{\mathrm{d}\phi}{\mathrm{d}z}\right)^2 H^2(1+z)^2 + \rho_M\right] \tag{2.15}$$

For the numerical implementation, we employ the expansion function $E = H/H_0$, since this is the relevant quantity for integrating the luminosity distance (1.42) of our model. The physical scaling with natural constants is done at the end. Replacing H by E in Eq. (2.15) and isolating its z-derivative hence yields

$$\Rightarrow \frac{\mathrm{d}E(z)}{\mathrm{d}z} = 4\pi G E(z)(1+z)\left(\frac{d\phi}{\mathrm{d}z}\right)^2 + \frac{3\Omega_M^0}{2E(z)}(1+z)^2 \tag{2.16}$$

where we used $\rho_M(z)/\rho_{\mathrm{crit}} = \Omega_M^0(1+z)^3$ from Eqs. (1.16) and (1.19).

In the next step, we also transform the equation of motion (1.64) for the quintessence field. We start by changing the time dependence of $\ddot{\phi}$ to redshift z, i.e. we need to apply the transformation (2.14) twice under consideration of the chain rule

$$\ddot{\phi} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}}{\mathrm{d}t}\phi\right) = -H(1+z)\frac{\mathrm{d}}{\mathrm{d}z}\left(-H(1+z)\frac{\mathrm{d}}{\mathrm{d}z}\phi\right) \tag{2.17}$$

$$= -H(1+z)\left(-\frac{\mathrm{d}H}{\mathrm{d}z}(1+z)\frac{\mathrm{d}\phi}{\mathrm{d}z} - H\frac{\mathrm{d}\phi}{\mathrm{d}z} - H(1+z)\frac{\mathrm{d}^2\phi}{\mathrm{d}z^2}\right) \tag{2.18}$$

$$= H(1+z)^2\frac{\mathrm{d}H}{\mathrm{d}z}\frac{\mathrm{d}\phi}{\mathrm{d}z} + H^2(1+z)\frac{\mathrm{d}\phi}{\mathrm{d}z} + H^2(1+z)^2\frac{\mathrm{d}^2\phi}{\mathrm{d}z^2} \tag{2.19}$$

The friction and the force term in the equations of motion (1.64) simply transform to

$$-3H\frac{d\phi}{\mathrm{d}t} - \frac{\mathrm{d}V(\phi)}{\mathrm{d}\phi} = 3H^2(1+z)\frac{\mathrm{d}\phi}{\mathrm{d}z} - \frac{\mathrm{d}V(\phi)}{\mathrm{d}\phi} \tag{2.20}$$

Combining Eq. (2.19) and Eq. (2.20), we can now isolate ϕ'' (using $'$ as a short notation for a z-derivative):

$$\Rightarrow \phi'' = \left(\frac{2}{1+z} - \frac{H'}{H}\right)\phi' - \frac{\mathrm{d}V(\phi)}{\mathrm{d}\phi}\frac{1}{H^2(1+z)^2} \tag{2.21}$$

Defining the "velocity" ψ of the field ϕ with respect to the redshift z

$$\psi := \frac{\mathrm{d}\phi}{\mathrm{d}z}\,, \tag{2.22}$$

we can now rewrite the second-order differential equation (2.21) into two first-order differential equations and additionally the background expansion

$$\phi' = \psi \tag{2.23}$$

$$\psi' = \left(\frac{2}{1+z} - \frac{E'(z)}{E(z)}\right)\psi - \frac{\mathrm{d}V(\phi)}{\mathrm{d}\phi}\frac{1}{E^2(z)(1+z)^2} \tag{2.24}$$

$$E' = 4\pi G E(z)(1+z)\psi^2 + \frac{3\Omega_M^0}{2E(z)}(1+z)^2 \tag{2.25}$$

This system of differential equations can be integrated numerically (analogously to sec. 2.3) and is the heart of the following algorithm.

As one can see, the shape of the acceleration of ϕ can take a variety different shapes, depending on the derivative of the interaction potential $V(\phi)$. Constructing the differential equations for the algorithm involves therefore a symbolic differentiation of the potential which is done by a Sympy function. This constrains the form of the individual potentials, as they have to be differentiable at least once, which is already a condition for the slow-roll parameters (1.65). To protect the algorithm of such individuals, exceptions for non-differentiable potentials are formulated. As in section 2.3, such faulty individuals get the worst fitness assigned.

