Diabatization via Gaussian Process Regression

Dissertation
zur Erlangung des wissenschaftlichen Grades
Doctor rerum naturalium

vorgelegt von

Stefan Benjamin Rabe

Institut für Theoretische Physik
Fakultät Physik
Bereich Mathematik und Naturwissenschaften
Technische Universität Dresden
2024

Eingereicht am 30. Januar 2024

1. Gutachter: Prof. Dr. Jan-Michael Rost
2. Gutachter: Prof. Dr. Roman V. Krems

Verteidigt am 17. Juni 2024

Abstract

Modern supervised machine learning (ML) techniques have taken a prominent role in academia and industry due to their powerful predictive capabilities. While many large-scale ML models utilize deep artificial neural networks (ANNs), which have shown great success if large amounts of data are provided, ML methods employing Gaussian processes (GPs) outperform ANNs in cases with sparse training data due to their interpretability, resilience to overfitting, and provision of reliable uncertainty measures. GPs have already been successfully applied to pattern discovery and extrapolation. The latter can be done in a controlled manner due to their small numbers of interpretable hyperparameters.

In this work we develop an approach based on GPs to extract diabatic patterns from energy spectra, adiabatic under variation of a parameter of the Hamiltonian. The emerging diabatic manifolds (or energy surfaces) exhibit crossings where the original (adiabatic) energy spectra avoid to cross.

In the context of highly excited, classically chaotic dynamics, we demonstrate that our GP regression approach can generate complete diabatic energy spectra with two exemplary systems: two coupled Morse oscillators and hydrogen in a magnetic field. For both we train GPs with few classical trajectories in order to inter- and extrapolate actions throughout the whole energy and parameter range to identify all points where the semiclassical Einstein-Brillouin-Keller (EBK) quantization condition is fulfilled. While the direct EBK method is restricted to regular classical dynamics, the interpretability of the GPs allow for controlled extrapolation into regions where no more regular trajectories exist due to irregular motion. Hence, semiclassical diabatic spectra can be continued into chaotic regions, where such manifolds are no longer well-defined.

Further, we investigate the origin of resonant motion in the coupled Morse oscillator system and their contributions to the semiclassical spectra, which provide energies along strongly repelled adiabatic surfaces. For the hydrogen atom in a magnetic field we show that a proper scaling of the coordinates by the magnetic field strength allows for the extraction of an infinite series of semiclassical energies with one single trajectory which fulfills the EBK condition. The implementation of boundary conditions into GPs, as well as scaling techniques to higher dimensions and their properties are discussed.

Kurzfassung

Moderne Methoden für maschinelles Lernen (ML) spielen heutzutage eine wichtige Rolle in der Wissenschaft und Industrie. Viele umfangreiche ML-Modelle basieren auf tiefen künstlichen neuronalen Netzen (KNN), welche großartige Erfolge erzielen, wenn große Datenmengen zur Verfügung stehen. In Fällen von spärlichen Datenmengen werden KNNe übertroffen von ML-Methoden, welche auf Gaußschen Prozessen (GP) basieren, aufgrund ihrer Interpretierbarkeit, Widerständigkeit gegenüber Überanpassung (Overfitting) und der Bereitstellung von verlässlichen Fehlermaßen. GPe wurden bereits erfolgreich angewandt für Mustererkennung und deren Extrapolation. Letztere ist kontrollierbar aufgrund der kleinen Anzahl von interpretierbaren Hyperparametern.

In der vorliegenden Arbeit entwickeln wir eine Methode basierend auf GPen für die Extraktion von diabatischen Mustern aus Energiespektren, welche sich adiabatisch unter der Variation eines Parameters des Hamiltonoperators verhalten. Die resultierenden diabatischen Mannigfaltigkeiten (oder Energieflächen) weisen Kreuzungen auf, wohingegen die originalen (adiabatischen) Energiespektren Kreuzungen vermeiden.

Im Bezug auf hoch angeregte, klassisch chaotische Dynamik demonstrieren wir, dass unsere Methode vollständige diabatische Spektren generiert anhand von zwei Beispielsystemen: zwei gekoppelte Morse-Oszillatoren und Wasserstoff im Magnetfeld. In beiden Fällen werden GPe trainiert anhand weniger klassischer Trajektorien, um deren Wirkungen zu inter- und extrapolieren über den gesamten Energie- und Parameterraum, und Punkte identifiziert, an denen die semiklassische Einstein-Brillouin-Keller (EBK)-Quantisierungsbedingung erfüllt ist. Obwohl die EBK-Methode auf reguläre klassische Dynamik beschränkt ist, erlaubt die Interpretierbarkeit von GPen eine kontrollierte Extrapolation zu Regionen, in denen keine Regularität mehr vorhanden ist. Dadurch können semiklassische diabatische Spektren ins chaotische Regime fortgesetzt werden, in welchem diese nicht mehr wohldefiniert sind.

Des Weiteren untersuchen wir den Ursprung resonanter Dynamik im System zweier gekoppelter Morse-Oszillatoren und deren Beitrag zu den semiklassischen Spektren, welche Energien entlang stark abgestoßener adiabatischer Flächen liefern. Im Fall von Wasserstoff im Magnetfeld zeigen wir, dass eine geeignete Skalierung der Koordinaten durch die Feldstärke die Generierung einer unendlichen Folge von semiklassischen Energien mit nur einer EBK-quantisierten Trajektorie erlaubt. Die Implementierung von Randbedingungen in GPen, sowie Skaliermethoden für höhere Dimensionen und deren Eigenschaften werden diskutiert.

Contents

1 Introduction

The year 2022 will go down in history as the year when MACHINE LEARNING (ML) tools were finally made utilizable for every last person with free access to the internet. In that year, the pioneering research organization OpenAI released their Chatbot ChatGPT-3.5 as a free version of their gigantic generative AI model pre-trained on a vast corpus of diverse text sources on the internet. Struck by the overwhelming variety of new possibilities, numerous companies still scramble to find ways to apply this new sensation in order to generate money. The novelty about this ML product is the fact that this extremely powerful tool has the mere simple goal of generating informed and coherent output in text form. In order to keep its application as general as possible, no specific task was designated for this ML model to solve and everybody's first attempts at using it revolved around making sense of it.

A similar rush on ML tools was seen in the physical sciences [1] about ten years earlier, when software libraries became available which allowed for easy implementation of ML methods while using the available computing power effectively, two of them being SCIKIT-LEARN and TENSORFLOW, released in 2007 and 2015, respectively. Building powerful predictive algorithms was then no longer reserved for dedicated experts in the field, causing a flood of scientific publications applying ML. Its success was driven by the high demand of new techniques for efficient, large-scale problem solving, as well as the ability to gain insight from data, leading to paradigm shifts in a number of areas. At the same time the broad application of ML models lead to better understanding of some of their internal workings which are difficult to comprehend. This triggered in turn the interest in further analyzing the mathematical foundations of the ML tools in order to improve upon their capabilities.

Evidently, each field of application in academia and industry requires different methods and tools tailored to its specific needs and tasks. Therefore, different ML tools hold contrasting features and handle supplied data differently. However, what all ML methods do have in common is the learning from "experience" in form of data. This manner of gaining insight and making predictions displays an inverse approach to the traditional scientific way of solving problems, i.e., the well-controlled mathematical implementation of physical laws. Instead, the ML philosophy of solving tasks is the extraction of unknown latent structures or relationships using an abundance of data. The underlying mechanics of such ML tools are simple in essence but can reveal complex features. The generality of such data-driven tools encouraged many researchers to apply these methods to their problems at hand, causing a similar hype as we see now with end consumer products like ChatGPT. Although some of these ML models constitute "black-boxes", their capabilities of learning patterns and features from data is astonishing. However, we are still scratching the surface of possibilities and much of the benefit is yet to be uncovered.

One of the countless fields that experiences significant advancements through ML developments is chemical physics. [2] At the base of numerical molecular simulations and the computation of observables under the Born-Oppenheimer approximation lies the representation of potential energy surfaces. The dimensionality of such surfaces grows with the size of the considered molecule and computations of individual points on the surface require expensive solutions of the Schrödinger equation. In order to reduce the computational burden, interpolation methods using ML tools are used to find local approximations for potential energy surfaces using relatively few points as training data. The methods employed for this task are twofold: earlier approaches involved ARTIFICIAL NEURAL NETWORKS (ANNs) [3,4], while later ones made use of GAUSSIAN PROCESS REGRESSION (GPR). [5,6] As both methods have quite distinct properties, this is a showcase for the question about the proper domains of each respective ML method. Typical ANNs are large parametric models, while GPR uses a non-parametric, Bayesian approach. For the particular case of fitting potential energy surfaces, Reference [7] demonstrates that GPR shows superior results over ANNs.

Inspired by those results, the present work employs GPR for the investigation of energy spectra of simple atomic and molecular systems under variation of a parameter. Our main focus lies on the construction of non-adiabatic, or diabatic, energy manifolds, which emerge when avoided crossings in the original (adiabatic) spectra are traversed. Human observation often reveals such diabatic manifolds as patterns in spectra. Hence, we approach this task of pattern extraction using GPR, as they have been identified as powerful tools for pattern discovery and extrapolation. [8,9] In order to generate complete diabatic energy spectra we exploit a semiclassical approach known as the EINSTEIN-BRILLOUIN-KELLER (EBK) quantization method, which allows for the computation of isolated eigenenergies of a system using solely regular trajectories. This method fails to describe quantum effects like tunneling properly, leading to the neglect of energy level repulsion. When combining with regression methods like GPR, complete diabatic energy spectra can be computed by interpolation between few regular trajectories.

In addition, we explore the properties of GPR in order to extrapolate the semiclassical energy spectra into classically irregular regions, where regular trajectories as the backbone for semiclassical quantization no longer exist. It has been brought up in Ref. [2] that Bayesian methods are suited for the extrapolation of physical properties, while ANNs often fall short of delivering adequate results. We demonstrate that GPR is indeed capable of providing semiclassical results in this regime, yielding complete diabatic spectra for systems with mixed classical phase spaces.

The outline of this dissertation is as follows: we begin with an in-depth introduction to GPR in Chapter 2. Its application to the extraction of diabatic manifolds in energy spectra is demonstrated in Chapter 3. A semiclassical way of generating diabatic spectra using the EBK quantization method in combination with GPR is discussed in Chapter 4. Chapters 5 and 6 apply the semiclassical method to two physical systems: two coupled Morse oscillators and the hydrogen atom in a magnetic field. Therein, it is demonstrated how the ML-aided method is capable of extrapolating results into otherwise inaccessible regions due to chaos. Finally, our conclusions are presented in Chapter 7.

2 Gaussian process regression

2.1 Supervised machine learning

The term ARTIFICIAL INTELLIGENCE covers a broad spectrum of tasks that traditionally required human decision making. A more concrete sub-field thereof is termed MACHINE LEARNING (ML) which covers the generation of algorithms that improve their performance in solving specific problems by "learning from experience". This "experience" can either be of the form of labeled data, e.g., classified pictures of cats and dogs, unlabeled data, e.g., shopping behaviors of online customers, or a history of rewards for taking particular actions, e.g., winning games of checkers. Respectively, these three forms of input give rise to the disciplines of SUPERVISED ML, UNSUPERVISED ML, and REINFORCEMENT LEARNING. Evidently, each discipline has its separate field of applications.

Supervised ML can be categorized into REGRESSION and CLASSIFICATION. If each data point can be assigned to one of several discrete labels, or classes, one speaks of classification. On the other hand, if each data point can have a label given by one or more continuous numbers, we speak of regression. The task of supervised ML is to find a generalization of an input-output relationship that necessarily describes the given data the algorithm has been trained on (previously referred to as "experience") and is able to predict labels for new inputs the algorithm hasn't seen before. For the previous example of pictures of cats and dogs, the goal is to correctly distinguish between the two species, or classes, on any given new picture. An example for a regression task may be to predict the age of a person, in form of a real number, from photographs.

Both tasks, classification and regression (the latter is also referred to as CURVE FITTING), have existed long before the term MACHINE LEARNING has first emerged. Early regressions involved the method of LEAST SQUARES [10], where an empirical model function is fitted to the data by varying its parameters to minimize the distances. However, as the complexity or dimensionality of the problems at hand grew, more sophisticated methods were sought after. In the middle of the 20th century ARTIFICIAL NEURAL NETWORKS (ANNs) [11, 12] came about and have taken the lead role in supervised ML by now.

Complementing ANNs, other ML methods have been developed over time that show different qualities and perform better than ANNs in certain applications. One such method is GAUSSIAN PROCESS REGRESSION (GPR) [13], which will be extensively applied within this work. For the purpose of introducing GAUSSIAN PROCESSES (GPs) we will first discuss the structure of ANNs and then show that GPs emerge as a special case when applying the central limit theorem.

Artificial neural networks

ANNs are inspired by the biological neuron structure inside a brain. [11] An ANN consists of layers of artificial neurons which are interconnected, as depicted in Figure 2.1. A real-

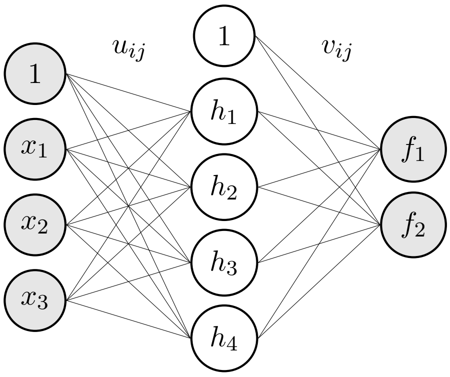

Figure 2.1: A feed-forward MULTILAYER PERCEPTRON (MLP) for a 3-dim. input, a 2-dim. output, and one hidden layer.

valued, D-dimensional input, $\mathbf{x} \in \mathbb{R}^D$, is given as values of D nodes of a first layer, the INPUT LAYER. Each node of the input layer is connected to each node of the second layer. The value of each node of the second layer is then a nonlinear function of the weighted sum of all the values of the first layer plus a bias,

$$h_j(\mathbf{x}) = \alpha\left(u_{0j} + \sum_{i=1}^{D} x_i u_{ij}\right). \tag{2.1}$$

The function $\alpha(\cdot)$ is called an ACTIVATION FUNCTION (or transfer function), the values u_{ij} are the WEIGHTS. A DEEP NEURAL NETWORK consists of more than three layers. The final layer is the OUTPUT LAYER and layers between input and output layers are called HIDDEN LAYERS. The number of nodes of the output layer is given by the dimensionality of the output in case of regression, or the number of classes in case of classification. The values of the output nodes are

$$f_j(\mathbf{x}) = \alpha\left(v_{0,j} + \sum_{i=1}^{N_H} h_i(\mathbf{x}) v_{ij}\right), \tag{2.2}$$

where N_H is the number of nodes of the last hidden layer. For regression, the activation function for the output layer is usually chosen to be linear in order to be unbounded. For classification, an activation function is chosen that is bounded between 0 and 1, such as the sigmoid function, in order to assign a score to each output node, i.e., each class.

Evidently, the chosen numbers of layers and nodes in each hidden layer give rise to the complexity of the input-output relationship such a model can offer. While the design of the ANN, including the choice of the activation functions, are chosen a priori and require refinement, the choice of all the weights connecting the neurons concern the TRAINING, or LEARNING, of the ANN. The latter is done by minimizing a cost function using the BACKPROPAGATION algorithm. [14]

In recent years, ML has been a rapidly evolving field due to the available computational power and has brought miraculous breakthroughs, most of which involve more sophisticated variations of the MULTILAYER PERCEPTRON (MLP), an ANN as depicted in Fig. 2.1 and described above. However, the vast number of parameters, i.e., the weights, together with the choice of the design of an ANN constitute a “black box” and it is so far impossible to fully understand or predict the success of a particular model. The process of designing an ANN is often trial and error as they are prone to OVERFITTING, i.e., the effect of over-explaining seen data at the cost of predictive power, and can converge into an sub-optimal parameter set if, e.g., an inappropriate learning rate is chosen. There is no principled framework for constructing an ANN; its advocates call its designing an "art".

Clearly, while ANNs are without doubt a powerful tool for ML, there are cases where other methods are more appropriate. One of those alternatives are GPs which can be considered as a special branch of ANNs.

From Multilayer Perceptron to Gaussian process

Following Refs. [13,15], a particular design of a ANN shall be investigated, which will coincide with a GP. Let us consider a MLP with a single hidden layer and a single output node. The activation function for the hidden layer shall be bounded. Further, the activation function for the output shall be linear, which yields the simplified expression for Equation (2.2)

$$f(\mathbf{x}) = v_0 + \sum_{i=1}^{N_H} h_i(\mathbf{x}) v_i. \tag{2.3}$$

As the next step, we make a crucial modification to our MLP by promoting all of its weights, which we collectively denote as $\mathbf{w} = (u_{ij}, v_i)$, to be random variables. This transforms our MLP to a BAYESIAN NEURAL NETWORK (BNN). For any choice of probability distribution over all weights, $\mathcal{P}(\mathbf{w})$, Eq. (2.3) yields a probability distribution for the output, $\mathcal{P}(f)$. Let us assume that the bias, v_0, is Gaussian (or normal) distributed, all other v_i $(i = 1, \ldots, N_H)$ be independently and identically distributed with finite variance, and all u_{ij} also be independently and identically distributed[1]. Then, in the limit of infinitely many hidden nodes, $N_H \to \infty$, the central limit theorem applies:

Theorem (Central limit theorem). *Let $\xi_1, \xi_2, \ldots, \xi_N$ be independent and identically distributed random variables with finite mean and variance, μ, σ^2. For $N \to \infty$, the probability distribution of $S_N = \sum_{n=1}^{N} \xi_n$ converges to a Gaussian distribution with mean $N\mu$, and variance $N\sigma^2$.*

Hence, the central limit theorem states that the sum in the second term in Eq. (2.3) converges to a Gaussian distribution. The output is then a sum of two Gaussian distributed random variables, which is again Gaussian[2]; we write $f \sim \mathcal{N}(\mu, \sigma^2)$.

Analogously, the joint distribution of the outputs $\left(f\left(\mathbf{x}^{(1)}\right), \ldots, f\left(\mathbf{x}^{(N)}\right)\right)$ for several input vectors, $\left(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}\right)$, converges towards a multivariate Gaussian distribution, $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with mean vector

$$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)^\top \tag{2.4}$$

and covariances

$$\Sigma_{pq} = \mathrm{cov}\Big(f\left(\mathbf{x}^{(p)}\right), f\left(\mathbf{x}^{(q)}\right)\Big) = E\Big[\Big(f\left(\mathbf{x}^{(p)}\right) - \mu_p\Big)\Big(f\left(\mathbf{x}^{(q)}\right) - \mu_q\Big)\Big], \tag{2.5}$$

---

[1] Due to the bounded activation function for the hidden layer, the variance of each h_i is also finite.