2.4.2 Initial Conditions

At this point, we can infer the initial conditions for the equations of motion by MCMC and maximum likelihood techniques as in section 2.3. Particularly for the procedure presented

here, this is a very time consuming exercise. Apart from the fact, that automatically choosing a dependable prior for every individual potential seems impossible, most individuals need up to 10^5 steps to get the relevant parts of the posterior distribution sampled by the MCMC algorithm. This involves a couple of seconds in every step to evaluate the posterior and hence, it can take hours to calculate the initial conditions of just a single individual. Even on a large cluster, a rich population over many generations may take days and weeks to terminate.

The goal of this thesis is the construction of a working symbolic regression algorithm. Therefore, we will exploit the simplification of choosing the initial conditions of the equations of motion by constrains of measurements of Ω_M^0 and w_ϕ, sufficient to demonstrate the underlying principles.

To match the measurements of the parameter values today we have to "translate" them into the initial conditions for ϕ and ψ. We can transform Eq. (1.60) for the equation of state parameter of the quintessence field to

$$w_\phi^0 = \frac{\frac{1}{2}\psi_0^2 H_0^2 - V(\phi_0)}{\frac{1}{2}\psi_0^2 H_0^2 + V(\phi_0)} \tag{2.26}$$

with ϕ_0 and ψ_0 being the field and its velocity values today. For simplicity we will assume

$$w_\phi^0 = -1 \tag{2.27}$$

Now, we can find the first initial condition easily, since ψ_0 needs to vanish. This follows from Eq. (2.26) under assumption (2.27):

$$\Rightarrow \quad \frac{1}{2}\psi_0^2 H_0^2 - V(\phi_0) = -\frac{1}{2}\psi_0^2 H_0^2 - V(\phi) \tag{2.28}$$

$$\Leftrightarrow \quad \psi_0^2 H_0^2 = 0 \qquad \Leftrightarrow \qquad \psi_0 = 0 \tag{2.29}$$

Another constraint is given by the density parameters. Since the sum of all parameters have to be equal to one by definition (See Eq. (1.21)), we can express Ω_ϕ^0 by its corresponding energy density today

$$\Omega_M^0 + \Omega_\phi^0 = \Omega_M^0 + \frac{\frac{1}{2}\psi_0^2 H_0^2 + V(\phi_0)}{\rho_{\text{crit}}^0} = 1 \tag{2.30}$$

To determine the initial condition for ϕ_0 with this relation, we can use the result of ψ_0 from Eq. (2.29). ϕ_0 can be obtained by inverting the potential V

$$\Rightarrow \quad \phi_0 = V^{-1}\left(\rho^0_{\text{crit}}(1-\Omega^0_M) - \frac{1}{2}\psi_0^2 H_0^2\right) \overset{\psi_0=0}{=} V^{-1}\left(\rho^0_{\text{crit}}(1-\Omega^0_M)\right) \tag{2.31}$$

This demands the invertibility of the potential and is thus another constraint on the possible function candidates.

The simple case of assuming $w^0_\phi = -1$ is the "vanilla" version of finding the initial conditions, because for this choice, good fitting solutions are generated at a very low computational cost. The inversion of the potential can be done symbolically with Sympy. However, the assumption (2.27) only recovers the cosmological constant case, i.e. a constant equation of state parameter w_ϕ. We can achieve an enhancement in the dynamics of $w_\phi(z)$ by setting w^0_ϕ slightly off minus one. This involves solving a non-linear system of equations, namely Eq. (2.26) and (2.30) simultaneously, but with $w_\phi \gtrsim -1$. This is done by another Scipy module called fsolve, an advanced Newton-algorithm for finding the roots of non-linear equations. Unfortunately, achieving convergence of the Newton-algorithm in a given amount of iterations is again governed by the choice of initial values, and even then it is not guaranteed.

2.4.3 Integration and Interpolation

We want to select our individual potentials based on their likelihood to produce the SNIa data, given as a set of n distance moduli $\hat{\mu}_j$ at redshifts $\hat{z}_j$, where $j = 0, ..., n$. For this matter, we need to calculate μ with Eq. (1.43) for any given individual potential, i.e. it is inevitable to interpolate both, the inverse expansion function and the resulting integral involved in the computation of the luminosity distance (1.42) and the distance modulus.

Reasons for this complication are predominantly found in the random distribution of the redshifts $\hat{z}_j$ of the SNIa data and the implicit z-dependence of the integral in d_L. To suppress error accumulation while integrating numerically, it is essential to take small and even step sizes, i.e.

the linear space of integration z_i is an evenly distributed array with step-size $\Delta z = (\hat{z}_n - \hat{z}_n)/N$ and $N > n$. This array does evidently not match with the random SNIa redshifts $\hat{z}_j$, since $i \neq j$ in general.

However, interpolation and numerical integration are cheap operations and can be performed with nearly arbitrary precision. To implement this, it is sufficient to interpolate our solution for the expansion function with cubic splines.

Integrating the equations of motion (2.25) with the initial conditions calculated from w_ϕ^0 and Ω_M^0 returns a set E_i at a given grid of z_i with $i = 0, ..., N$. We will use the same z_i as a support for the first interpolation

$$\frac{1}{E(z)} \approx \text{Spline}\left(\{z_i\}, \frac{1}{\{E_i\}}\right) \quad \text{for } z \in [z_0, z_N] \tag{2.32}$$

Note that this function approximates the data points well only within the boundaries given by z_i.

We can now construct a set of integration support points $\{x_k\}$ within $[\hat{z}_0, \hat{z}_n]$, such that $\hat{z}_0 = x_0 < x_1 < ... < x_{M-1} < x_M = \hat{z}_n$ and step-size $\Delta x = (\hat{z}_n - \hat{z}_j)/M$, to make use of the Trapezoidal rule and approximate the integral over the whole redshift region

$$\int_0^z \frac{\mathrm{d}z'}{E(z')} \approx \frac{\Delta x}{2} \sum_{k=1}^{M} \left[\frac{1}{E(x_{k-1})} + \frac{1}{E(x_k)}\right] \tag{2.33}$$

The redshift dependency of Eq. (1.42) is in the upper boundary limit, i.e. every single point z of $d_L(z)$ involves implicit integration. We achieve a functional expression of the integral by another cubic spline interpolation on the previously generated support points x_k

$$\text{Integral}(z) \approx \text{Spline}\left(\{x_k\}, \int_0^{\{x_k\}} \frac{\mathrm{d}z'}{E(z')}\right) \quad \text{for } z \in [\hat{z}_0, \hat{z}_n] \tag{2.34}$$

This finally yields the function for calculating the luminosity distance

$$d_L(z) = \frac{c(1+z)}{H_0}\text{Integral}(z) \tag{2.35}$$

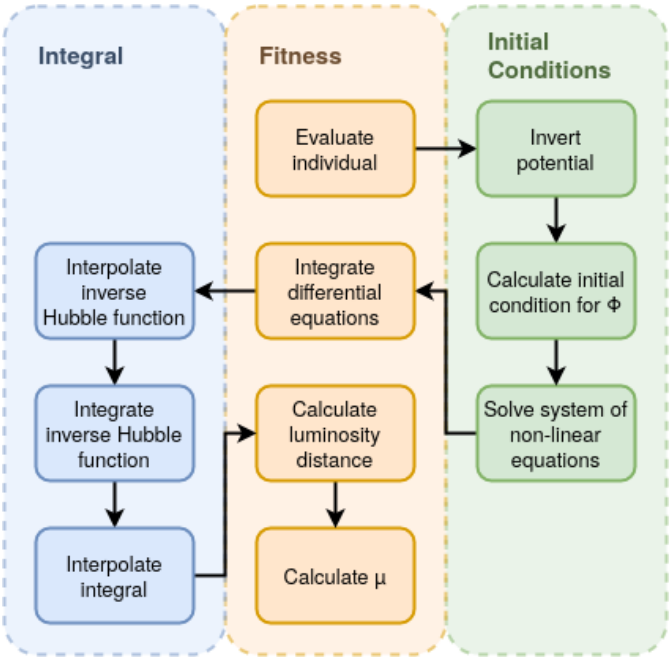

Figure 2.4 Flowchart of the pipeline to calculate the distance modulus μ from a given individual potential. In the middle is the relevant part for the fitness, on the right the initial conditions for a given potential get calculated and on the left the inverse integral of the expansion function is interpolated