[2] If $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, then $(X + Y) \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

where $E[\cdot]$ is the expected value and the notation $f\big(\mathbf{x}^{(p)}\big) \sim \mathcal{N}\big(\mu_p, \sigma_p^2\big)$ was used. Since the output is a function of a continuous argument, $\mathcal{P}(f(\mathbf{x}))$ produces a GP. In other words, $\mathcal{P}(f(\mathbf{x}))$ is a multivariate Gaussian distribution of infinite dimensions or a Gaussian distribution over functions, $f(\mathbf{x})$.

We have thus seen that a BNN with one infinitely wide hidden layer and a single output node can produce a GP. (Ref. [15] discusses cases with multiple hidden layers of infinite width as well, giving rise to DEEP GAUSSIAN PROCESSES.) Multiple output nodes would produce mutually independent values due to the assumption that the hidden-to-output weights be independent. Therefore, more than one output node yield no additional information.[3]

The transformation from a MLP to a GP has brought up a new set of properties for our ML technique. While a conventional ANN has a large but finite number of parameters/weights that need to be tuned, a GP has an infinite number of parameters and is therefore a NON-PARAMETRIC model. And it may seem surprising, but even though the number of parameters is infinite, GPs are not prone to overfitting. Further, Refs. [16–18] showed that a feed-forward ANN with a single hidden layer can uniformly approximate any continuous function, given the number of hidden units is sufficient, making single-layer ANNs universal approximators.

Moreover, since we consider Gaussian probabilities for the output, $\mathcal{P}(f)$, a GP naturally provides an UNCERTAINTY MEASURE with its prediction. This allows a whole new spectrum of applications and is one of, if not *the* most useful property of GPs. Being a non-parametric model, GPs also do not require adjustment of any weights. Instead, the model is fitted to a training data set using Bayesian inference. The choice of the activation function for the hidden layer corresponds to the choice of covariance function of the GP, which will be discussed in the next section. Lastly, there is no longer any question about the choice of design since we considered one hidden layer of infinite width. One drawback so far is the limitation to one single output node, which corresponds to scalar functions in the regression case. However, multi-output (or multi-task) GPs are discussed in the literature. [19]

2.2 From multivariate Gaussian to Gaussian process

Of all the probability distributions one can think of, the Gaussian (or normal) distribution is arguably one of the simplest and has the most practical properties. One of those properties is the unique characterization through the mean (or expected) value and the variance – the first and second cumulant, respectively. All higher cumulants vanish.

The importance of the Gaussian distribution results from the central limit theorem, which states that the sum of a large number of independent random variables is again a random variable that tends to be Gaussian distributed.[4] Thus, in situations where a large number of stochastic factors have an additive effect on a measurable quantity it is often suitable to assume that this quantity is sampled from a Gaussian distribution.

Due to its simplicity and stochastic importance it is desirable to use Gaussian distributions in order to make probabilistic predictions about quantities where little is known about their characteristics. In particular, Gaussian distributions may turn out useful in order to conduct

---

[3] Although one could consider correlations between the mean and variance of the weights for one output node and those of another. This would lead to characteristics of multi-output GPs.

[4] The Lyapunov variant of the central limit theorem generalizes to independent random variables, without the condition of being identically distributed.

supervised ML in a probabilistic manner, using Bayesian inference.

A (univariate) Gaussian probability distribution is uniquely defined by a mean value, μ, and a variance, σ^2, and we write for a Gaussian distributed random variable

$$f \sim \mathcal{N}\big(\mu, \sigma^2\big) := \frac{1}{\sqrt{2\pi}\sigma} \exp\bigg(-\frac{1}{2}\frac{(f-\mu)^2}{\sigma^2}\bigg) = \mathcal{P}(f), \tag{2.6}$$

where $\mathcal{P}$ denotes the PROBABILITY DENSITY FUNCTION (PDF).

The generalization to a vector-valued random variable (or random vector), $\mathbf{f} \in \mathbb{R}^N$, yields a joint probability distribution that matches a multivariate Gaussian distribution, which is uniquely characterized by a mean vector, $\boldsymbol{\mu} \in \mathbb{R}^N$, and a covariance matrix, $\boldsymbol{\Sigma} \in \mathbb{R}^{N\times N}$. Analogously to the univariate case we write

$$\mathbf{f} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) := \frac{1}{\sqrt{(2\pi)^N \det(\boldsymbol{\Sigma})}} \exp\bigg(-\frac{1}{2}(\mathbf{f}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{f}-\boldsymbol{\mu})\bigg) = \mathcal{P}(\mathbf{f}). \tag{2.7}$$

In the limit of $N \to \infty$, i.e., infinitely many components of a random vector, we can think of each component as values of a continuous function, $f_j \xrightarrow{N\to\infty} f(\mathbf{x})$. This means, $\mathcal{P}(f(\mathbf{x}))$ becomes a Gaussian distribution over functions, more precisely a GP, which is uniquely characterized by a mean function, $\mu(\mathbf{x})$, and a covariance function, $\Sigma(\mathbf{x}, \mathbf{x}')$,

$$f(\mathbf{x}) \sim \mathcal{GP}\big(\mu(\mathbf{x}), \Sigma\big(\mathbf{x}, \mathbf{x}'\big)\big). \tag{2.8}$$

Note that both mean and covariance function are scalar, while the arguments need not be scalar.

Stochastic process

Loosely speaking, if our random variable of interest is a continuous function, we speak of a STOCHASTIC PROCESS (or random process). Historically, this continuous function was often a function of time. If the evolution of a physical process is governed by an underlying random or uncertain behavior, each state at a given point in time could be considered a random variable. A discrete and finite time sequence therefore yields a random vector with a joint probability distribution. If this joint distribution is a multivariate Gaussian, our stochastic process is a GP. As mentioned before, a stochastic process needs not be a function of only one variable (such as time), but it can be a function of multiple variables.

Definition (Gaussian process (GP)). *A GP is a collection of random variables, any finite number of which have a joint Gaussian distribution. In other words: $\mathcal{P}(f(\mathbf{x}))$ is a GP if the marginal distribution, $\mathcal{P}\big(f\big(\mathbf{x}^{(1)}\big), \ldots, f\big(\mathbf{x}^{(N)}\big)\big) = \int \mathcal{P}(f(\mathbf{x}))\,\mathrm{d}(f(\mathbf{Y}))$, for any finite subset, $\big\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}\big\} \subset \mathcal{X}$, is a multivariate Gaussian. Here, $\mathcal{X} \subseteq \mathbb{R}^D$ is the domain of $\mathbf{x}$ and $\mathbf{Y} = \mathcal{X} \setminus \big\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}\big\}$.*

Note that this definition includes finite collections of random variables as well, i.e., multivariate Gaussians, but these are for obvious reasons not the cases of our interest.

Marginalization of Gaussians

In a practical supervised ML application, one would represent a continuous function in a discretized fashion on a computer. Hence, if a function needs to be represented as values of a finite subset, $\{f(\mathbf{x}^{(1)}), \ldots, f(\mathbf{x}^{(N)})\}$, it is necessary to work with the marginalized distribution, $\mathcal{P}(f(\mathbf{x}^{(1)}), \ldots, f(\mathbf{x}^{(N)})) = \int \mathcal{P}(f(\mathbf{x}))\,\mathrm{d}(f(\mathbf{Y}))$. This integral, however, seems intimidating, as the integration ought to be performed over the probability distributions of an infinite number of random variables, $f(\mathbf{Y})$. To our rescue comes the important marginalization property of Gaussian distributions:

Theorem (Marginal distribution of Gaussians)**.** *For a Gaussian random vector, partitioned as* $(\mathbf{f}_1, \mathbf{f}_2)^\top$*, with correspondingly partitioned mean and (symmetric) covariance,*

$$\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)^\top, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}{}^\top & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \tag{2.9}$$

the marginal distribution of $\mathbf{f}_1$ *is another multivariate Gaussian with* $\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$*. For an unordered subset,* $\mathbf{f}_1$*, the components of* $\boldsymbol{\mu}_1$ *are just the components of* $\boldsymbol{\mu}$ *for the subset and* $\boldsymbol{\Sigma}_{11}$ *are the corresponding rows and columns of* $\boldsymbol{\Sigma}$*.*

Evidently, marginalizing Gaussians is trivial. For a given GP, $\mathcal{GP}(\mu(\mathbf{x}), \Sigma(\mathbf{x}, \mathbf{x}'))$, the marginalized (discretized) probability distribution at $\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}\}$ is simply $\mathcal{N}(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s)$ with

$$\boldsymbol{\mu}_s = \begin{pmatrix} \mu(\mathbf{x}^{(1)}) \\ \vdots \\ \mu(\mathbf{x}^{(N)}) \end{pmatrix}, \quad \boldsymbol{\Sigma}_s = \begin{pmatrix} \Sigma\left(\mathbf{x}^{(1)}, \mathbf{x}^{(1)}\right) & \ldots & \Sigma\left(\mathbf{x}^{(1)}, \mathbf{x}^{(N)}\right) \\ \vdots & \ddots & \vdots \\ \Sigma\left(\mathbf{x}^{(N)}, \mathbf{x}^{(1)}\right) & \ldots & \Sigma\left(\mathbf{x}^{(N)}, \mathbf{x}^{(N)}\right) \end{pmatrix}, \tag{2.10}$$

which is in accordance with above definition of a GP. So instead of sampling a function from a true GP, $f(\mathbf{x}) \sim \mathcal{GP}(\mu(\mathbf{x}), \Sigma(\mathbf{x}, \mathbf{x}'))$, in practice we sample a vector from a multivariate Gaussian, $\mathbf{f} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, giving us all function values at the locations of our interest.

2.3 Mean and covariance functions

Just like a multivariate Gaussian distribution is uniquely characterized by its mean vector and covariance matrix, a GP is fully specified by its mean, $\mu(\mathbf{x})$, and covariance function (or KERNEL), $\Sigma(\mathbf{x}, \mathbf{x}')$. Each sample function, $f(\mathbf{x})$, drawn from a GP will be Gaussian distributed around the mean function. One writes

$$f(\mathbf{x}) \sim \mathcal{GP}\left(\mu(\mathbf{x}), \Sigma(\mathbf{x}, \mathbf{x}')\right). \tag{2.11}$$

Fig. 2.2 shows various sample functions drawn from a GP with a particular choice of mean and covariance function. The mean function “guides” all sample functions by forcing them to be distributed nearby. The covariance function, similar to a covariance matrix, measures the correlation between two function values, $f(\mathbf{x})$ and $f(\mathbf{x}')$. If the covariance, $\Sigma(\mathbf{x}, \mathbf{x}')$, is large, the two corresponding function values will be forced to be very similar. If it is zero, the two values will be uncorrelated and, since they are jointly Gaussian, independent. If the covariance is negative, large values at $\mathbf{x}$ will correspond to small values at $\mathbf{x}'$. The

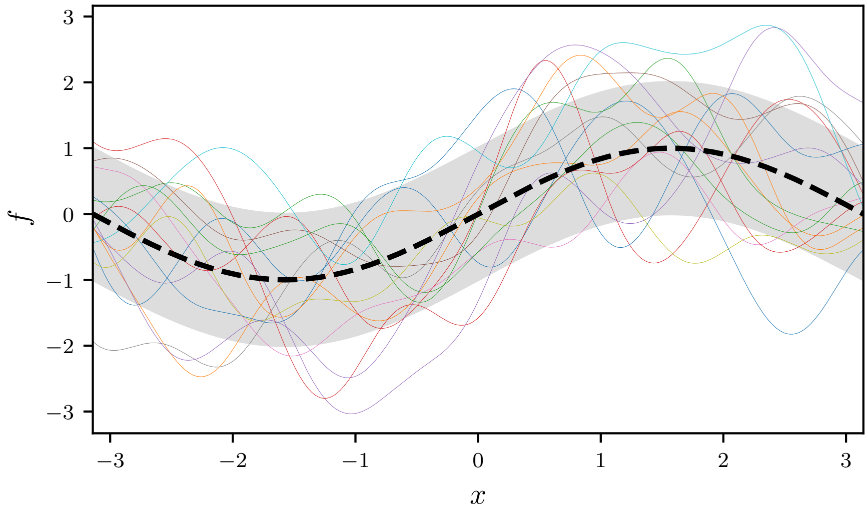

Figure 2.2: Sample functions (colored) drawn from a GP with sinusoidal mean (dashed) and RADIAL BASIS FUNCTION (RBF) covariance function with $\sigma_f = 1$ and $l = 0.5$. The range within the standard deviation is shown as gray shade.

diagonal of the covariance function, $\Sigma(\mathbf{x}, \mathbf{x}) = \sigma^2(\mathbf{x})$, yields the variance of $f(\mathbf{x})$. A valid covariance function is symmetric and positive definite. The first property directly follows from the definition of a covariance, Eq. (2.5), while the second property ensures positive variance everywhere, $\sigma^2(\mathbf{x}) > 0$.

In the following we will discuss one particular covariance function very commonly used for regression purposes, namely the RADIAL BASIS FUNCTION (RBF) covariance function,

$$\Sigma_{\mathrm{RBF}}(\mathbf{x}, \mathbf{x}') = {\sigma_f}^2 \exp\left(-\frac{1}{2}\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{l^2}\right). \tag{2.12}$$

This function is sometimes also referred to as SQUARED EXPONENTIAL or Gaussian covariance function. It is a stationary function, as it only depends on the (Euclidean) distance between the input vectors, $\|\mathbf{x}-\mathbf{x}'\|$. The two values σ_f and l shall be called OUTPUT-SCALE (also referred to as SIGNAL VARIANCE) and LENGTH-SCALE, respectively. They fall under the category of HYPERPARAMETERS, which will be discussed in Sec. 2.4.2, and differ significantly from regular parameters, e.g., the weights in ANNs. The output-scale measures the spread of the sample functions along the output dimension and directly yields the standard deviation, σ, if $\mathbf{x} = \mathbf{x}'$. The length-scale measures the correlation length in the input space. Large length-scales lead to a stronger correlation between two points, causing any sample function to vary slowly. Shorter length-scales cause sampling functions to be more wiggly, as two points are less correlated. As the distance between two points grows, $\|\mathbf{x}-\mathbf{x}'\| \to \infty$, their covariance becomes smaller, $\Sigma_{\mathrm{RBF}}(\mathbf{x}, \mathbf{x}') \to 0$, and their corresponding output values, $f(\mathbf{x})$ and $f(\mathbf{x}')$, tend to be independent. In Fig. 2.2 a RBF covariance function with an output-scale of one and a length-scale of $l = 0.5$ is used. For higher input dimensions, $D > 1$, one can introduce one length-scale for each direction by replacing $\|\mathbf{x}-\mathbf{x}'\|^2/l^2 \to (\mathbf{x}-\mathbf{x}')^\top \mathbf{M}^{-1}(\mathbf{x}-\mathbf{x}')$ with the diagonal matrix $\mathbf{M} = \mathrm{diag}(l_1^2, l_2^2, \dots, l_D^2)$.

If a GP is characterized by a smooth mean and covariance function, its sample functions will be smooth as well. If either function shows any kinks, the sample functions will inherit such kinks at the same positions. The RBF covariance function is infinitely often differentiable, and therefore smooth. It is also stationary, meaning that its local properties do not change when varying the input region.

Other stationary covariance functions frequently used in the literature shall be given at this point. A generalization of the RBF covariance function is given by the RATIONAL QUADRATIC (RQ) function,

$$\Sigma_{\mathrm{RQ}}(\mathbf{x}, \mathbf{x}') = {\sigma_f}^2 \left(1 + \frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\alpha l^2}\right)^{-\alpha}, \tag{2.13}$$

where an additional length-scale appears: $\alpha > 0$. Ref. [13] describes the RQ covariance function as an infinite sum of RBF covariance functions with different length-scales, l. In the limit of $\alpha \to \infty$, the RQ covariance function converges to an RBF covariance function. Like the RBF, the RQ covariance function is infinitely differentiable.

One class of covariance functions of finite differentiability is the class of MATÉRN covariance functions. Ordered by differentiability, the first three simplified members are

$$\Sigma^{(0)}_{\mathrm{Mat}}(\mathbf{x}, \mathbf{x}') = {\sigma_f}^2 \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|}{l}\right), \tag{2.14}$$

$$\Sigma^{(1)}_{\mathrm{Mat}}(\mathbf{x}, \mathbf{x}') = {\sigma_f}^2 \left(1 + \frac{\sqrt{3}\|\mathbf{x} - \mathbf{x}'\|}{l}\right) \exp\left(-\frac{\sqrt{3}\|\mathbf{x} - \mathbf{x}'\|}{l}\right), \tag{2.15}$$

$$\Sigma^{(2)}_{\mathrm{Mat}}(\mathbf{x}, \mathbf{x}') = {\sigma_f}^2 \left(1 + \frac{\sqrt{5}\|\mathbf{x} - \mathbf{x}'\|}{l} + \frac{5\|\mathbf{x} - \mathbf{x}'\|^2}{3l^2}\right) \exp\left(-\frac{\sqrt{5}\|\mathbf{x} - \mathbf{x}'\|}{l}\right). \tag{2.16}$$

The first covariance function is not differentiable; GPs corresponding to the second and third are respectively once and twice mean square differentiable. Many more covariance functions exist and new exemplars can be constructed, e.g., by adding and multiplying different valid covariance functions together. [9, 13]

Two particular covariance functions have become prominent in modern GP applications and merit special recognition: the SPECTRAL MIXTURE (SM) [8, 9] and the DEEP KERNEL LEARNING (DKL) [20] covariance functions. The SM covariance is particularly powerful in discovering and extrapolating periodic patterns occurring in data. It is constructed by superposing multiple periodic covariance functions in order to match and extrapolate data; its hyperparameters represent the underlying periods and their corresponding scales.

DKL is the term for the combination of a deep ANN and a GP as the output layer. This method inherits the expressive power of a deep ANN and the flexibility and meaningful predictive uncertainty of a GP. Additionally, DKL allows for high scalability, i.e., learning from data sets including up to millions of training points, which will be discussed in Sec. 2.5.