Note that here the natural constants are reintroduced, so d_L is scaled physically again. As a last step we generate the set of distance moduli with our given set of SNIa redshifts $\hat{z}_j$

$$\mu_j = \mu(\hat{z}_j) = 5 \log_{10}\left[d_L(\hat{z}_j)\right] + 5 \tag{2.36}$$

These μ_j, generated from our quintessence model, can now finally be used for the statistical inference from the SNIa data $\hat{\mu}_j$. This concludes the necessary preparations for our experiments with the symbolic regression algorithm searching for suitable quintessence potentials. The pipeline of the evaluation algorithm is schematically shown in Fig. 2.4

2.4.4 Experiments

In order to test our construction, we first study the results for a given potential. We choose the cosmon potential [28], given by

$$V(\phi) = M^4 \exp\left(-\frac{\alpha\phi}{M}\right) \tag{2.37}$$

Primitives	$+, -, \times, \div, \exp, \cos, \sin$
Terminals	ϕ, M, π

Table 2.2 Function primitives and terminals for symbolic regression.

where we define $M^2 \equiv M_p^2/16\pi$ and $\alpha > 0$ is a free parameter of the model. The solution for the distance modulus μ of the cosmon potential can be seen in Fig. 2.5.

The equation of state parameter w_ϕ for this potential can be calculated with Eq. (1.60) and is plotted as a function of redshift z in Fig.2.5. As expected, w_ϕ approaches -1 asymptotically for $z = 0$ from above and hence acts like the cosmological constant Λ at late times. However, the time at which w_ϕ crosses $-1/3$ and thus starts the cosmic acceleration is very dependent on the choice of w_ϕ^0. The smaller the difference of w_ϕ^0 to -1 (See condition (1.47), the bigger is the redshift z, at which quintessence becomes repulsive. The energy density for our quintessence field starts at $\Omega_\phi^0 = (1 - \Omega_M^0)$, as intended (See Fig. 2.5). This demonstrates, that the algorithm works and that it generate reasonable results.

We continue our experiments by letting the algorithm search for quintessence potentials freely. We choose a simple set of functions and constants as primitives and terminals of our individual potential presented in Table 2.2.

Most potentials, resulting from combination of those building blocks by the symbolic regression algorithm, are nonsensical. Not only are most of them undefined in a mathematical sense, but for many individual potentials, the non-linear system of equation (2.26) and (2.30) is mostly not even solvable, i.e. our Newton-algorithm will not converge. We have to sort those potentials out, which inevitably shrinks the size of the population dramatically. Only a few potentials survive and can therefore reproduce. To counter this problem we need to scale the parameters of the symbolic regression algorithm up. Unfortunately, the limit of this scaling is reached quickly and hence only some simple potentials can be generated with limited computing resources. A collection of the best fit individuals can be found in Table 2.3

Many individual potentials are invertible and finding the right initial conditions may be possi-