Both methods for constructing covariance functions are complementary. However, due to the potentially large number of hyperparameters, such covariance functions need to be constructed carefully.

2.3.1 Incorporating non-stationary features

All of the examples in the previous section were stationary covariance functions. But one might request non-stationary behavior in form of boundary conditions or cusps, i.e., positions of non-differentiability. Incorporating such properties requires a different mean and/or covariance function. Generally, one finds a suiting GP by choosing a mean function which satisfies the desired conditions and modifying a stationary covariance in order to include such features.

Example: Dirichlet boundary conditions

As a demonstration, let us start with a GP, $\mathcal{GP}(\mu, \Sigma)$, where the sample functions do not fulfill any boundary condition, such as the one depicted in Fig. 2.2, where a sinusoidal mean and a stationary RBF covariance function is used. The goal shall be to implement Dirichlet boundary conditions at the edges of a Cartesian box: $f(\mathbf{x}) = 0$ at $|x_d| = L_d$ $(d = 1, \ldots, D)$. In order for both, mean and covariance function, to vanish at the edges, one may choose to employ a linear function which contains above boundary conditions,

$$B(\mathbf{x}) = \prod_{d=1}^{D} |x_d - L_d||x_d + L_d|. \tag{2.17}$$

As discussed in Ref. [21], a GP with the implemented boundary conditions is then given as $\mathcal{GP}\big(\mu^{\text{Box}}, \Sigma^{\text{Box}}\big)$ with

$$\mu^{\text{Box}} = B(\mathbf{x})\mu(\mathbf{x}), \tag{2.18}$$

$$\Sigma^{\text{Box}} = B(\mathbf{x})\Sigma\big(\mathbf{x}, \mathbf{x}'\big)B(\mathbf{x}'). \tag{2.19}$$

A modified version of the GP from Fig. 2.2 with incorporated Dirichlet boundary conditions at $|x| = 2$ is shown in Fig. 2.3, left. As requested, the mean of the transformed GP satisfies the boundary condition. But more importantly, the uncertainty of the GP approaches zero towards the edges of the box. This behavior is necessary because every sample function, $f(\mathbf{x}) \sim \mathcal{GP}\big(\mu^{\text{Box}}, \Sigma^{\text{Box}}\big)$, shall satisfy the boundary condition. More general boundary shapes are discussed in Ref. [22].

Example: a single electron in a molecule

An electron bound around a molecular ion experiences an attractive potential due to the Coulomb force. Eigenstates of the electron, $\psi(\mathbf{x})$, given as solutions of the TIME-INDEPENDENT SCHRÖDINGER EQUATION, have kinks at the positions of the nuclei, $\mathbf{R}_i$, where the Coulomb potential is singular, and are smooth elsewhere. At large distances, they decay away from the molecular ion like a hydrogenic eigenstate, $\psi(\mathbf{x}) \sim \exp\big(-\frac{Z}{n}\|\mathbf{x} - \mathbf{R}\|\big)$, where Z is the number of protons inside the molecule, centered at $\mathbf{R}$, and n is the principal quantum number of the hydrogenic state.

Such properties can be implemented into a stationary GP by transforming the mean and

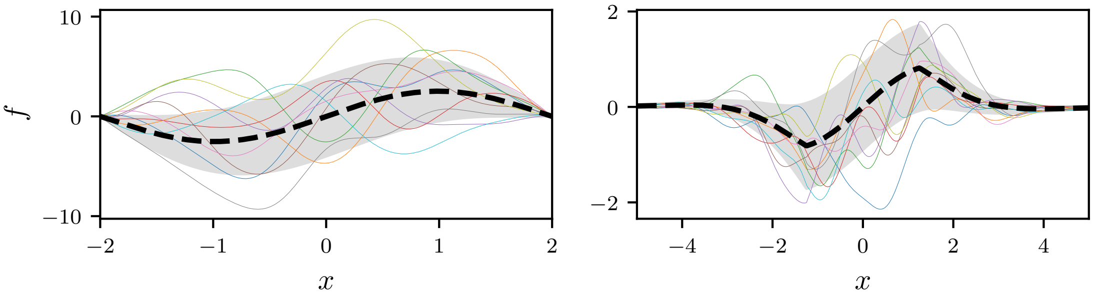

Figure 2.3: Transformations of the GP in Fig. 2.2: **left**, with Dirichlet boundary conditions at $|x| = 2$ using Equations (2.17) – (2.19); **right**, with molecular features from Eqs. (2.20) – (2.22), using $\gamma_1 = \gamma_2 = 1$, $\eta = 2$, and $\mathbf{R}_{1,2} = \pm 1.25$.

covariance function using

$$B(\mathbf{x}) = \prod_i \exp\left(-\frac{\gamma_i}{\eta}\|\mathbf{x} - \mathbf{R}_i\|\right) \tag{2.20}$$

to yield, similar to Eqs. (2.18) and (2.19),

$$\mu^{\text{Mol}} = B(\mathbf{x})\mu(\mathbf{x}), \tag{2.21}$$

$$\Sigma^{\text{Mol}} = B(\mathbf{x})\Sigma(\mathbf{x}, \mathbf{x}')B(\mathbf{x}'). \tag{2.22}$$

These modifications of the mean and covariance functions exhibit kinks at the positions of the nuclei, $\mathbf{x} = \mathbf{R}_i$. Regarding the standard deviation of the corresponding GP far away from the molecular ion,

$$\begin{aligned} \sigma^{\text{Mol}}(\mathbf{x}) &\equiv \sqrt{\Sigma^{\text{Mol}}(\mathbf{x}, \mathbf{x})} = \sigma_f B(\mathbf{x}) && (2.23) \\ &= \sigma_f \prod_i \exp\left(-\frac{\gamma_i}{\eta}\|\mathbf{x} - \mathbf{R}_i\|\right) = \sigma_f \exp\left(-\frac{1}{\eta}\sum_i \gamma_i\|\mathbf{x} - \mathbf{R}_i\|\right) && (2.24) \\ &\xrightarrow{\|\mathbf{x}\| \to \infty} \sigma_f \exp\left(-\frac{\sum_i \gamma_i}{\eta}\|\mathbf{x} - \mathbf{R}\|\right), && (2.25) \end{aligned}$$

we find that it decays, as desired, like an electronic bound state. Evidently, the values of γ_i represent the charges of the nuclei and η is a hyperparameter relating to the principal quantum number. Arbitrarily, the output-scale, σ_f, could be chosen such that it normalizes the area of the standard deviation, i.e., $\int \left(\sigma^{\text{Mol}}\right)^2 \mathrm{d}\mathbf{x} = 1$. However, each sample function from the GP needs to be normalized individually.

In Fig. 2.3, right, the H_2^+ molecular ion in one dimension is considered: one electron around two protons with an intermediate distance of 2.5 ATOMIC UNITS (a.u.). Sample functions from the modified GP are shown, $f(\mathbf{x}) \sim \mathcal{GP}\left(\mu^{\text{Mol}}, \Sigma^{\text{Mol}}\right)$, for $\gamma_1 = \gamma_2 = 1$, $\eta = 2$, normalized σ_y and $l = 0.5$. Each sample function exponentially decays for $\|\mathbf{x}\| \to \infty$. They all possess kinks at the locations of the nuclei and are smooth elsewhere.

The non-stationary features that have been treated in this section were simple, pedagogical cases that could be handled analytically by mindful considerations. However, many other cases may be more complicated and would require more involved procedures for implementation. Such cases are discussed in, e.g., Refs. [21, 22].

2.4 Bayesian inference

In the previous section we have discussed GPs by exploring different mean and covariance functions and drawing samples from the corresponding distributions. We have seen that one can adjust a GP by incorporating features one expects due to empirical observations, such as boundary behavior. So far, the effect of each hyperparameter appearing in the covariance functions could be interpreted by observing different sample functions drawn from the GP. Hence, we are able to construct a whole distribution over functions that exhibit a desired behavior, before having seen any training data.

In this section, we tend to our original task of supervised ML, i.e., training a model on given data in order to make predictions. For an ANN, this process involves optimizing the many parameters, i.e., the weights, using the backpropagation algorithm in order to fit the model function to the data. But as we have noticed, no such weights appear in a GP model. The few variables that are left, the hyperparameters, do not determine a particular functional form, but still yield a whole distribution of infinitely many functions. Instead, inference is done by conditioning a PRIOR probabilistic model on the training data. The conditional distribution, or POSTERIOR, then provides a new distribution that carries the information of the observed data, as well as the features of the prior.

In the philosophy of probabilities the employment of a prior belief falls into the school of BAYESIAN (or evidential) probability. In the Bayesian interpretation probability expresses an empirical or subjectively plausible belief about an event to occur. A prior probability may include knowledge about an event gained from physical insight or previous experiments. After a number of trials/measurements the prior belief is then “corrected” by applying BAYES' THEOREM, yielding the posterior probability. This interpretation of probability opposes the FREQUENTIST school, in which probability is defined as the physical limit of the relative frequency of an event after a large number of trials.

Bayesian statistics is a practical tool for supervised ML since its goal is the inference of a predictive probabilistic model from limited available data. The employment of a prior probability, however, is a blessing and curse at the same time. While the incorporation of prior knowledge, e.g., smoothness or stationarity, is invaluable, an unreasonable choice could lead to futile results. To one's benefit comes the fact that in the limit of large data sets the role of a stationary prior becomes negligible.

In what follows, we will discuss GPs for regression purposes only. For classification, we refer the reader to Ref. [13].

Training data

Let us assume a set of N observations is given, referred to as the TRAINING DATA SET, in form of tuples of an input vector, $\mathbf{x}_j \in \mathbb{R}^D$, and a corresponding scalar output, y_j: $\{(\mathbf{x}_j, y_j); j = 1, \ldots, N\}$. Concatenating all training input vectors into columns of a matrix, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N) \in \mathbb{R}^{D \times N}$, and all training outputs into a vector, $\mathbf{y} = (y_1, \ldots, y_N)^\top$, we can write $\{\mathbf{X}, \mathbf{y}\}$ for the training set. $\mathbf{X}$ is referred to as the DESIGN MATRIX. The observed output values shall be noisy evaluations of the true input-output relationship,

$$y_j = f(\mathbf{x}_j) + \epsilon, \tag{2.26}$$

with additive independent Gaussian noise with zero mean, $\epsilon \sim \mathcal{N}\left(0, {\sigma_n}^2\right)$. We call the true input-output relationship, $f(\mathbf{x})$, the LATENT FUNCTION. A prior GP is defined by choosing a mean and covariance function that resembles our prior belief about $f(\mathbf{x})$, i.e., smoothness, boundary conditions, etc.

In order to do inference, the prior model must be conditioned on the training data. This step can be interpreted as restricting our prior distribution over functions to only those functions which go through the training points under consideration of the underlying noise, ϵ.

2.4.1 Conditioning on training data

For a random vector partitioned into two subvectors, $(\mathbf{w}, \mathbf{y})^\top$, the conditional distribution of $\mathbf{w}$, conditioned on the event of a fixed $\mathbf{y}$, is defined as

$$\mathcal{P}(\mathbf{w}|\mathbf{y}) := \frac{\mathcal{P}(\mathbf{w}, \mathbf{y})}{\mathcal{P}(\mathbf{y})}. \tag{2.27}$$

Here, $\mathcal{P}(\mathbf{w}, \mathbf{y})$ is the joint probability and the marginal distribution,

$$\mathcal{P}(\mathbf{y}) = \int \mathcal{P}(\mathbf{w}, \mathbf{y}) \, \mathrm{d}\mathbf{w} > 0, \tag{2.28}$$

is positive because we do not condition on impossible events. If the conditional distribution equals the marginal distribution, $\mathcal{P}(\mathbf{w}|\mathbf{y}) = \mathcal{P}(\mathbf{w})$, then $\mathbf{w}$ and $\mathbf{y}$ are independent. Using the definition of the conditional distribution and substituting $\mathcal{P}(\mathbf{w}, \mathbf{y}) = \mathcal{P}(\mathbf{y}|\mathbf{w})\mathcal{P}(\mathbf{w})$ one obtains BAYES' THEOREM:

$$\mathcal{P}(\mathbf{w}|\mathbf{y}) = \frac{\mathcal{P}(\mathbf{y}|\mathbf{w}) \, \mathcal{P}(\mathbf{w})}{\mathcal{P}(\mathbf{y})}. \tag{2.29}$$

We denote $\mathcal{P}(\mathbf{w})$ as the prior distribution and $\mathcal{P}(\mathbf{y}|\mathbf{w})$ is referred to as the LIKELIHOOD of $\mathbf{w}$ given $\mathbf{y}$, i.e., $L(\mathbf{w}; \mathbf{y}) = \mathcal{P}(\mathbf{y}|\mathbf{w})$. $\mathcal{P}(\mathbf{w}|\mathbf{y})$ is the posterior distribution – the prior conditioned on some fixed $\mathbf{y}$.

Starting from a probabilistic supervised ML model with an initial/prior distribution over the parameters one can use Bayes' theorem to infer from training data. This yields the posterior distribution over the parameters, which corresponds to the optimized weights in a non-probabilistic ANN. In the case of a BNN the prior and posterior are the probability distributions of the weights before and after conditioning on the training data, respectively. (This is suggested by the nomenclature used above.) Predictions are made by marginalizing over all weights using the posterior distribution,

$$\mathcal{P}(f(\mathbf{x})|\mathbf{y}) = \int \mathcal{P}(f(\mathbf{x})|\mathbf{w}) \, \mathcal{P}(\mathbf{w}|\mathbf{y}) \, \mathrm{d}\mathbf{w}. \tag{2.30}$$

In Sec. 2.1 we have seen that a GP corresponds to a BNN with at least one infinitely wide hidden layer. So instead of considering a prior distribution for infinitely many weights, we directly consider a prior GP with a mean and covariance function, $\mathcal{P}(f(\mathbf{x})) = \mathcal{GP}(\mu(\mathbf{x}), \Sigma(\mathbf{x}, \mathbf{x}'))$.

Bayes' rule then writes for the posterior

$$\mathcal{P}(f(\mathbf{x})|\mathbf{y}) = \frac{\mathcal{P}(\mathbf{y}|f(\mathbf{x}))\,\mathcal{P}(f(\mathbf{x}))}{\mathcal{P}(\mathbf{y})}. \tag{2.31}$$

We see that the posterior is already our predictive distribution over functions. No marginalization over weights is necessary. On top of that, another useful property of Gaussian distributions comes to aid:

Theorem (Conditional distribution of Gaussians)**.** *For a Gaussian random vector, partitioned as* $(\mathbf{f}_1, \mathbf{f}_2)^\top$*, with correspondingly partitioned mean and (symmetric) covariance,*

$$\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)^\top, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}{}^\top & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \tag{2.32}$$

the conditional distribution of $\mathbf{f}_1$*, conditioned on a fixed* $\mathbf{f}_2 = \mathbf{y}$*, is another multivariate Gaussian,* $\mathcal{P}(\mathbf{f}_1|\mathbf{f}_2 = \mathbf{y}) = \mathcal{N}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$*, with*

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{y} - \boldsymbol{\mu}_2), && (2.33) \\ \hat{\boldsymbol{\Sigma}} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}{}^\top. && (2.34) \end{aligned}$$

Comparing this result to the marginal distribution of Gaussians, i.e. $\mathcal{P}(\mathbf{f}_1) = \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$, one sees that knowledge of observed training data, $\mathbf{y}$, alters the distribution of $\mathbf{f}_1$. This is of course what we expect from doing inference. It is straightforward to generalize Eqs. (2.33) and (2.34) for GPs by simply substituting the prior mean vector and covariance matrix by continuous mean and covariance functions,

$$\begin{aligned} \hat{\mu}(\mathbf{x}) &= \mu(\mathbf{x}) + \boldsymbol{\Sigma}(\mathbf{x};\mathbf{X})^\top\boldsymbol{\Sigma}(\mathbf{X},\mathbf{X})^{-1}\big(\mathbf{y} - \boldsymbol{\mu}(\mathbf{X})\big), && (2.35) \\ \hat{\Sigma}(\mathbf{x},\mathbf{x}') &= \Sigma(\mathbf{x},\mathbf{x}') - \boldsymbol{\Sigma}(\mathbf{x};\mathbf{X})^\top\boldsymbol{\Sigma}(\mathbf{X},\mathbf{X})^{-1}\boldsymbol{\Sigma}(\mathbf{x}';\mathbf{X}). && (2.36) \end{aligned}$$

Here, $\boldsymbol{\Sigma}(\mathbf{x};\mathbf{X})$ is the vector of functions $\big(\Sigma(\mathbf{x};\mathbf{x}_1), \ldots, \Sigma(\mathbf{x};\mathbf{x}_N)\big)^\top$, $\boldsymbol{\Sigma}(\mathbf{X},\mathbf{X})$ is the matrix with entries $\Sigma_{ij} = \Sigma(\mathbf{x}_i, \mathbf{x}_j) + \sigma_n{}^2\delta_{ij}$ with the diagonal variance following from the noise assumption in Eq. (2.26), and $\boldsymbol{\mu}(\mathbf{X})$ is the vector $\big(\mu(\mathbf{x}_1), \ldots, \mu(\mathbf{x}_N)\big)^\top$. Our predictive posterior GP is then $\mathcal{GP}\big(\hat{\mu}(\mathbf{x}), \hat{\Sigma}(\mathbf{x},\mathbf{x}')\big)$. We have thus seen that doing inference using GPs is analytically tractable; we speak of GPR.

For the same reason why we marginalize in practical ML applications on a computer, we also compute only finite values of our predictive functions. Therefore, although considering continuous GPs, we only evaluate the conditional distribution, i.e., the posterior, using finite, multivariate Gaussians by applying Eqs. (2.33) and (2.34).

The mayor drawback of Eqs. (2.33) – (2.36) is the necessity of computing the inverse of a square covariance matrix of the size of the number of training points, N. Matrix inversion is a computationally expensive task and scales in general cubically with the size of the matrix, $\mathcal{O}(N^3)$. This matrix inversion states *the* major bottleneck of using GPs for supervised ML tasks. Hence, in order to work with large training data sets, ancillary methods are necessary to improve upon the scaling of ordinary matrix inversion. Such methods will be discussed in Sec. 2.5.