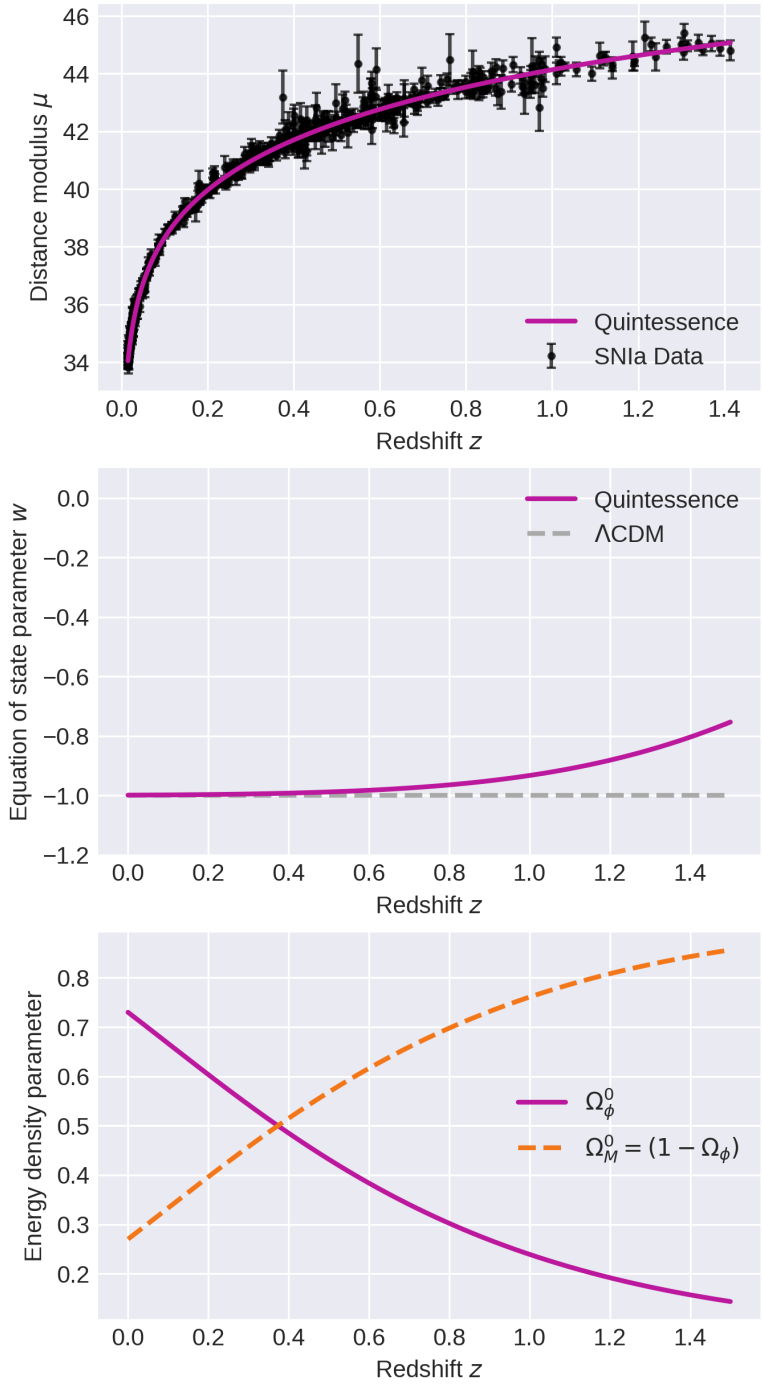

Figure 2.5 Solution for the Cosmon potential $V(\phi) = M^4 \exp(-\alpha\phi/M)$ with $\alpha = 1$, $\Omega_M^0 = 0.27$ and $w_\phi^0 = -0.999$. Top: Distance modulus against redshift with SNIa data. Middle: Equation of state parameter against redshift. Bottom: Energy density parameter with $\Omega_M(z) + \Omega_\phi(z) = 1$

Potential	Likelihood $\mathcal{L}$
$M^4\phi^{-2}$	117.0568012
$M\phi^{-1}$	117.0568012
$M\exp(-\phi)$	117.0548650
$M^2\phi^2$	117.0050210
$M\phi$	116.9334976

Table 2.3 Solutions of five runs with $n = 1000$, $g = 10$, $P_{\text{mate}} = 0.75$ and $P_{\text{mutate}} = 0.25$. Ordered by their maximum likelihood with Eq. (1.77) from top to bottom. Likelihood for ΛCDM is 117.34608028.

ble, but this does not guarantee the physical plausibility. Some potentials possess a minimum at an unfortunate region for the field ϕ. This often results in oscillatory phenomena. The condition (1.47) must hold true for the potential to realize the desired properties. Reaching automation in avoiding other potentials is difficult due to the rich dynamics, which the equations of motion (2.25) can produce with different potentials.

Over the experiments with the algorithm, no potential demonstrated significant advantage over the likelihood of the ΛCDM model with the same value for Ω_M^0. This finding favors Λ as an explanation for dark energy, despite its problems mentioned in section 1.3.1. However, the differences of the quintessence models to the ΛCDM case are much lower than the 1σ uncertainty of the SNIa measurements (See Fig. 2.6). Hence, at the moment we cannot decide conclusively from the the data which model describes our cosmos the best, until observational precision improves.