2.4.2 Gaussian process regression

Using Eqs. (2.33) – (2.36), we can finally make predictions about a latent function describing an input-output relationship after the observation of some training data serving as "experience" for the ML model. Fig. 2.4, left, shows a posterior GP using the prior in Fig. 2.3, left, conditioned on three training points. The non-stationary features from the prior, namely the boundary conditions, are inherited in the posterior. However, the posterior mean function carries almost no resemblance to the prior mean, except for the smoothness and the zero boundaries. In fact, after conditioning on the training data, the mean has taken an opposite trend. This behavior is indeed desirable, meaning that the detailed choice for the prior mean is almost arbitrary, which makes inference robust. Following this observation, in most practical cases the prior mean function can be set to be identically zero, $\mu(\mathbf{x}) = 0$. Not only does this choice slightly simplify Eqs. (2.33) and (2.35), but often a mean function which lies far away from the training data prevents underfitting, on which will be elaborated in Sec. 2.4.3. Following this insight we will solely consider zero mean priors from now on.

The posterior GP provides again a mean and covariance function. For regression tasks one would use the posterior mean as the regression function, as it constitutes the values of the highest probability density. The diagonal elements of the posterior covariance provide the variance, or uncertainty, at each point, $\hat{\sigma}^2(\mathbf{x}) = \hat{\Sigma}(\mathbf{x}, \mathbf{x})$. Regarding the equation for the posterior mean more closely, Eq. (2.35), one notices that it can be rewritten as (using a zero prior mean)

$$\hat{\mu}(\mathbf{x}) = \mathbf{\Sigma}(\mathbf{x}; \mathbf{X})^\top \underbrace{\mathbf{\Sigma}(\mathbf{X}, \mathbf{X})^{-1}\mathbf{y}}_{:=\mathbf{c}} \tag{2.37}$$

$$= \sum_{j=1}^{N} c_j \Sigma(\mathbf{x}; \mathbf{x}_j), \tag{2.38}$$

which is simply an expansion using N basis functions, $\Sigma(\mathbf{x}; \mathbf{x}_j)$, where N is the number of training points. If, for instance, we choose the RBF covariance function for the prior, above expression would yield

$$\hat{\mu}(\mathbf{x}) = {\sigma_f}^2 \sum_{j=1}^{N} c_j \exp\left(-\frac{1}{2} \frac{\|\mathbf{x} - \mathbf{x}_j\|^2}{l^2} \right), \tag{2.39}$$

which is a simple superposition of Gaussian functions. These functions are centered on the positions of the training points, $\mathbf{x}_j$, with a common width, l, and individual coefficients given by the j-th component of the matrix-vector product: $c_j = \left[\mathbf{\Sigma}(\mathbf{X}, \mathbf{X})^{-1}\mathbf{y}\right]_j$. (Note that the coefficients depend linearly on the training values.) The effect of the hyperparameters, σ_f and l, on the regression function becomes clear now as well. If we use a prior covariance function with implemented non-stationary features instead, each basis function in Eq. (2.38) would inherit such features and therefore the regression function as well, being just a weighted sum thereof. The crucial information of the training data, embedded in $\mathbf{y}$, only contributes to the values of the coefficients, c_j. Those coefficients are fine-tuned in order to ensure that this superposition of functions precisely goes through the whole set of training values irregardless of the values of the hyperparameters, e.g., σ_f and l.

The posterior covariance, Eq. (2.36), serves as a corrected prior covariance after the model

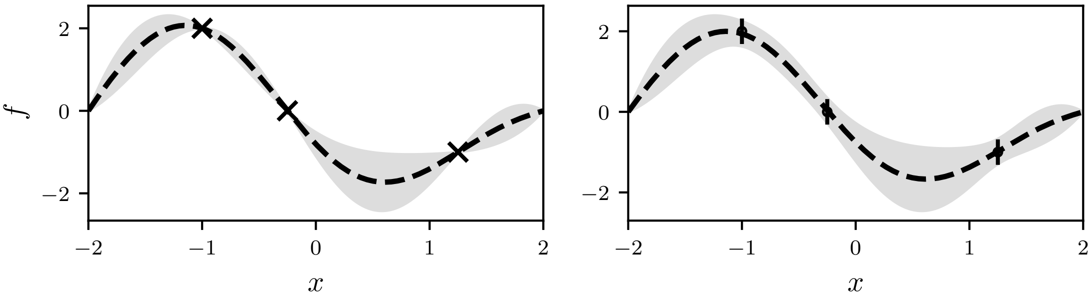

Figure 2.4: Two posterior GPs are shown using the prior from Fig. 2.3, left, conditioned on three data points. In the **left** panel, noiseless training data is assumed (crosses), while one the **right**, noise is added by setting ${\sigma_n}^2 = 0.1$ (errorbars).

has "learned" from training observations. The first term on the right hand side of Eq. (2.36) is the prior covariance. Subtracted from that is the second term representing the "gain in knowledge" about the latent input-output relationship coming from the training data. Note that it is independent of the actual values of the observations, $\mathbf{y}$. The diagonal of the posterior covariance yields the posterior variance, $\hat{\Sigma}(\mathbf{x},\mathbf{x}) = \hat{\sigma}^2(\mathbf{x})$, serving as the predicted uncertainty to the posterior mean. Depending on the observation noise, σ_n, it decreases towards the locations of the observations, as expected. For the stationary covariance functions discussed in Section 2.3, it approaches the value of the output-scale with increasing distance from training data, $\hat{\sigma}(\mathbf{x}) \xrightarrow{\|\mathbf{x}-\mathbf{x}_j\|\to\infty} \sigma_f$, and it approaches the noise on the positions of the observations, $\hat{\sigma}(\mathbf{x}_j) = \sigma_n$.

A new and somewhat special hyperparameter has sneaked in with the noise assumption in Eq. (2.26). As already briefly mentioned with Eqs. (2.35) and (2.36), Gaussian noise on the observations, $\mathbf{y}$, is implemented in the diagonal elements of the matrix $\mathbf{\Sigma}(\mathbf{X},\mathbf{X})$. More accurately, one should write $\mathbf{\Sigma}(\mathbf{X},\mathbf{X}) + {\sigma_n}^2\mathbf{I}$ instead, where $\mathbf{I}$ is the identity matrix and σ_n the hyperparameter for symmetric, Gaussian noise. One then yields for the coefficients of the posterior mean

$$c_j = \left[\left(\mathbf{\Sigma}(\mathbf{X},\mathbf{X}) + {\sigma_n}^2\mathbf{I}\right)^{-1}\mathbf{y}\right]_j . \tag{2.40}$$

Fig. 2.4, right, demonstrates the effect of Gaussian noise on the training observations by choosing a positive variance, ${\sigma_n}^2 = 0.1$. The posterior mean is no longer forced to precisely pass through the values of $\mathbf{y}$, but merely stays nearby, taking the uncertainty of the observations into consideration. It is further possible to consider different noise magnitudes for each data point by replacing the constant diagonal noise, representing SPHERICAL NOISE, with a diagonal noise matrix, ${\sigma_n}^2\mathbf{I} \to \mathbf{N}$, where each diagonal entry specifies the individual noise on each data point, $N_{ij} = {\sigma_{n,j}}^2\delta_{ij}$. The ability to implement noise on the training data is essential whenever the data origins from physical measurements or stochastic analyses.

The diagonal noise also plays an important role from a numerical aspect. When performing matrix inversion or solving a system of linear equations using a coefficient matrix, numerical inaccuracy can cause two rows or columns to be nearly linear dependent, which renders the matrix *singular*. In terms of Eq. (2.38), this can be caused by too large overlaps between two basis functions, $\int \Sigma(\mathbf{x};\mathbf{x}_i)\Sigma(\mathbf{x};\mathbf{x}_j)\,\mathrm{d}\mathbf{x}$. In order to avoid such errors when inverting the matrix in Eq. (2.40), little diagonal noise needs to be added.

Non-Gaussian noise on the training data corresponds to non-Gaussian likelihoods in Eq. (2.31), causing inference to be no longer tractable in closed form. This is the case for classification tasks.

Comparing ANNs and GPR

Comparing GPR to conventional regression methods, e.g., ANNs (including BNNs), it appears that the two are fundamentally different when inferring a regression function. In ANNs, as well as in other parametric ML models, the analytic form of the regression function is defined with the design of the model. This design defines the CAPACITY of the model and limits the upper bound for the complexity a regression function may have. In typical cases, this complexity may be far beyond that of the true latent input-output relationship. To approach overfitting of a given data set, ANNs can either be PRUNED, i.e., reducing the capacity by lowering the number of nodes, or one employs regularization methods, such as weight regularization, which causes the weights to stay small, and thus, reducing the capacity of the model. [12]

In contrast, the posterior mean function of a GP, Eq. (2.38), gains one term in the sum for every training point in the data set. For few training points the posterior mean will show little structure due to the little information given, as in Fig. 2.4. A large number of data will lead to a large number of basis functions, and hence, allows for more complexity in the regression function. This is a desirable behavior as it prevents overfitting at few data points and at the same time enhances flexibility for rich-structured posterior means at large data sets. As mentioned earlier, the issue of the choice of the right design never occurs. Other kernel methods used for supervised ML, such as SUPPORT VECTOR MACHINES [23], share this property. On the downside, multi-layered ANNs possess the ability to hierarchically learn complicated features in high-dimensional training data that simple single-layer GPs fail to resolve. Therefor an extension to DEEP GAUSSIAN PROCESSES appears necessary, constituting multi-layer GPs.

2.4.3 Learning hyperparameters

Being a non-parametric method for supervised ML, GPs require no adjustment of parameters in order to fit training data. The variables inside the covariance function, e.g., σ_f and l, together with the standard deviation of the Gaussian noise on the training data, σ_n, form the set of HYPERPARAMETERS and play a special role. Other than the parameters in a parametric model, e.g., the weights in an ANN, each reasonable set of values for the hyperparameters will provide a meaningful regression function that fits all data points. By varying the hyperparameters the characteristics of the posterior can be adjusted. This is also referred to as MODEL SELECTION.

As a demonstrative example, let us consider a GPR using the RBF covariance function, Eq. (2.12), similar as in Fig. 2.2, but with a zero mean prior. The hyperparameters shall be collectively denoted as $\boldsymbol{\theta} = (\sigma_f, l, \sigma_n)$. Each hyperparameter has a different effect on the posterior. The effect of Gaussian noise, σ_n, is illustrated in Fig. 2.4. The squared outputscale, ${\sigma_f}^2$, is the prefactor of the covariance and therefore proportionally effects the variance of both, prior and posterior; see Eq. (2.36). Its effect on the posterior mean is being counteracted by the coefficients which depend, neglecting noise, inversely on the square of the output-scale, $c_j \propto {\sigma_f}^{-2}$; see Eqs. (2.39) and (2.40). The length-scale, l, directly effects both posterior mean and covariance. For all stationary covariance functions discussed in Sec. 2.3, the length-scale inversely scales the distance, $\|\mathbf{x} - \mathbf{x}'\|/l$. This leads to longer or shorter correlations between inputs, $\mathbf{x}$ and $\mathbf{x}'$. For the RBF covariance, Eq. (2.39), larger length-scales correspond to wider Gaussian basis functions. The posterior mean therefore varies more slowly with $\mathbf{x}$ when the length-scale is large. This effect is illustrated in the left panel of Fig. 2.5. Other covariance functions may bring new hyperparameters with different effects on the posterior. Their effects can be interpreted by inspecting Eqs. (2.38) and (2.36).

Though for each set of hyperparameters a posterior mean can infer a proper regression on training data, the question arises whether one set of hyperparameter values performs better than another. As a posterior GPs gains flexibility with the size of the training data, it can become vulnerable to over- and underfitting. OVERFITTING is the term for precise fitting of training data, while performing poorly at predictions. Then, a too complex regression function is chosen for a less complex latent function. This occurs for too small length-scales, $l \to 0$. On the contrary, UNDERFITTING describes the phenomenon of fitting training data too loosely by not capturing the detailed features of the latent function. This happens if the noise variance is chosen too large, $\sigma_n \gtrsim \sigma_f$.

Maximum likelihood estimation

In statistics, the MAXIMUM LIKELIHOOD ESTIMATION (MLE) is a method for finding estimators for the parameters of probability distributions approximately modeling the unknown PDFs of stochastic phenomena. The likelihood is a function of the parameters, $\boldsymbol{\theta}$, given a fixed single measurement, $\mathbf{y}$,

$$L(\boldsymbol{\theta};\mathbf{y}) = c\,\mathcal{P}(\mathbf{y};\boldsymbol{\theta}). \tag{2.41}$$

For K independent and identically distributed measurements, the likelihood is given as the product of the likelihoods for each single measurement,

$$L(\boldsymbol{\theta};\mathbf{y}_1,\dots,\mathbf{y}_K) = \prod_{i=1}^{K} L(\boldsymbol{\theta};\mathbf{y}_i) = c\prod_{i=1}^{K} \mathcal{P}(\mathbf{y}_i;\boldsymbol{\theta}). \tag{2.42}$$

The constant prefactor, c, is merely a reminder that the likelihood function is not a valid PDF as it needs not be normalized. The relative value of the likelihood can be interpreted as how likely a set of measurements is under the assumption of a particular probabilistic model. The goal is to aim at high values for the likelihood by adjusting the model, namely by optimizing its parameters. The MLE method returns those estimators for the parameters for which the likelihood is maximal,

$$L(\hat{\boldsymbol{\theta}}) = \max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}). \tag{2.43}$$

Often it is computationally convenient to consider the natural logarithm of the likelihood instead: $\mathcal{L}(\boldsymbol{\theta};\mathbf{y}) := \log L(\boldsymbol{\theta};\mathbf{y})$. Appendix A.1 demonstrates as an example the well-known arithmetic mean as a maximum likelihood estimator.

For our GPR case, the probabilistic model is our prior GP and its parameters are the hy-

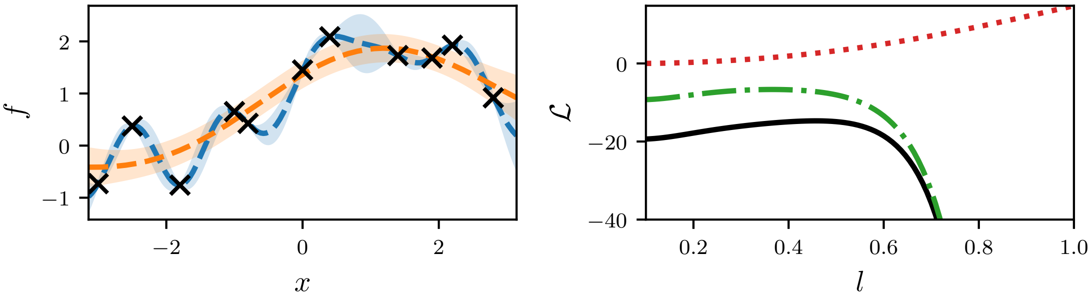

Figure 2.5: The effect of different length-scales, l, of the prior RBF covariance is illustrated. **Left**, two posteriors are shown using different hyperparameter sets, $\boldsymbol{\theta} = (\sigma_f, l, \sigma_n)$. The blue posterior uses $\boldsymbol{\theta} = (1, 0.5, 0)$, while the orange posterior uses $\boldsymbol{\theta} = (1, 2, 0.5)$. The data is generated using the GP from Fig. 2.2. The hyperparameters of the blue posterior are optimized via MAXIMUM LIKELIHOOD ESTIMATION (MLE) and therefore correctly reproduce the hyperparameters of the generating GP. In the **right** panel the corresponding log-likelihood is shown (solid black) together with its data-fit (dash-dotted green) and negative complexity penalty (dotted red) contributions. The other hyperparameters are fixed at $(\sigma_f, \sigma_n) = (1, 0)$.

perparameters, $\boldsymbol{\theta}$. The training observations, $\mathbf{y}$, serve as one single measurement vector, assuming we "measured" only once at each point. Using the MLE method we can find the set of hyperparameters under which the observed training data is most likely. The procedure of optimizing the hyperparameters is the LEARNING step in a GPR task. For a zero mean GP the likelihood is given by (omitting the arbitrary factor c) the marginal distribution of $\mathbf{y}$, which is again Gaussian according to the marginalization theorem. The log-likelihood is then

$$\mathcal{L}(\boldsymbol{\theta}; \mathbf{y}) = \log \mathcal{P}(\mathbf{y}; \boldsymbol{\theta}) = \underbrace{-\frac{1}{2}\mathbf{y}^\top {\boldsymbol{\Sigma}_n}^{-1}\mathbf{y}}_{\text{data-fit}} \underbrace{- \frac{1}{2}\log\det(\boldsymbol{\Sigma}_n)}_{\text{complexity penalty}} - \frac{N}{2}\log 2\pi, \tag{2.44}$$

where $\boldsymbol{\Sigma}_n \equiv \boldsymbol{\Sigma}(\mathbf{X}, \mathbf{X}) + {\sigma_n}^2\mathbf{I}$. The argmax of $\mathcal{L}(\boldsymbol{\theta}; \mathbf{y})$ provides good estimators for $\boldsymbol{\theta}$. Evidently, this states a global optimization problem with the dimensionality given by the number of hyperparameters.

The three terms in Eq. (2.44) play different roles in the optimization of the hyperparameters. [13] Only the first term contains the observed data and corresponds to the data-fit; the second term serves as the complexity penalty depending only on the covariance matrix at the training points. The third term is merely a normalization constant. The dependence of the log-likelihood and its constituents on the length-scale is depicted in Fig. 2.5, right, where the output-scale and noise were kept fixed at $(\sigma_f, \sigma_n) = (1, 0)$. For large length-scales the data-fit term monotonically decreases as it loses the flexibility needed to describe the noiseless training values, $\mathbf{y}$. The negative complexity penalty increases with the length-scale due to the model becoming less complex. This contribution is essential for preventing overfitting. The log-likelihood experiences a peak around $l = 0.5$, which constitutes the maximum likelihood estimator for l (at fixed σ_f and σ_n). This value accurately reproduces the length-scale of the GP from which the training data was generated.

In practice all hyperparameters of a GP prior are simultaneously optimized using to the MLE method. In order to find the optimum numerically, any gradient-based optimizer will perform well. We choose the 'L-BFGS-B' algorithm, which is the default optimizer in the SCIPY package for python. This optimizer has proven to give satisfactory results if the gradient of the log-likelihood is provided. The gradient is given in Ref. [13],

$$\frac{\partial}{\partial \theta_j}\mathcal{L}(\boldsymbol{\theta};\mathbf{y}) = \frac{1}{2}\operatorname{tr}\left((\boldsymbol{\alpha}\boldsymbol{\alpha}^\top - {\boldsymbol{\Sigma}_n}^{-1}) \frac{\partial \boldsymbol{\Sigma}_n}{\partial \theta_j} \right), \tag{2.45}$$

where $\boldsymbol{\alpha} = {\boldsymbol{\Sigma}_n}^{-1}\mathbf{y}$. The provision of the gradient is optional; if no gradient is provided, the optimizer will locally approximate the gradient using additional function evaluations.