However, a scalar field governing the accelerated expansion of space is still a reasonable model for dark energy, since we recover the constant equation of state, i.e. a field behaving "just like" Λ, for a suitable interaction potential $V(\phi)$. Combination of different measurements (SNIa + CMB + BAO) could bring more statistical certainty and thus more reliable candidates for the potential in the process of symbolic regression .

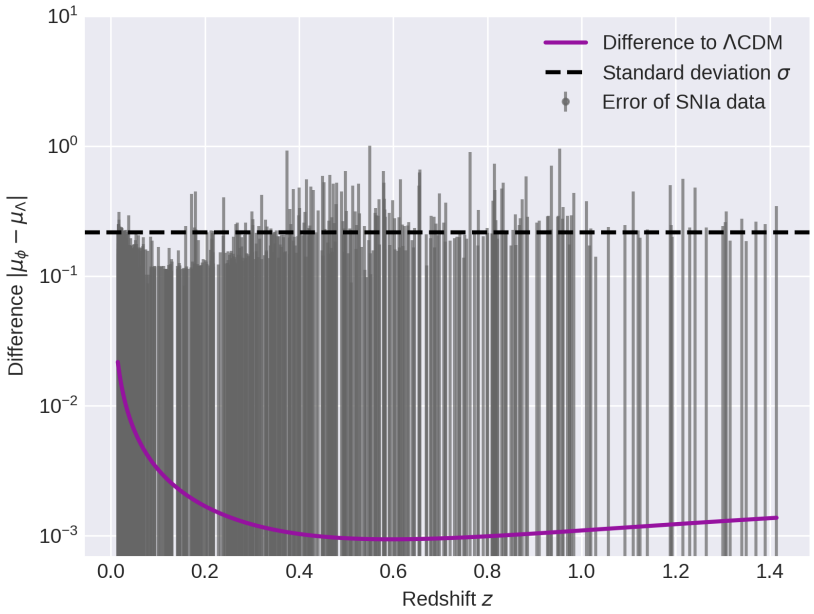

Figure 2.6 Plot of the difference between the ΛCDM model and the quintessence model for the cosmon potential $V(\phi) = M^4 \exp(-\alpha\phi/M)$ with $\alpha = 1$, $\Omega_M^0 = 0.27$ and $w_\phi^0 = -0.999$. Dashed black is the 1σ uncertainty of the SNIa measurements.

Chapter Three

CONCLUSION

In order to advance our understanding of the fundamental nature of our Universe, we must employ the most powerful statistical tools available. Symbolic regression and Bayesian inference are methods to automatically find reasonable mathematical expressions constrained by observational data. Without doubt, symbolic regression is a promising approach and with a sufficient amount of computing resources, it should be possible to solve many problems also beyond the ones presented in this thesis. If the type of problem can be expressed as an optimization task, e.g. finding the right functional expression to a given set of data, symbolic regression has the ability to find the right symbolic term with high precision. This is a great advantage over traditional regression techniques, since the choice of the function is determined in a completely data-driven approach. In our quest for insight about the cosmos, this framework can potentially guide our path to theoretical advancements, through a statistical analysis of observational data.

As discussed in this thesis, a combination of symbolic regression and Bayesian inference is capable of finding symbolic potential terms for both classical mechanical systems and the quintessence field. Its reconstruction capability, however, does not by itself prove the physical "truth" of the fittest individual potentials. The differences between the best quintessence potentials presented in section 2.4.4 are orders of magnitude below the 1σ uncertainty of the supernova type Ia data. Regarding classical systems such as the pendulum in section 2.3, several potentials fulfill the constraints by the data nearly equally well, despite their diverse genotypes. Hence, we have to include prior knowledge unconditionally, which involves a choice of a suitable set of function primitives, boundaries for the initial conditions and free

model parameters.

Having a reliable symbolic regression application at hand, further research should explore extended evolutionary algorithms with age-based optimization [29] and advanced recombination and mutation algorithms.