One must always be cautious when searching for the global maximum to not get stuck in local maxima. This can be prevented by choosing sensible initial values for the hyperparameters. Choosing a prior mean function too close to the training data will create a local maximum of the likelihood around large values for the length-scale and the noise, as the data can be described as white noise around the prior mean. Therefore, it is generally not recommended to adjust prior mean functions to closely resemble observed data, as their benefits are neglectable anyway.

The computational burden caused by Eqs. (2.44) and (2.45) is dominated by the inversion of $\boldsymbol{\Sigma}_n$, which scales cubically with the number of training points, and is therefore equal to that of doing inference using Eqs. (2.35) and (2.36). However, throughout the optimization of hyperparameters the likelihood and its gradients must be computed in each step. For this reason, learning the hyperparameters causes much more computational expense than a single inference.

Notes on sampling from a GP

In supervised ML applications GPs are used to infer from data to make probabilistic predictions. For that the practice of creating actual samples is not needed. Samples from a customized GP could be used in order to generate data for the training of other ML models or in a generative adversarial network. The crucial point to consider is that the number of discrete evaluations of the sample function represents the dimension of the input for the ML model. An analytic expression of a single sample function is not provided, but could be approximated, e.g., via GPR.

2.5 Scaling to large data sets in higher dimensions

The major drawbacks of GPR are the computational cost of inference, Eqs. (2.35) and (2.36), and hyperparameter learning, Eqs. (2.44) and (2.45). They involve the inversion of the covariance matrix, ${\boldsymbol{\Sigma}_n}^{-1} = \left(\boldsymbol{\Sigma}(\mathbf{X},\mathbf{X}) + {\sigma_n}^2\mathbf{I}\right)^{-1}$, as well as the computation of the complexity penalty for hyperparameter learning, $\log\det(\boldsymbol{\Sigma}_n)$. Both scale cubically with the number of training data, $\mathcal{O}(N^3)$. For data sets containing more than about 10.000 points those operations may take computation time of the order of minutes. Sadly, the required amount of training data suffers badly from the notorious CURSE OF DIMENSIONALITY: to thoroughly cover higher-dimensional input spaces, $D > 1$, the required amount of data grows exponentially with the dimension, $\mathcal{O}(N^D)$. Hence, for larger data sets one is compelled to find approaches that improve upon the cubic scaling. Such approaches can be classified into two groups. The first group relies on properties of the covariance matrix and the training data that simplify computations while keeping the results exact; see Sec. 2.5.1. The second group uses approximations while keeping generality in the data; see Sec. 2.5.2.

2.5.1 Exact inference on structured data

Underlying structure in the positions of the training data can allow for enhancing alterations when computing either ${\boldsymbol\Sigma_n}^{-1}$ directly, or instead solving ${\boldsymbol\Sigma_n}^{-1}\mathbf{y}$. In particular, two structural properties allow for major simplifications for computations involving the inverse, namely Toeplitz and Kronecker structure. Both are complementary and are used in modern frameworks for scaling GPs to large multidimensional data sets.

Toeplitz structure

For the case of one-dimensional input, the matrix of a stationary covariance function simplifies significantly if the data is equidistantly spaced. Since $\Sigma(\mathbf{x}, \mathbf{x}') = \Sigma(\mathbf{x}-\mathbf{x}')$ for a stationary covariance, each diagonal of the covariance matrix then has equal entries, $\Sigma_{ij} = \Sigma_{i+1,j+1}$. Such a matrix is called Toeplitz (or diagonal-constant) matrix. Because of the symmetry of the covariance, $\Sigma_{ij} = \Sigma_{ji}$, the full information of $\boldsymbol\Sigma_n = \boldsymbol\Sigma + {\sigma_n}^2\mathbf{I}$ is stored in one single row or column. Exploiting Toeplitz structure, matrix-vector multiplication can be done in $\mathcal{O}(N \log N)$ operations. [24] Ref. [9] provides the corresponding implementation using fast Fourier transforms. Thus, inference is efficiently done by solving the system of linear equations

$$\boldsymbol\Sigma_n \mathbf{z} = \mathbf{y} \tag{2.46}$$

for $\mathbf{z} = {\boldsymbol\Sigma_n}^{-1}\mathbf{y}$ via the conjugate gradient (CG) method [25], which exclusively applies matrix-vector products. The total cost of inference then drops from cubic to $\mathcal{O}(JN \log N)$, where J is the number of iterations of the CG method.

Less efficiently provided is the complexity penalty. Exploiting Toeplitz structure, Ref. [9] shows that $\log\det(\boldsymbol\Sigma_n)$ can be computed in $\mathcal{O}(N^2)$ operations.

Kronecker structure

Matrices that decompose into a Kronecker product pose as the most desirable category,

$$\boldsymbol\Sigma = \boldsymbol\Sigma_1 \otimes \boldsymbol\Sigma_2 \otimes \cdots \otimes \boldsymbol\Sigma_D = \bigotimes_{d=1}^{D} \boldsymbol\Sigma_d. \tag{2.47}$$

Although the inverse of such a Kronecker matrix can be easily computed as

$$\boldsymbol\Sigma^{-1} = \boldsymbol\Sigma_1^{-1} \otimes \cdots \otimes \boldsymbol\Sigma_D^{-1}, \tag{2.48}$$

our actual target is the inverse of $\boldsymbol\Sigma + {\sigma_n}^2\mathbf{I}$, which incorporates additional noise. Ref. [26] shows that by exploiting the Kronecker form one can efficiently solve $(\boldsymbol\Sigma + {\sigma_n}^2\mathbf{I})^{-1}\mathbf{y}$ and

$\log\det\left(\boldsymbol{\Sigma}+{\sigma_n}^2\mathbf{I}\right)$. Using the eigendecompositions, $\boldsymbol{\Sigma}_d = \mathbf{Q}_d\boldsymbol{\Lambda}_d\mathbf{Q}_d^\top$, one can rewrite

$$\left(\bigotimes_{d=1}^{D}\boldsymbol{\Sigma}_d+{\sigma_n}^2\mathbf{I}\right)^{-1}\mathbf{y} = \left(\bigotimes_{d=1}^{D}\mathbf{Q}_d\boldsymbol{\Lambda}_d\mathbf{Q}_d^\top+{\sigma_n}^2\mathbf{I}\right)^{-1}\mathbf{y} \tag{2.49}$$

$$= \left(\left[\bigotimes_{d=1}^{D}\mathbf{Q}_d\right]\left[\bigotimes_{d=1}^{D}\boldsymbol{\Lambda}_d\right]\left[\bigotimes_{d=1}^{D}\mathbf{Q}_d\right]^\top+{\sigma_n}^2\mathbf{I}\right)^{-1}\mathbf{y} \tag{2.50}$$

$$= \left(\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\top+{\sigma_n}^2\mathbf{I}\right)^{-1}\mathbf{y} \tag{2.51}$$

$$= \mathbf{Q}\left(\boldsymbol{\Lambda}+{\sigma_n}^2\mathbf{I}\right)^{-1}\mathbf{Q}^\top\mathbf{y}. \tag{2.52}$$

Therefor the orthogonality of $\mathbf{Q}$ (since the $\boldsymbol{\Sigma}_d$ are symmetric) and the properties of the Kronecker product were used. Computing the inverse of the diagonal matrix in Eq. (2.52) is trivial. What is left are the eigendecompositions of each much smaller factor matrix, $\boldsymbol{\Sigma}_d$, and the more expensive matrix-vector products. Latter can be efficiently implemented exploiting the Kronecker form, reducing the cost from $\mathcal{O}(N^2)$ to $\mathcal{O}(DN^{\frac{D+1}{D}})$, with $D \ll N$, see Appendix A.2. Using the eigenvalues of $\boldsymbol{\Sigma}$, which are the entries of the diagonal matrix $\boldsymbol{\Lambda}$, namely λ_j, one can easily compute the complexity penalty,

$$\log\det\left(\boldsymbol{\Sigma}+{\sigma_n}^2\mathbf{I}\right) = \log\prod_{j=1}^{N}\left(\lambda_j+{\sigma_n}^2\right) = \sum_{j=1}^{N}\log\left(\lambda_j+{\sigma_n}^2\right). \tag{2.53}$$

In total, the computational expense drops from cubic to $\mathcal{O}\left(DN^{\frac{D+1}{D}}\right)$ and the covariance matrix can be stored in $\mathcal{O}(DN^{2/D})$ instead of $\mathcal{O}(N^2)$. This is a huge advantage; the complexity then scales roughly linearly with the dimension and the number of points in the training data set.

The question arises when a covariance matrix becomes a Kronecker product: a product of one-dimensional covariance functions,

$$\Sigma(\mathbf{x},\mathbf{x}') = \prod_{d=1}^{D}\Sigma_d(x_d,x_d'), \tag{2.54}$$

can form a Kronecker product if the training data lies on a multidimensional Euclidean lattice,

$$\left\{\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(N)}\right\} \subset \mathcal{X} = \mathcal{X}_1\times\cdots\times\mathcal{X}_D \subseteq \mathbb{R}^D. \tag{2.55}$$

The grid spacing in each dimension needs not be equidistant, but if it is, the Toeplitz structure of the factor matrices can be exploited additionally. The prerequisite of Eq. (2.54) is minor restrictive, as many popular multidimensional covariance functions possess this property, in particular the RBF and the SM covariance functions. The major restriction states the grid structure of the data. Such data can appear in image analysis, sensor arrays, and, more importantly, numerical data assimilation due to the data being often generated on discrete grids of parameters. However, training sets forming complete grids are rarely provided.

The full grid requirement can be relaxed to incomplete grids [27,28] (or almost grid-like data sets) by adding data points with arbitrary output value and infinite Gaussian noise in order to complete a full grid. Those additional points serve as dummies and permit a Kronecker decomposition as in Eq. (2.47). The original training data set forming the incomplete grid, $\{(\mathbf{x}_j, y_j); j = 1, \ldots, N\}$, is then expanded by M points with $(y_{N+1}, \ldots, y_{N+M}) = \mathbf{0}$ and positions, $(\mathbf{x}_{N+1}, \ldots, \mathbf{x}_{N+M})$, such that together they fill all $N+M$ grid sites. The covariance matrix and training vector for the completed grid shall be denoted as $\mathbf{\Sigma}^{(N+M)}$ and $\mathbf{y}^{(N+M)}$, respectively. The diagonal noise matrix is then correspondingly expanded to

$$\mathbf{D}_n = \begin{pmatrix} {\sigma_n}^2 \mathbf{I}^N & 0 \\ 0 & \epsilon \mathbf{I}^M \end{pmatrix}, \tag{2.56}$$

where $\mathbf{I}^N$ stands for the identity matrix of dimension $\mathbb{R}^{N\times N}$ and the dummy noise tends to infinity, $\epsilon \to \infty$. As before, variable individual noise on the original observations can be implemented by substituting ${\sigma_n}^2 \mathbf{I}^N \to \mathbf{N}$, where $N_{ij} = {\sigma_{n}}_{,j}^2 \delta_{ij}$. Refs. [27,28] show that the augmentation of the dummy set has no corrupting effect on the inference result,

$$\left(\mathbf{\Sigma}^{(N+M)} + \mathbf{D}_n\right)^{-1} \mathbf{y}^{(N+M)} = \left(\mathbf{\Sigma} + {\sigma_n}^2 \mathbf{I}\right)^{-1} \mathbf{y}. \tag{2.57}$$

Since the diagonal noise matrix, Eq. (2.56), is no longer a multiple of the identity matrix, the previous decomposition, Eqs. (2.49) – (2.52), is no longer possible. Instead, inference can be efficiently done by solving the system of linear equations

$$\left(\mathbf{\Sigma}^{(N+M)} + \mathbf{D}_n\right) \mathbf{z} = \mathbf{y} \tag{2.58}$$

for $\mathbf{z} = \left(\mathbf{\Sigma}^{(N+M)} + \mathbf{D}_n\right)^{-1} \mathbf{y}$ via the CG method. This method iteratively solves Eq. (2.58) by solely applying matrix-vector products. Same as for Eq. (2.52), products of Kronecker matrices and vectors can be efficiently computed in $\mathcal{O}(D(N+M)^{\frac{D+1}{D}})$ operations, see Appendix A.2. The computational burden of the CG solver is then $\mathcal{O}(JD(N+M)^{\frac{D+1}{D}})$, where J is the number of iterations. In Refs. [27,28] the preconditioner matrix for the CG method, ${\mathbf{D}_n}^{-\frac{1}{2}}$, is suggested. However, this choice of preconditioner has not improved convergence in our experiments.

On the downside, the complexity penalty for hyperparameter learning, Eq. (2.53), must be approximated. Two methods were proposed, where the first method is limited to spherical noise on the original data on the incomplete grid, ${\sigma_n}^2 \mathbf{I}$. For this case, Ref. [28] used the observation that the dummy variables do not alter the result, Eq. (2.57), and suggests to consider only the original data. They apply Eq. (2.53) and approximate the eigenvalues of $\mathbf{\Sigma}$ by using the N largest eigenvalues of $\mathbf{\Sigma}^{(N+M)}$,

$$\log\det\left(\mathbf{\Sigma} + {\sigma_n}^2 \mathbf{I}\right) = \sum_{j=1}^{N} \log\left(\lambda_j + {\sigma_n}^2\right) \approx \sum_{j=1}^{N} \log\left(\frac{N}{N+M} \lambda_j^{(N+M)} + {\sigma_n}^2\right), \tag{2.59}$$

where $\lambda_j^{(N+M)}$ are the eigenvalues of $\mathbf{\Sigma}^{(N+M)}$, which can be cheaply computed due to its Kronecker form.

On the other hand, for the more general case of variable, individual noise, $N_{ij} = {\sigma_{n}}_{,j}^2 \delta_{ij}$, Ref. [27] suggests to consider the completed-grid data set and approximates the noise matrix with a scaled identity matrix,

$$\log\det\Big(\mathbf{\Sigma}^{(N+M)}+\mathbf{D}_n\Big)\approx\log\det\Big(\mathbf{\Sigma}^{(N+M)}+\gamma(\mathbf{D}_n)\mathbf{I}\Big)=\sum_{j=1}^{N+M}\log\Big(\lambda_j^{(N+M)}+\gamma(\mathbf{D}_n)\Big), \tag{2.60}$$

where $\gamma(\mathbf{D}_n)=\Big(\prod_{j=1}^{N+M}\mathbf{D}_{n,jj}\Big)^{\frac{1}{N+M}}$ is the geometric mean. In either case, only the complexity penalty, hence the learning of the hyperparameters, undergoes a small approximation while inference remains exact. The total computational cost for data on incomplete grids is now $\mathcal{O}(JD(N+M)^{\frac{D+1}{D}})$ instead of $\mathcal{O}(DN^{\frac{D+1}{D}})$ for the data producing a complete lattice. Again, the number of operations scales superlinearly with the number of training points. The storage cost remains the same.

Above considerations are suited to handle data that exhibit partial grid structure. If a data set has even one dimension along which it shows no grid structure at all the computational cost becomes $\mathcal{O}(N^3)$ again, since one Kronecker factor matrix for the completed grid will have dimension N.

2.5.2 Approximative inference on generic data

In the previous section Toeplitz and Kronecker properties are used to immensely improve upon the scalability of GPR. While keeping inference exact, both rely on lattice structures in the training data, rendering such improvements rarely applicable. For generic data, inference and learning can be sped up by employing approximations to the covariance matrix or its inverse. Early popular methods rely on subsets of training locations, so called INDUCING POINTS, in order to reduce the size of the covariance matrix to be inverted. Such methods are coined GREEDY or SPARSE approximations, many of which are summarized in Refs. [29,30]. They all have in common that the number of inducing points, M, must be chosen much smaller than the size of the training data set, $M \ll N$, in order to improve upon the computational effort, at the expense of accuracy for the regression. The computational cost then reduces to $\mathcal{O}(M^2N+M^3)$. Still, the curse of dimensionality remains unbroken since the number of necessary inducing points exponentially grows with the dimension.

Refs. [31,32] propose to combine such approximations using inducing points with the efficiency gains coming from structured data, discussed in Sec. 2.5.1, by placing the inducing points, $(\mathbf{u}_1,\dots,\mathbf{u}_M)$, on a regular grid. Due to the lattice structure a much larger set of inducing points can be treated, $M>N$. The covariance at training locations is then cubically interpolated between two neighboring inducing points, resulting in a sparse matrix with interpolations weights, $\mathbf{W}\in\mathbb{R}^{N\times M}$. The original covariance matrix of the training inputs is then approximated by a simple matrix product of sparse weight matrices and the covariance matrix of the inducing point lattice with Toeplitz and Kronecker forms, $\mathbf{\Sigma}(\mathbf{X},\mathbf{X})\approx\mathbf{W}\mathbf{\Sigma}(\mathbf{U},\mathbf{U})\mathbf{W}^\top\equiv\mathbf{\Sigma}_{\mathrm{SKI}}$. The authors call this method STRUCTURED KERNEL INTERPOLATION (SKI). With fast matrix-vector multiplication enabled, inference is done by solving the system of linear equations

$$\left(\mathbf{\Sigma}_{\mathrm{SKI}}+\sigma_n\mathbf{I}\right)\mathbf{z}=\mathbf{y} \tag{2.61}$$

for $\mathbf{z}=\left(\mathbf{\Sigma}_{\mathrm{SKI}}+\sigma_n\mathbf{I}\right)^{-1}\mathbf{y}\approx\left(\mathbf{\Sigma}+\sigma_n\mathbf{I}\right)^{-1}\mathbf{y}$ via the CG method. The complexity penalty for hyperparameter learning can be computed using Eq. (2.59), where $\mathbf{\Sigma}_{\text{SKI}}$ takes the role of $\mathbf{\Sigma}^{(N+M)}$. Using a large number of inducing points reduces the magnitude of the approximation without sacrificing scalability into multiple dimensions. The restrictions on the training data set are completely lifted. The total computational complexity is thus reduced to $\mathcal{O}(N + JM \log M)$ operations and $\mathcal{O}(N + M)$ storage when exploiting Toeplitz structure for $D = 1$, and $\mathcal{O}(JDM^{\frac{D+1}{D}})$ operations and $\mathcal{O}(N + DM^{\frac{2}{D}})$ storage when exploiting Kronecker structure for $D > 1$.