The Bayesian parameter estimation of initial conditions for the dynamical differential equations provides promising results for classical mechanical systems. Since the executing Monte Carlo Markov Chain depends on previous calculations, it is not possible to exploit parallelization schemes as with an evolutionary algorithm. We introduced the construction of a pipeline of algorithms in order to evaluate quintessence potentials in section 2.4. The central limitation of the success is the computing strength of the resources available. Since it is desirable to guarantee reasonable and physical results in the landscape of otherwise nonsensical potentials generated by the algorithm, an implementation for multi-processor computing on larger clusters will be necessary. The parallel computation of many individuals over several generations will ensure the diversity in the population of potentials necessary for evolutionary algorithms to select the best individuals provided by the data and, at the same time, an efficient time/complexity ratio.

Further, it may be an interesting endeavor to infer the initial conditions of the equations of motion of the quintessence field for every individual potential separately from the data. The parameters w_ϕ^0 and Ω_M^0 can have different values, depending on the cosmological model we assume. The importance of finding the right initial conditions for the quintessence equations of motion should therefore not be underestimated. The initial conditions are in general not the only free parameters of the system. For example, the parameter in the cosmon potential (2.37) is free and should in principle be determined together with the initial conditions for the field. Despite all the difficulties discussed in this thesis, the success of automatically modelling potentials for a given set of data is motivation to further facilitate and advance the statistical techniques as well as machine learning applications to gain more insight in physical questions.

REFERENCES

[1] A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, B. Leibundgut, M. M. Phillips, D. Reiss, B. P. Schmidt, R. A. Schommer, R. C. Smith, J. Spyromilio, C. Stubbs, N. B. Suntzeff, and J. Tonry, "Observational evidence from supernovae for an accelerating universe and a cosmological constant," *The Astronomical Journal* **116** no. 3, (Sep, 1998) 1009–1038. https://doi.org/10.1086%2F300499.

[2] S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, and et al., "Measurements of and from 42 high-redshift supernovae," *The Astrophysical Journal* **517** no. 2, (Jun, 1999) 565–586. http://dx.doi.org/10.1086/307221.

[3] B. Ratra and P. J. E. Peebles, "Cosmological consequences of a rolling homogeneous scalar field," *Phys. Rev. D* **37** (Jun, 1988) 3406–3427. https://link.aps.org/doi/10.1103/PhysRevD.37.3406.

[4] C. Wetterich, "Cosmology and the fate of dilatation symmetry," *Nuclear Physics B* **302** no. 4, (Jun, 1988) 668–696. http://dx.doi.org/10.1016/0550-3213(88)90193-9.

[5] M. Schmidt and H. Lipson, "Distilling free-form natural laws from experimental data," *Science* **324** no. 5923, (2009) 81–85, `https://science.sciencemag.org/content/324/5923/81.full.pdf`. https://science.sciencemag.org/content/324/5923/81.

[6] A. A. Penzias and R. W. Wilson, "A Measurement of Excess Antenna Temperature at 4080 Mc/s.," **142** (July, 1965) 419–421.

[7] D. J. Eisenstein, I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, R. C. Nichol, R. Scranton, H. Seo, M. Tegmark, Z. Zheng, and et al., "Detection of the baryon acoustic peak in the large-scale correlation function of sdss luminous red galaxies," *The Astrophysical Journal* **633** no. 2, (Nov, 2005) 560–574. http://dx.doi.org/10.1086/466512.

[8] M. Bartelmann, *General Relativity*. Heidelberg University Publishing, Heidelberg, 2019. http://doi.org/10.17885/heiup.534.

[9] L. Amendola and S. Tsujikawa, *Dark Energy: Theory and Observations*. Cambridge University Press, 2010.

[10] G. F. Hinshaw, "Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Results," .

[11] M. Bartelmann, *Cosmology*. Lecture Notes, Heidelberg, 2018. https://heibox.uni-heidelberg.de/f/e1e57faba9a44eb88692/.

[12] S. Chandrasekhar, "Xlviii. the density of white dwarf stars," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **11** no. 70, (1931) 592–596, https://doi.org/10.1080/14786443109461710. https://doi.org/10.1080/14786443109461710.

[13] A. Einstein and W. de Sitter, "On the Relation between the Expansion and the Mean Density of the Universe," *Proceedings of the National Academy of Science* **18** no. 3, (Mar., 1932) 213–214.