Evidently, the initial cost of $\mathcal{O}(N^3)$ for GPR is immensely reduced. The curse of dimensionality, however, remains in the number of inducing points, M, which grows exponentially with D. In order to enable GPR in very high input dimensions, $D \gg 5$, Ref. [33] further employs the Lanczos approximation to the covariance matrix resulting in truly linear runtime with the dimension. An alternative approach is DKL [20], where an SKI GP is used as an output layer of a deep ANN, improving upon expressive power and scalability at the same time.

For the predictive covariance, Eq. (2.36) remains to be efficiently solved. Building upon the SKI method in Ref. [31], Ref. [34] supplies an approximate method for computing the uncertainty and sampling from the posterior distribution using a Lanczos decomposition of the prior covariance matrix.

2.5.3 On the convergence of the CG method

In this section it was shown that many methods for lowering the computational complexity efficiently use fast matrix-vector multiplication. The costly computations involving the inverse of the covariance matrix are then executed using the CG method for solving systems of linear equations, since it relies solely on matrix-vector products. The total computational cost then scales like $\mathcal{O}(JM \log M)$ for Toeplitz and $\mathcal{O}(JDM^{\frac{D+1}{D}})$ for Kronecker matrices, where J is the number of iterations of the CG method and M the number of training/inducing points, i.e., the dimension of the covariance matrix.

The CG method iteratively solves a system of linear equations, $\mathbf{A}\mathbf{z} = \mathbf{y}$. From an (arbitrarily) chosen starting point, $\mathbf{z}_0$, the norm of the residual, $\|\mathbf{r}_i\| = \|\mathbf{y} - \mathbf{A}\mathbf{z}_i\|$, is decreased by moving, step by step, in $\mathbf{A}$-conjugate directions, i.e., $\mathbf{d}_i^\top \mathbf{A}\mathbf{d}_j = 0 \ \forall i \neq j$. Neglecting round-off errors, the method arrives at its solution, $\|\mathbf{r}_J\| = 0$, in $J \leq M$ iterations, where $\mathbf{A} \in \mathbb{R}^{M\times M}$.

In practice, however, the number of iterations heavily depends on the condition of matrix $\mathbf{A}$. For well-conditioned covariance matrices, $J \ll M$, rendering the effective computational costs of inference from $\mathcal{O}(JM \log M)$ for Toeplitz and $\mathcal{O}(JDM^{\frac{D+1}{D}})$ for Kronecker structures to $\mathcal{O}(M \log M)$ and $\mathcal{O}(DM^{\frac{D+1}{D}})$, respectively. On the other hand, for ill-conditioned covariance matrices, J can become much larger than M. More precisely, we find that for longer correlations, i.e., larger length-scales, l, the CG algorithm converges weaker and the number of iterations become prohibitively large, $J > M$, worsening the overall computational burden.

In Fig. 2.6 the effect of the length-scale in the RBF covariance function, Eq. (2.12), on the convergence of the CG method is depicted. The underlying set of data points is located on a complete four-dimensional lattice with ten equidistant points in each dimension, summing up to a total of $M = 10^4$ points. The distance between two neighboring points is 1. The residual in each iteration is shown for the five different length-scales $l = 0.5, 0.75, 1.0, 1.25, 1.5$. The larger the length-scale is chosen, the slower the CG method converges. The other hyperparameters, σ_f and σ_n, as well as the values of the training data, $\mathbf{y}$, do not alter the scaling behavior of the convergence. Note that since the optimal value for the length-scale according to the MLE, Eq. (2.44), depends on the values of $\mathbf{y}$, the range of optimal l is in general unbound. As a consequence, the computational efficiency of the CG method heavily depends on the length-scale of the covariance function. In the experiment in Fig. 2.6, the number of iterations already exceeds the number of data points at length-scales larger than 1.25 times the nearest neighbor distance. Prohibitively large numbers of iterations of the CG solver, $J > M$, render many up-scaling methods discussed within this section counterproductive. This kind of behavior is not discussed in the cited literature.

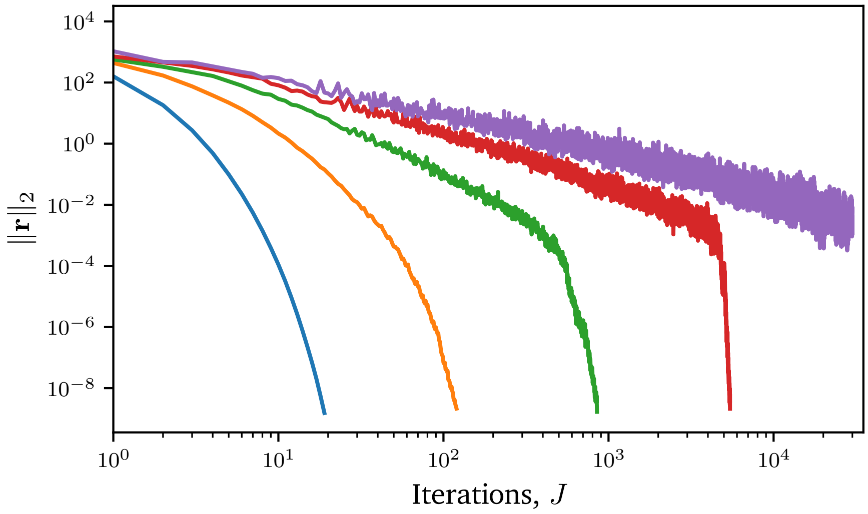

Figure 2.6: The effect of different length-scales, l, of the RBF covariance matrix on the convergence of the CG method is illustrated. The values are $l = 0.5$ (blue), 0.75 (orange), 1.0 (green), 1.25 (red), 1.5 (purple). The iterations were stopped when either a threshold of the residual, $\|\mathbf{r}\|_2$, was reached, or the number of iterations exceeded $3{\cdot}10^4$. The training points were located on the sites of a four-dimensional Euclidean lattice with 10 points in each dimension and distances of 1.

2.6 Active learning

A GP naturally provides uncertainty measures. A practical task could not only be the fitting of a training data set, but also come with the question of where additional training points are needed the most. We shall call such a task ACTIVE LEARNING (AL). The uncertainty of a posterior GP can be a sensible measure for the lack of information about the true latent function, thus enabling AL procedures.

If one considers any stationary covariance function, such as those discussed in Sec. 2.3, one finds that the uncertainty grows with the distance to the training data. This behavior is exactly what one expects and was already discussed in Sec. 2.4.2. However, for an open range in the input space, this is generally little useful for AL, as new training locations, $\mathbf{x}^\star$, would be suggested at $\|\mathbf{x}^\star - \mathbf{x}_n\| \to \infty$, where $\mathbf{x}_n$ is the nearest training point. For the case of finite, discrete test points GPs significantly improve the sampling for additional training points over random sampling, as shown in Ref. [35]. For continuous input spaces one has to define boundaries by incorporating boundary conditions into the GP. [36]

Coming back to our two examples for incorporating non-stationary features into a GP in Sec. 2.3.1, we can demonstrate AL procedures using the uncertainty to successively find locations where new training points ought to be added.

Example: a particle in a box

One of the simplest, exactly solvable quantum systems to consider is a free particle in a Cartesian box with infinitely high potential walls,

$$V(\mathbf{x}) = \begin{cases} 0 & \text{if} \quad 0 \leq x_i \leq L_i \quad \forall i, \\ \infty & \text{else.} \end{cases} \tag{2.62}$$

Since the potential is time-independent, the eigenstates of this system are given as solutions of the time-independent Schrödinger equation,

$$H\psi(\mathbf{x}) = \left[-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{x}) \right] \psi(\mathbf{x}) = E\psi(\mathbf{x}), \tag{2.63}$$

all of which have the properties of being continuously differentiable inside the box and zero outside the box. Such properties were already considered in Sec. 2.3.1, where Dirichlet boundary conditions were implemented into a GP, Eqs. (2.18) and (2.19). Using such a prior (with zero mean) we can conduct an AL procedure for reconstructing an eigenstate by efficiently placing training points in a subsequent manner.

Fig. 2.7 illustrates such a procedure for the eigenstate with quantum numbers $(2, 1)$. Before possessing any training data, the standard deviation of the prior GP, σ, having the proper boundary features implemented, suggests placing the first training point in the center of the box, where the standard deviation is maximal; see the top right panel. After evaluating the true wave function at this point, we compute the first posterior GP by conditioning the prior GP on this first training point. The second training point is then placed, where the posterior uncertainty is maximal, yielding the second training point. Repeating this a number of times one efficiently collects a sensible training data set for the given eigenstate. The two bottom panels of Fig. 2.7 show the GPR after ten iterations. The two optimized length-scales of the

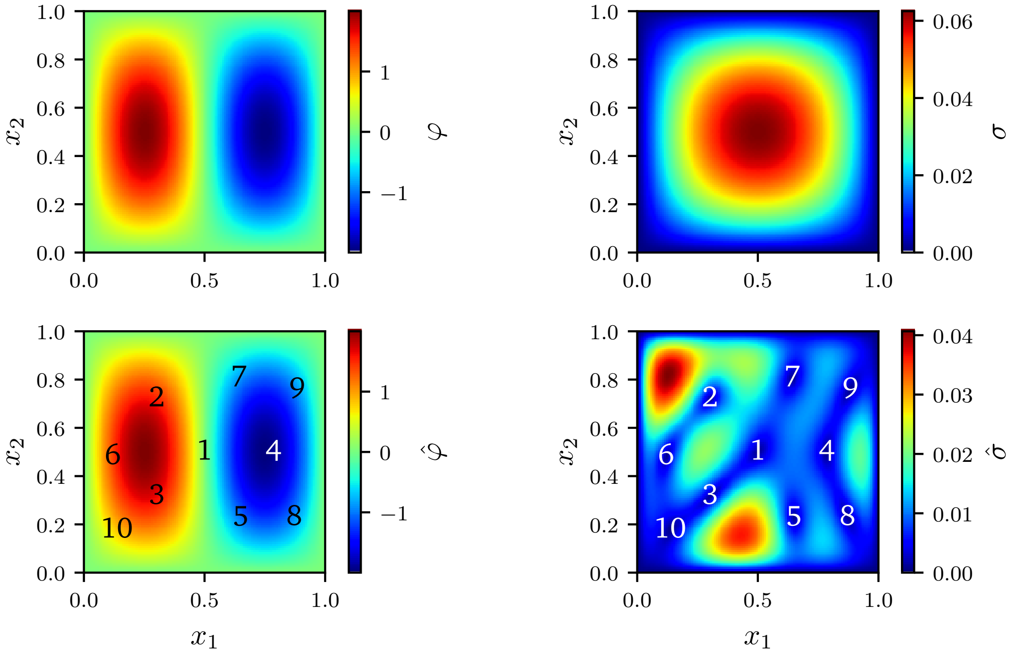

Figure 2.7: An AL procedure for the $(2, 1)$ eigenstate of a particle in a 2-dim. box is illustrated. The **top left** panel shows the true eigenstate; the **bottom left** panel shows the posterior mean using the ten training points depicted as their number of inclusion into the training set. The two **right** panels show the uncertainties of the prior (**top**) and posterior GP (**bottom**).

posterior are $l_1 \approx 0.4$ for the x_1-direction and $l_2 \approx 1.0$ for the x_2-direction, expressing the fact that ψ has a node in x_1-direction but none in x_2-direction.

Note that it was assumed that the true wave function can be evaluated. In a realistic scenario this could possibly not be the case. Still, AL can be conducted since the posterior covariance, Eq. (2.36), does not depend on the training values, $\mathbf{y}$, but only on the positions, $\mathbf{x}_j$. After actively learning N positions of a training set, the corresponding $\mathbf{y}$-values can be evaluated by diagonalizing the Hamiltonian, H, in the basis of the GP basis functions, see Eq. (2.38). This would state a generalized eigenvalue problem, the solutions of which yield the first N approximated eigenenergies, E_i, and the corresponding eigenvectors, $\mathbf{v}_i$, whose elements are the coefficients, c_{ij}, from Eq. (2.38). The only difference of AL without evaluating the true target function is the fact that the hyperparameters cannot be adjusted to the training data in each step, Eq. (2.44). This is, however, not problematic, since the diagonalization of the system Hamiltonian yields all N eigenstates, each of which will have different optimal hyperparameters anyway.

Example: the H_2^+ molecular ion

As a second example for AL we consider the case of an singly ionized hydrogen molecule, H_2^+. The properties of the electronic eigenstates of this system were discussed in Sec. 2.3.1, where a GP with same properties was developed. Specifically, we employ the prior (with zero mean) from the right panel of Fig. 2.3, where two nuclei with charges $\gamma_1 = \gamma_2 = 1$ at

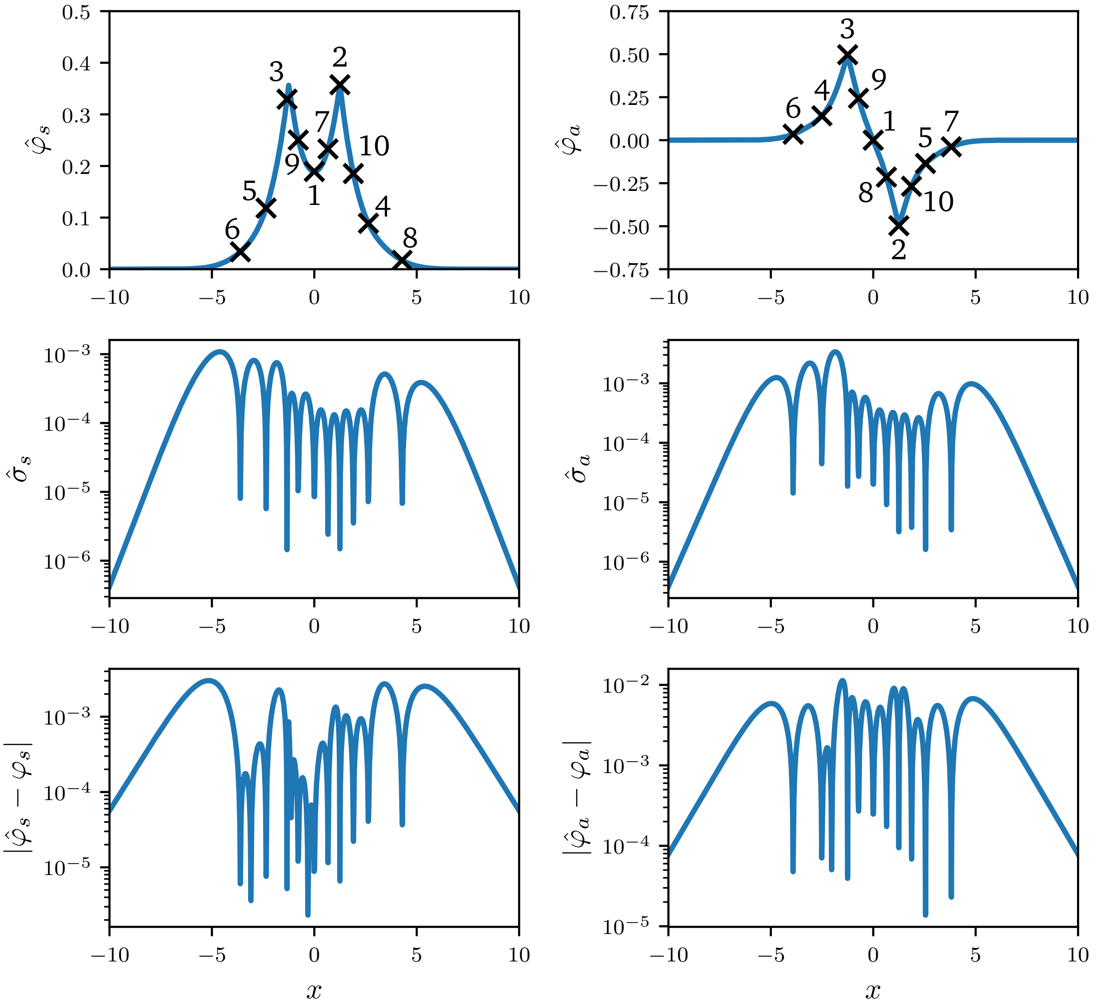

Figure 2.8: AL procedures for the symmetric (**left**) and anti-symmetric (**right**) LCAO eigenstates of the H_2^+ molecular ion are illustrated. To reduce the system to one dimension, the wave function is only considered along the molecular axis. The **top** two panels show the posterior means with the actively learned training sets depicted as crosses, numbered by their order of inclusion into the training set. The **middle** row shows the posterior uncertainties, approaching zero at each training point. The **bottom** two panels show the errors of the regressions.

the positions $\mathbf{R}_{1,2} = \pm 1.25$ were used.

As approximate solutions to the molecular ions lowest eigenstates we use the LINEAR COMBINATION OF ATOMIC ORBITALS (LCAO) ansatz [37] of two hydrogen ground states, $\psi_{1,2}$, at the positions of the two nuclei,

$$\psi_s = c_s(\psi_1 + \psi_2), \tag{2.64}$$

$$\psi_a = c_a(\psi_1 - \psi_2), \tag{2.65}$$

with

$$c_s = \sqrt{\frac{1}{2(1+S)}}, \tag{2.66}$$

$$c_a = \sqrt{\frac{1}{2(1-S)}}, \tag{2.67}$$

$$S = \left(1 + \|\mathbf{R}_1 - \mathbf{R}_2\| + \frac{\|\mathbf{R}_1 - \mathbf{R}_2\|^2}{3}\right) e^{-\|\mathbf{R}_1 - \mathbf{R}_2\|}, \tag{2.68}$$

where the subscripts 's' and 'a' stand for the symmetric and anti-symmetric solutions, respectively. As unit of length the Bohr radius, $a_0 = 1$ a.u., is used. For demonstration purposes we consider only the wave function along the molecular axis, rendering the problem one-dimensional.

Fig. 2.8 illustrates an AL procedure, similarly to Fig. 2.7. The non-stationary properties of the prior, namely the cusps at the nuclei and the exponential decay away from the nuclei, suit all eigenstates of the system. The exponential decay of the GP is crucial for reasonable data acquisition: the actively selected training points are located around the nuclei, where the wave functions exhibit their characteristic structures. No manual boundaries for the learning regions have to be set. For both eigenstate regressions the optimal hyperparameter relating to the principal quantum number approaches $\eta = 1$, which correctly restores the ground states in the LCAO ansatz.