[14] N. Suzuki, D. Rubin, C. Lidman, G. Aldering, R. Amanullah, K. Barbary, L. F. Barrientos, J. Botyanszki, M. Brodwin, N. Connolly, and et al., "Thehubble space

telescopecluster supernova survey. v. improving the dark-energy constraints abovez> 1 and building an early-type-hosted supernova sample," *The Astrophysical Journal* **746** no. 1, (Jan, 2012) 85. http://dx.doi.org/10.1088/0004-637X/746/1/85.

[15] S. Weinberg, "The cosmological constant problem," *Rev. Mod. Phys.* **61** (Jan, 1989) 1–23. https://link.aps.org/doi/10.1103/RevModPhys.61.1.

[16] A. M. Turing, "Computing machinery and intelligence," *Mind* **49** (Jan. 01, 1950) 433–460. http://www.cs.umbc.edu/471/papers/turing.pdf.

[17] C. Darwin, *On the Origin of Species by Means of Natural Selection*. Murray, London, 1859. or the Preservation of Favored Races in the Struggle for Life.

[18] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA, USA, 1992.

[19] T. Baeck, D. Fogel, and Z. Michalewicz, *Evolutionary Computation 1: Basic Algorithms and Operators*. Basic algorithms and operators. Taylor & Francis, 2000. https://books.google.de/books?id=4HMYCq9US78C.

[20] R. Trotta, "Bayes in the sky: Bayesian inference and model selection in cosmology," *Contemporary Physics* **49** no. 2, (Mar, 2008) 71–104. http://dx.doi.org/10.1080/00107510802066753.

[21] S. B.-m. Bartelmann, Matthias, *Statistics: The Logic of Science*. Lecture Notes, Heidelberg, 2019. https://heibox.uni-heidelberg.de/f/134be8f876aa4cac8669/.

[22] E. T. Jaynes, *Probability theory: The logic of science*. Cambridge University Press, Cambridge, 2003.

[23] F.-A. Fortin, F.-M. De Rainville, M.-A. Gardner, M. Parizeau, and C. Gagné, "DEAP: Evolutionary algorithms made easy," *Journal of Machine Learning Research* **13** (Jul, 2012) 2171–2175.

[24] PYTHON. https://wiki.python.org/moin/ParallelProcessing.

[25] M. Quade, M. Abel, K. Shafi, R. K. Niven, and B. R. Noack, "Prediction of dynamical systems by symbolic regression," *Physical Review E* **94** no. 1, (Jul, 2016) . http://dx.doi.org/10.1103/PhysRevE.94.012214.

[26] A. Meurer, C. P. Smith, M. Paprocki, O. Čertík, S. B. Kirpichev, M. Rocklin, A. Kumar, S. Ivanov, J. K. Moore, S. Singh, T. Rathnayake, S. Vig, B. E. Granger, R. P. Muller, F. Bonazzi, H. Gupta, S. Vats, F. Johansson, F. Pedregosa, M. J. Curry, A. R. Terrel, v. Roučka, A. Saboo, I. Fernando, S. Kulal, R. Cimrman, and A. Scopatz, "Sympy: symbolic computing in python," *PeerJ Computer Science* **3** (Jan., 2017) e103. https://doi.org/10.7717/peerj-cs.103.

[27] E. Jones, T. Oliphant, P. Peterson, *et al.*, "SciPy: Open source scientific tools for Python," 2001–. http://www.scipy.org/. [Online; accessed].

[28] C. Wetterich, "The cosmon model for an asymptotically vanishing time-dependent cosmological "constant"," 1994.

[29] M. D. Schmidt and H. Lipson, "Age-fitness pareto optimization." in *GECCO*, M. Pelikan and J. Branke, eds., pp. 543–544. ACM, 2010. http://dblp.uni-trier.de/db/conf/gecco/gecco2010.html#SchmidtL10.

Acknowledgements

Ich möchte außerordentlichen Dank für Herrn Prof. Dr. Björn-Malte Schäfer aussprechen. Sein stetiges Bemühen hat diese Bachelorarbeit erst ermöglicht.

Ich möchte darüber hinauß meiner Familie danken, für die emotionale und auch finanzielle Unterstüzung. Besonderes Dank für zahlreichen, hilfreichen und inspirierenden Diskussionen gilt meinen Freunden Julian Urban, Christian Schmidt und Jonas Massa.

Erklärung

Ich versichere, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den 17. März 2020,