2.6.1 Bayesian optimization

The possibility of actively collecting data in order to efficiently gain information about the true latent function suggests modifications that enable finding specific values of that function, such as extrema, i.e., minima and maxima. [38] Finding extrema of functions falls into the discipline of optimization. Local optimizers, e.g., the gradient descend algorithm, often follow the steepest descent to arrive at the local minimum of a latent function. More difficult to find are global extrema; local optimizers usually get stuck around local extrema. To find global extrema large numbers of function evaluations are usually required in order find all local extrema.

Due to their predictive power while requiring relatively few function evaluations, GPs have turned into a popular tool for global optimization. [39] The procedure of finding global extrema using the posterior mean and variance of a GP was coined BAYESIAN OPTIMIZATION (BO). BO is essentially an AL algorithm that iteratively adds training points. The difference to AL in the previous section is that new training points are added not where the uncertainty is largest, but instead where an ACQUISITION FUNCTION, such as

$$\alpha(\mathbf{x}) = \hat{\mu}(\mathbf{x}) \pm \kappa\hat{\sigma}(\mathbf{x}), \tag{2.69}$$

has a maximum or minimum, respectively. For the first case, above acquisition function is large in regions of high predictive mean, $\hat{\mu}$, and high uncertainty, $\hat{\sigma}$, and its maximum provides sensible candidates for global maxima. High values of the parameter κ lead to more EXPLORATION, i.e., searching in regions of high uncertainty, while small values of κ lead to more EXPLOITATION, i.e., sampling close around predicted maxima. A sensible choice for the value of κ is given in Ref. [40].

In Fig. 2.9 a snapshot of an BO task for maximization is shown. As prior GP for AL the one from the right panel of Fig. 2.3 was used, since it includes the prior knowledge that the latent function, depicted in blue, decays as $\|\mathbf{x}\| \to \infty$. If no prior knowledge about the latent function is at hand, a simple RBF covariance, Eq. (2.12), can be used instead. At each step, the acquisition function, Eq. (2.69), is maximized to locate the next training point, being the potential candidate for a global maximum. The posterior GP after six iterations is shown. The oscillatory character of the latent function causes local optimization procedures to likely get stuck in local maxima, as they usually follow gradients until converged. Instead, BO utilizes the uncertainty to sample unexplored regions where maxima can occur. Hence, Bayesian optimizers never get stuck locally and at the same time use much fewer latent function evaluations, which is particularly important if such evaluations are costly.

One notices that the optimization task has shifted from optimizing the true latent function to maximizing/minimizing an acquisition function, Eq. (2.69). The reason is that optimizing the acquisition function is much less costly, as its evaluations correspond to the evaluations of the posterior GP, Eqs. (2.38) and (2.36), which scales $\mathcal{O}(N)$.

While being very efficient, the application of BO requires no computations of gradients of the latent function. Latter needs not be analytic or possess any specific properties, e.g., being continuous. Both, the training GP and the acquisition function can be tuned to the problem at hand, e.g., to ignore divergences or define preferred regions for exploration. Recently, BO has become a prominent method for designing ANNs as well. [41]

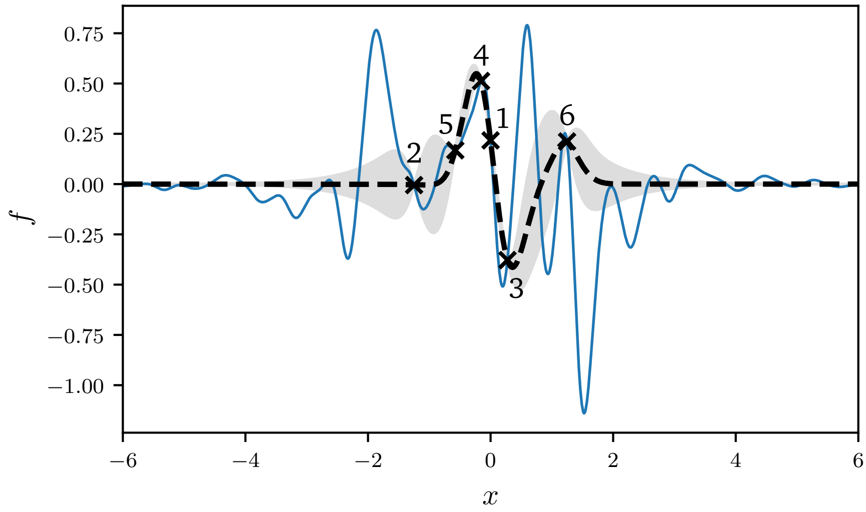

Figure 2.9: BO is illustrated on finding the global maximum of a sample function (blue) generated from a GP similar to that of the right panel of Fig. 2.3. The same GP is used as prior for the BO method. At six iterations the posterior mean (dashed) and standard deviation (gray shades) are shown together with the actively selected training points, numbered by their order of inclusion into the training set.

3 Adiabatic and diabatic dynamics

All states and processes of matter underlie the laws of quantum physics. The dynamical description of non-relativistic particles on the atomic scale is governed by the TIME-DEPENDENT SCHRÖDINGER EQUATION [42], which describes the time-evolution of a quantum system, represented by a wave function, Ψ,

$$i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{x},t) = \left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial\mathbf{x}^2} + V(\mathbf{x},t)\right]\Psi(\mathbf{x},t) \equiv H(\mathbf{x},t)\Psi(\mathbf{x},t). \tag{3.1}$$

The operator H is the HAMILTONIAN of the system – the observable containing all the terms that contribute to its energy. If an atom or molecule is completely isolated from its environment, it can be described by a time-independent Hamiltonian, $H(\mathbf{x})$. The eigenenergies, E_j, and eigenstates, ψ_j, of the Hamiltonian are governed by the TIME-INDEPENDENT SCHRÖDINGER EQUATION (TISE),

$$H(\mathbf{x})\psi_j(\mathbf{x}) = E_j\psi_j(\mathbf{x}). \tag{3.2}$$

Due to the linearity of Eq. (3.1), its general solution for the time-independent case can be written as the superposition of all eigenstates of H,

$$\Psi(\mathbf{x},t) = \sum_j a_j e^{-\frac{i}{\hbar}E_j t}\psi_j(\mathbf{x}). \tag{3.3}$$

Without any perturbation, a system, originally in an eigenstate, would remain forever in that state, i.e., $|\Psi(t)|^2 = |\psi_j|^2\ \forall t$. A change in the external or internal conditions will alter the state of the system. Such a change can be quantified by a, likely time-dependent, parameter variation in the Hamiltonian, $H(\lambda)$, causing its eigenenergies and eigenstates to vary as well: $E_j(\lambda)$, $\psi_j(\lambda)$.

Adiabatic theorem

Throughout a perturbation the evolution of an initial system state starting from an eigenstate depends on the relative time-scales. If we consider an electron inside an atom, the time-scale of its motion around the nucleus is of the order of femtoseconds, $10^{-15}s$. A much slower perturbation of the Hamiltonian will have a smaller effect on the system than a perturbation of similar time-scale. This is stated by the ADIABATIC THEOREM: [43]

Theorem (Adiabatic theorem). *A physical system remains in its instantaneous eigenstate if a given perturbation is acting on it slowly enough and if there is a gap between the eigenvalue and the rest of the Hamiltonian's spectrum.*

Slowly means that the time-scale of the intrinsic dynamics of the system is much smaller than that of the change of the perturbation, i.e., $\frac{\mathrm{d}\lambda}{\mathrm{d}t} \to 0$. An initial system state in an eigenstate,

$\psi_j(\lambda(t_0))$, will adapt to the slow changes of the system and end up in the corresponding eigenstate of the final Hamiltonian, $\psi_j(\lambda(t_1))$, assuming its eigenenergy has not crossed that of another. The initial and final eigenstates will in general not have the same shape, $\psi_j(\lambda(t_0)) \neq \psi_j(\lambda(t_1))$. Such behavior is called ADIABATIC.

For the case of quickly changing conditions, the system state will not be able to adapt to the changes and ceases being an eigenstate. In the limit of infinitely rapid perturbation, $\frac{\mathrm{d}\lambda}{\mathrm{d}t} \to \infty$, the shape of the wave function will remain unaltered, i.e., $|\Psi(t_0)|^2 = |\Psi(t_1)|^2$. Such behavior is called non-adiabatic or DIABATIC.

For intermediate perturbations, $0 < \frac{\mathrm{d}\lambda}{\mathrm{d}t} < \infty$, the system state will in general neither remain unchanged nor an eigenstate. When a measurement of the system energy is conducted the possible measurement outcomes are the eigenenergies of the Hamiltonian with probabilities given as the squared absolute value of the overlap between system state and the corresponding eigenstate: $P(E_j) = \left|\int \psi_j^* \Psi \,\mathrm{d}\mathbf{x}\right|^2$. For simple two-level systems the probability of the transition from one to the other eigenstate, $\int \psi_1^* \psi_0 \,\mathrm{d}\mathbf{x}$, can be approximated with the LANDAU-ZENER formula. [44–47]

Avoided crossings

Continuously changing perturbations of a system, described by a Hamiltonian depending on a parameter, $H(\lambda)$, are reflected in changing eigenenergies and eigenstates of the system Hamiltonian. The top left panel of Fig. 3.1 shows eigenenergies of bound states of an example Hamiltonian as functions of a perturbation. These spectra exhibit various crossings of energy surfaces, as well as avoided crossings, most prominent around $\lambda \approx 0.4$. The details of the underlying physical system, namely two kinetically coupled Morse oscillators, will be discussed in Chapter 5 and shall not be of concern for now.

The fact that energy surfaces cross, i.e., become degenerate, implies that the system must have more than one degree of freedom, since there are no degenerate bound eigenstates in one-dimensional non-singular systems. [48] Consequently, all bound eigenstates exhibit avoided crossings. A proof for this statement is provided in Appendix A.3.

For higher-dimensional systems the VON NEUMANN-WIGNER theorem [49] states that in general eigenenergies can only cross on a manifold of $N_\lambda - 3$ dimensions, where N_λ is the number of continuous real parameters of the Hamiltonian. For instance, a diatomic molecule, having only one parameter, namely the internuclear distance, will in general not exhibit any crossings. A triatomic molecule, having three configurational parameters, will in general exhibit a crossing of energies at isolated points, so-called CONICAL INTERSECTIONS. [50]

For systems that possess symmetries, eigenenergies corresponding to eigenstates of different symmetry can cross on manifolds of $N_\lambda - 1$ dimensions. Thus, the variation of one single parameter will locate intersections of eigenenergies. A derivation of the von Neumann-Wigner theorem is given in Appendix A.4.

The system of which the energy spectrum is depicted in Fig. 3.1 possesses a symmetry. The blue and orange energy surfaces in the top left panel correspond to symmetric and anti-symmetric eigenstates, respectively. While energies belonging to symmetric states can cross those of anti-symmetric states, they avoid all crossings with other symmetric states. I.e., blue and orange lines cross, while blue lines avoid crossing other blue lines. The same goes for orange anti-symmetric states. Considering only symmetric or only anti-symmetric states, the adiabatic theorem states that a system state starting on any energy surface will remain in that particular eigenstate if the perturbation is changing slowly. Hence, we shall denote the individual lines of one symmetry adiabatic manifolds or ADIABATS.

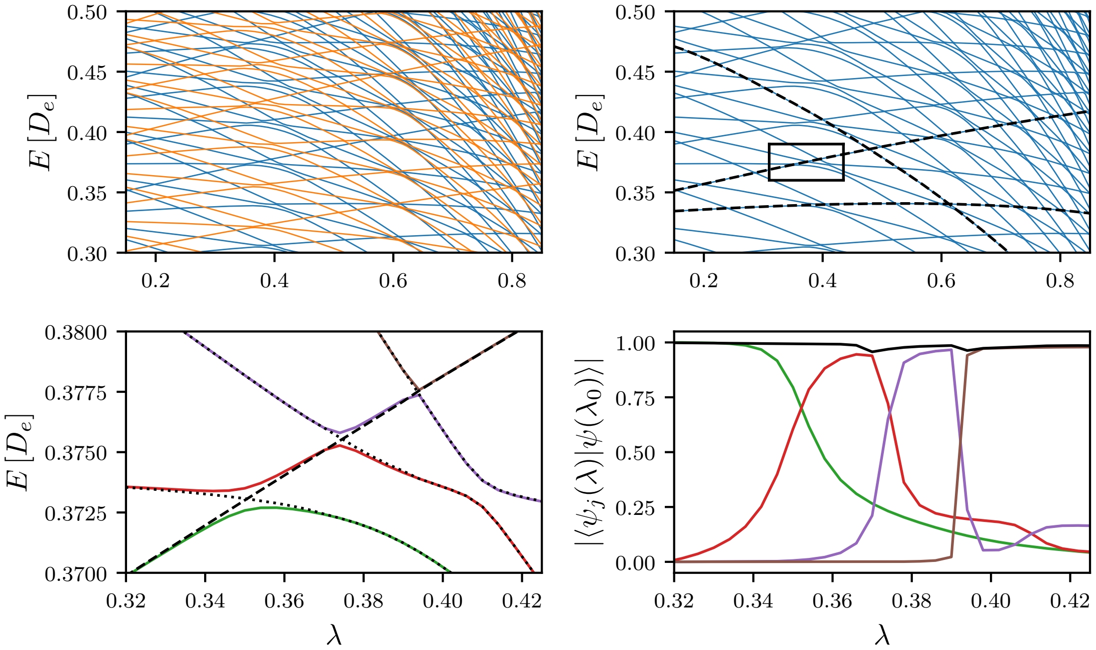

Figure 3.1: The **top left** panel shows the energy spectra of two identical, coupled Morse oscillators with dissociation threshold D_e for a range of the coupling parameter, λ. The blue lines correspond to symmetric and the orange lines to anti-symmetric eigenstates. The **top right** panel shows three diabats (dashed) constructed via GPR from the symmetric energy spectrum. The **bottom left** panel zooms into the black rectangle in the top right panel. The colored lines show the eigenenergies, the dashed line the constructed diabat, and the dotted lines the energies corresponding to the orthogonal states from the diabatic transformation, Eqs. (3.5) – (3.6). The **bottom right** panel shows the auto-correlation functions for each of the eigenstates in the bottom left panel in their associated colors, as well as the auto-correlation of the transformed diabatic state in black. $\psi(\lambda_0)$ is the initial state of the black diabat at $\lambda_0 = 0.32$.

3.1 Finding diabats using GPR

The energy spectra in the top panels of Fig. 3.1 show a rich abundance of avoided crossings, some of which are more prominent and others barely resolved. Human observation of the spectra reveals for each state a pattern which is continued after each avoided crossing by other states of the same symmetry. While such patterns, which occur for numerous spectra of other systems as well, are simply discovered with the naked eye, it is not straightforward to isolate them in an automated fashion using a computer.

Pattern recognition falls into the set of tasks of ML. In Ref. [8] GPR has been successfully applied to extrapolate complicated patterns, proving that GPs are a capable tool for pattern extrapolation. The crucial strength of GPs is, as we shall see later, the interpretability of the hyperparameters in the covariance function. This enables the user to extrapolate from data in a controlled manner.

GPR shall be applied to the problem of finding the correct continuation beyond avoided crossings for any eigenstate in Fig. 3.1, top left. The perturbation parameter represents the input space, $x = \lambda$, while the energy denotes the output values of the GP, $f(x) = E(\lambda)$. Ref. [8] employs a SM covariance function – a combination of various periodic covariances – in order to correctly reproduce complicated patterns in the training data. As no such periodicity is seen in our spectra, we apply the simple RBF covariance function, Eq. (2.12), instead.

Since all states of either symmetry are completely decoupled from states of the other symmetry, we shall only focus on symmetric (blue) states from now on. Starting from a point where the chosen adiabat is well-separated from other states, i.e., away from any avoided crossings, one initially collects points of the same manifold, comprising a first training data set. Inferring from this data set, the posterior GP will supply a smooth continuation of the data. In order to correctly extrapolate, the posterior mean, Eq. (2.38), must pick up the trend of the small set of points. For this to happen, the hyperparameters must be chosen appropriately. Luckily, we know precisely what effect each hyperparameter has on the posterior GP. The crucial hyperparameter for detecting long-range correlations in training data is the length-scale, l. For the RBF covariance, the length-scale corresponds to the widths of the Gaussian basis functions comprising the posterior mean, Eq. (2.39). For extrapolation it must generally be chosen much larger than the nearest neighbor distance in the data set, $l \gg \Delta\lambda$. Observation of our particular spectra in Fig. 3.1, top left, reveals that the length-scale of the long-ranged patterns are much larger than the given range of the perturbation parameter, i.e., $l \gg 1$. Such boundaries for the hyperparameters can be supplied to the optimizer for the MLE method.

To reconstruct the full pattern of the continued eigenstate, we successively add batches of new training points at regions of low uncertainty and whose energies lie closest to the extrapolating GPR. This is done, similarly to the active learning procedure from Sec. 2.6.1, by adding points where an acquisition function exhibits minimal values. The acquisition function to minimize for this procedure is

$$\alpha(\lambda, E_i) = |E_i(\lambda) - \hat{\mu}(\lambda)| + \kappa\hat{\sigma}(\lambda), \tag{3.4}$$

where E_i are the eigenenergies of the spectra, and $\hat{\mu}(\lambda)$ and $\hat{\sigma}(\lambda)$ are the mean and uncertainty of the posterior GP, respectively. The first term of the right hand side ensures that the test point lies close to the extrapolated regression function. The second term only depends on the test point locations and aids to select points with low uncertainties. Its parameter should be chosen rather large compared to the length-scale, $\kappa \gg l$. The acquisition function is evaluated at all points of the spectra and batches of points with lowest $\alpha(\lambda, E_i)$ are repeatedly added to the training set.

The top right panel of Fig. 3.1 shows three specimen of patterns of eigenenergies that were constructed by the presented method. Using the MLE method, the lower boundary for the optimization of the length-scales has been set to $l \geq 5$, while the other hyperparameters, σ_f and σ_n, were optimized freely. The size of each batch of points that is added subsequently to the data set can be chosen large if the length-scale of the pattern is large. The method is robust under variation of the batch size, but it should be kept smaller than the training data set.

Once all points for a full construction of the pattern are collected, a final GPR produces a smooth continuation of the manifold, as shown as the black dashed lines in the top right panel of Fig. 3.1. In order to prevent distortions at the avoided crossings, additional noise has been added at the jumps between the eigenstates by substituting ${\sigma_n}^2\delta_{ij} \rightarrow \sigma_{n,j}^2\delta_{ij}$ in Eq. 2.40.

The question that remains is what physical behavior is represented by such patterns? To answer this, let us consider one black line in Fig. 3.1, top right. Starting on an eigenstate at small $\lambda < 0.2$, it clearly follows the pattern throughout a number of avoided crossings. The bottom left panel zooms into the black rectangle inside the top right panel. Therein, all four states involved in the avoided crossing are shown in different colors. The bottom right panel shows the corresponding auto-correlation functions, i.e., the overlap with the initial eigenstate, for each state in the respective color. A high auto-correlation, i.e., $|\langle\psi_j(\lambda)|\psi(\lambda_0)\rangle| \approx 1$, means that it closely resembles the shape of the initial eigenstate. One can see that each intermediate state consecutively gains a high auto-correlation for a short λ-range, while after the avoided crossing only the final (brown) state shares a high auto-correlation with the initial eigenstate. Evidently, along the emergent pattern that is extracted with our method, we find that $\mathrm{d}_\lambda |\psi(\lambda)\rangle \simeq 0$. For this reason, we shall denote such curves as diabatic manifolds or DIABATS.

3.2 Diabatic transformation

Although the constructed diabatic manifolds do not resemble eigenenergies at the regions of avoided crossings, they represent a likely path for system states under diabatic parameter variation, i.e., for perturbations that happen rather fast. Along such diabats the wave functions carry their invariant characteristics; one could argue that the diabatic spectra are associated with constants of motion of the system, as we shall discuss in the following chapters.

Another case where a diabatic representation is advantageous is when molecules are considered. Computation of the eigenenergies of molecules is regularly done under the application of the BORN-OPPENHEIMER (or adiabatic) approximation leading to adiabatic potential energy surfaces that depend parametrically on the nuclear coordinates. This approximation exploits the fact that the masses of the atoms are much larger than that of the electrons, causing their dynamics to evolve at different timescales. However, in regions of avoided crossings this approximation fails and a diabatic transformation of states is needed. [51,52]

The simplest way of assembling the corresponding diabatic wave functions to each diabatic energy surface is by mixing the lower and upper adiabatic states such that they fulfill some criterion. For molecules, this criterion is usually the diagonal representation of the kinetic energy operator for the nuclear coordinates. Ref. [53] supplies another way of mixing two eigenstates if no particular criterion needs to be met. In our case, our constructed diabatic surface provides a criterion for the diabatic wave function.

A diabatic state, $\psi^d(\lambda)$, can be constructed as a mixture of the corresponding lower and upper adiabatic states, $\psi^a_{0,1}(\lambda)$, whose crossing is avoided. The diabatic state is then given as a superposition of the two adiabatic states. At the same time an orthogonal diabatic state emerges as the complementary superposition. Such a mixing can be written as a unitary rotation around a diabatic mixing angle $\vartheta(\lambda) \in [0, \frac{\pi}{2}]$,

$$\begin{pmatrix} \psi^d_0(\lambda) \\ \psi^d_1(\lambda) \end{pmatrix} = \underbrace{\begin{pmatrix} \cos\vartheta(\lambda) & \sin\vartheta(\lambda) \\ -\sin\vartheta(\lambda) & \cos\vartheta(\lambda) \end{pmatrix}}_{U(\lambda)} \begin{pmatrix} \psi^a_0(\lambda) \\ \psi^a_1(\lambda) \end{pmatrix} . \tag{3.5}$$

The two-level Hamiltonian in this diabatic basis has the non-diagonal form

$$H^d(\lambda) = U \begin{pmatrix} E_0(\lambda) & 0 \\ 0 & E_1(\lambda) \end{pmatrix} U^\top = \begin{pmatrix} \cos^2\vartheta E_0 + \sin^2\vartheta E_1 & \cos\vartheta\sin\vartheta(E_1 - E_0) \\ \cos\vartheta\sin\vartheta(E_1 - E_0) & \sin^2\vartheta E_0 + \cos^2\vartheta E_1 \end{pmatrix} . \tag{3.6}$$

The minimal energy level splitting of $2\left|H^d_{01}\right| = E_1 - E_0$ occurs at $\vartheta = \frac{\pi}{4}$.

Using GPR for the discovery of a diabat, $E^d_{\text{GP}}(\lambda)$, the emerging regression gives the condition for the energy of the corresponding diabatic state, $\psi^d_0(\lambda)$. One can extract a mixing angle by minimizing the energy difference

$$\min_{\vartheta}\left\{ \left| E^d_{\text{GP}} - H^d_{00}(\vartheta) \right|^2 \right\} \tag{3.7}$$

at each λ. Once that mixing angle is extracted both diabatic states can be constructed.

The advantage of our proposed procedure over that of Ref. [53] is the following: our procedure is more general in the sense that the mixing angle is not fixed, but must be supplied by another criterion, in our case the GPR. Thus, the provided GPR enables diabatization over multiple avoided crossings at once, such as in Fig. 3.1. The method in Ref. [53] can only handle pairwise diabatization, since their procedure provides a fixed mixing angle between two states. The special choice for the mixing angle of

$$\vartheta(\lambda) = \arccos\sqrt{\frac{1 + S(\lambda)}{2}} \tag{3.8}$$

leads to the method in Ref. [53], where $S(\lambda)$ is defined in therein.

The bottom left panel of Fig. 3.1 shows the constructed diabats in dashed black. The mixing angle is extracted according to Eq. (3.7), considering the immediate lower and upper adiabatic states at each λ. These two adiabatic states change whenever the diabat crosses an adiabatic surface. The energies of the orthogonal diabatic states are depicted as black dotted lines.

The bottom right panel of Fig. 3.1 shows the auto-correlation with the initial eigenstate of the constructed diabatic state in black. It closely resembles the initial eigenstate but has small dips every time the diabatic energy crosses an adiabatic surface. Such dips are unavoidable if the diabatic state jumps multiple surfaces at one avoided crossing. In fact, these fluctuations of the diabatic wave function display a shortcoming of the assumption that only two states couple at any given point. In reality, multiple states couple whenever they collectively undergo an avoided crossing. The coupling of more than two states involves more than one mixing angle. However, the diabatic surface constitutes only one single condition for choosing the mixing angles, leaving the resulting problem underdetermined.

Although considering only pairwise couplings, the constructed diabatic state carries the properties of the initial eigenstate throughout the whole parameter range. It shall be emphasized that such a state was assembled by solely considering the adiabatic energy spectra. The information of the shape of the eigenstates was not used for the diabatization.

4 Semiclassical diabatization via Gaussian process regression

In the previous chapter we investigated diabatic states and their construction using GPR. There we considered each energy surface individually, which leads to a huge computational overload if the full spectrum shall be transformed. Instead, one would seek a way to compute the diabatic energy spectra of a system directly. Such spectra would ignore any couplings between eigenstates. Rather, each eigenenergy should be determined in an isolated fashion. Such methods for determining eigenenergies of quantum systems bring us back to the very beginning of quantum mechanics.

4.1 EBK quantization

Very early attempts to describe the quantized energy spectra of atoms relied on the heuristic BOHR-SOMMERFELD quantization condition, [54]

$$\oint p_i \,\mathrm{d}q_i = \int_0^{T_i} p_i \frac{\mathrm{d}q_i}{\mathrm{d}t} \,\mathrm{d}t = 2\pi\hbar \cdot n_i \ , \quad i = 1, \ldots, D, \tag{4.1}$$

which quantizes the allowed path a bound particle may take in D degrees of freedom if the motion is SEPARATE in its coordinates, i.e., the motion along each coordinate is independent from another. Here, p_i is the momentum canonically conjugate to the coordinate q_i and n_i serves as a quantum number assuming integer values. The integral is taken along a full period, T_i, in each degree of freedom.

Albert Einstein realized very early that the condition of separate variables, which depends on the choice of the coordinate system, causes the quantization rules to have no invariant meaning and therefore proposed a coordinate-invariant formulation. [55,56] He pointed out that classical trajectories, under certain conditions which will be specified later, are confined to a D-dimensional torus in phase space. Such a torus offers precisely D topologically distinct ways of closing loops on its surface, C_i, along which the phase integral $\oint_{C_i} \mathbf{p} \cdot \mathrm{d}\mathbf{q}$ is invariant for different but topologically similar choices of C_i. Since these integrals depend on the energy but are coordinate-independent, Einstein proposed the invariant modification for the quantization rules,

$$\oint_{C_i} \mathbf{p} \cdot \mathrm{d}\mathbf{q} = \oint_{C_i} \sum_{d=1}^{D} p_d \,\mathrm{d}q_d = 2\pi\hbar \cdot n_i \ , \quad i = 1, \ldots, D. \tag{4.2}$$

Note that the paths, C_i, are generally not the path of the trajectory but any closed path on the torus.

In one dimension the Bohr-Sommerfeld quantization condition is derived from the TISE by using the semiclassical WENTZEL-KRAMERS-BRILLOUIN (WKB) approximation [57–59], while Einsteins quantization rule corresponds to its multidimensional generalization. However, the WKB method yields a small correction to the quantization rule, namely

$$\oint_{C_i} \mathbf{p} \cdot \mathrm{d}\mathbf{q} = 2\pi\hbar \left(n_i + \frac{\nu_i}{4} \right) \ , \quad i = 1, \ldots, D, \tag{4.3}$$

where ν_i is the MASLOV index – an integer that counts the encounters of CAUSTICS along a path. This modified version of Einsteins quantization rule was derived 40 years later by Keller [60] and was coined EINSTEIN-BRILLOUIN-KELLER (EBK) quantization by Percival. [61] This quantization condition constitutes a semiclassical approximation to the exact solutions of the TISE, i.e., they emerge in the limit of $\hbar \to 0$.

The EBK quantization condition serves as a tool to identify eigenenergies of quantum mechanical systems by locating classical trajectories which fulfill this condition, namely the n_i being integers for every topologically distinct loop around their phase space tori. The energy of such a trajectory provides the eigenenergy of the quantum system corresponding to the quantum numbers, i.e., $E(n_1, \ldots, n_D)$. The major advantage of such a procedure is the numerical simplicity of computing classical trajectories, which can be done parallelly and independently in large numbers. In addition, methods for the generation of wave functions from such "eigentrajectories" (which fulfill the EBK quantization condition) have been brought forward, e.g., in Ref. [62].

While the EBK method gives the impression that for any given system the eigeneneries can be computed using classical trajectories only, there are a number of downsides to this semiclassical method. Many physical systems show classical trajectories whose dynamics is IRREGULAR, i.e., *not* confined to a D-torus in phase space. Generic systems may experience both regular and irregular (or chaotic) classical dynamics. [63] Unfortunately, the EBK quantization can only be applied to trajectories of the regular type, although attempts have been made to push this method into the irregular regime. [64] Semiclassical treatment of quantum systems with chaotic behavior was initialized by Martin Gutzwiller. [65, 66] Instead of considering individual states, he derived an equation for the density of states of a chaotic system from the sole knowledge of closed periodic trajectories, which is now referred to as the GUTZWILLER TRACE FORMULA. For the case of purely regular systems, both procedures for the calculation of energy spectra are directly related. [67, 68]

The property of regularity or irregularity/chaoticity of a Hamiltonian system is determined by its number of constants of motion. [69] If the number of independent, Poisson-commuting constants of motion is equal or greater than the number of degrees of freedom, the classical dynamics is regular. Such systems are called INTEGRABLE. If there are fewer Poisson-commuting constants of motion than degrees of freedom, the motion is irregular. Such systems are called NONINTEGRABLE. For autonomous (or time-independent) systems the energy is conserved and therefore constitutes a constant of motion. For this reason, there is no irregularity in one-dimensional autonomous systems.

Integrable systems – including separable systems as a special subgroup – are much rarer than nonintegrable systems. However, the KOLMOGOROV-ARNOL'D-MOSER (KAM) theorem [70–73] implies that for integrable systems under perturbation regular dynamics often partly remains. In such regular regions the EBK method can still be applied, even though the system is nonintegrable.

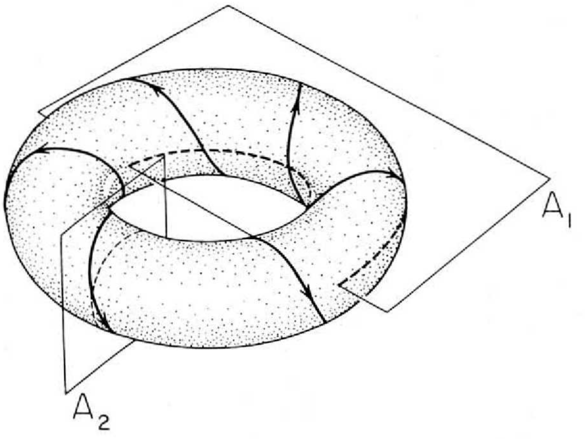

Figure 4.1: An illustration of a torus with a trajectory winding around its surface. The two areas, A_1 and A_2, show two topologically distinct ways of cutting through the torus. The figure is taken from Ref. [74].

4.2 Quasiperiodic and periodic trajectories

A trajectory (or ORBIT) of a system of D degrees of freedom is regular if it is confined to a D-dimensional torus in $2D$-dimensional phase space. A two-dimensional torus is illustrated in Fig. 4.1. Note that this is merely an illustration with the reduced phase space dimensionality of three; for two degrees of freedom the torus would be a two-dimensional submanifold in *four*-dimensional phase space. A regular trajectory will almost always densely fill the surface of that torus when winding around it an infinite number of times. One denotes such trajectories as QUASIPERIODIC. However, there are special trajectories that are confined to a torus but instead of densely filling its surface, they wind around it periodically, i.e., they repeat to follow one and the same path that draws a single line looping around the torus. The arrows winding around the torus in Fig. 4.1 represent a path such a trajectory may take. These trajectories are called PERIODIC and they fill a subspace in phase space of measure zero and are therefore hard to find.

A torus filled by a quasiperiodic orbit can be cut in D topologically distinct ways, each of which providing an invariant action variable given as the phase integral along the edge of the slice,

$$S_i = \oint_{C_i} \mathbf{p} \cdot \mathrm{d}\mathbf{q} \ , \quad i = 1, \ldots, D. \tag{4.4}$$

The two topologically distinct ways of slicing the torus in Fig. 4.1 are suggested by the areas A_1 and A_2. The possibilities of slicing are not unique. The loops created by the slicing are candidates for the paths, C_i, along which the phase integral, Eq. (4.4), is invariant under different but topologically similar ways of slicing.

The left panel of Fig. 4.2 shows a projection onto position space of a quasiperiodic trajectory for $D = 2$. One can see that this trajectory fills only a portion of the energetically allowed region. The four "edges" formed by the trajectory locate the CAUSTICS, which can be thought of as higher-dimensional generalizations of the one-dimensional turning points, although there the momentum does in general not vanish. A periodic trajectory would also be confined by caustics, but it would not densely fill the area. Instead, it would close on itself after a number of periods, yielding a single line only.

Irregular/chaotic trajectories are not confined to a torus in phase space and their projections onto position space would not show clear edges but fill the energetically allowed region almost entirely.

4.3 Poincaré surfaces of section

One way of investigating the properties of trajectories is the POINCARÉ SURFACES OF SECTION (SOS) method. [75] In this section the use of the SOS method to extract action variables [76] will be presented. A Poincaré SOS is a $(2D-1)$-dimensional representation of a trajectory moving in a $2D$-dimensional phase space. It can be constructed by defining a plane in phase space, often by choosing a fixed value for one of the $2D$ phase space coordinates, e.g., $q_2 = 0$. Every time a trajectory pierces through the predefined plane in one direction, e.g., $p_2 > 0$, all values of phase space coordinates are registered. The plane can be freely positioned, although the velocity vector should not be tangent to the section at any point. [77]

The method is illustrated in Fig. 4.2. The two planes are chosen at the constant values of $q_2 = 0$ and $q_1 = 0$ and are drawn as blue and orange dashed lines, respectively. Every time the trajectory penetrates the blue plane in negative direction, i.e., from positive to negative values of q_2, the position on that plane, q_1, and the tangential component of the momentum, p_1, are recorded as a blue point in the right panel. A linear interpolation scheme for the positions on the SOS is shown in Appendix A.5. After many cycles the scattered points will supply a SOS plot such as the ones in the right panel of Fig. 4.2. The values of the normal components of the momenta, $p_2 < 0$, are implicitly known from the values of q_1, p_1 and $q_2 = 0$, given the constant energy E, and are therefore omitted in all SOS figures.

A quasiperiodic trajectory will yield a pattern of scattered points on the SOS that form one or more narrow lines, as shown in the right panel. A periodic trajectory will create finitely many isolated points on a SOS and an irregular/chaotic trajectory will fill an area.

If a SOS is chosen such that it cuts once through two opposing caustics of a quasiperiodic trajectory, such as in Fig. 4.2, left, then the Poincaré SOS depicts a single closed loop, such as the ones in Fig. 4.2, right. For a trajectory in two degrees of freedom, there are two ways of cutting in such manner. These two cuts correspond to the two topologically distinct torus cuts in Fig. 4.1. The areas inside the loops in the right panel of Fig. 4.2 are the actions in Eq. (4.4) and can be computed using the "shoelace" formula for the quadrature of polygon areas. For this example the SOS are located at $q_2 = 0$ and $q_1 = 0$ and the two actions then yield $J_1 = \oint p_1(q_1)\,\mathrm{d}q_1$ and $J_2 = \oint p_2(q_2)\,\mathrm{d}q_2$. This procedure can only be applied to quasiperiodic trajectories, since periodic and irregular trajectories do not yield closed loops on their SOS.

While the Poincaré SOS method for the extraction of all D action variables is a straightforward and robust procedure, it requires the correct positioning of D planes in phase space. For non-resonant separable and nearly separable systems a quasiperiodic trajectory can be represented as a "box" in position space, such as the one in Fig. 4.2, left. For these box-like trajectories it is relatively easy to correctly place the section planes. For resonant quasiperiodic trajectories, which will be discussed in the next chapter, a full set of D section planes is hard to find. However, it is always possible to replace one of the action variables of the torus cuts with the action along the trajectory for one or more periods and closing both

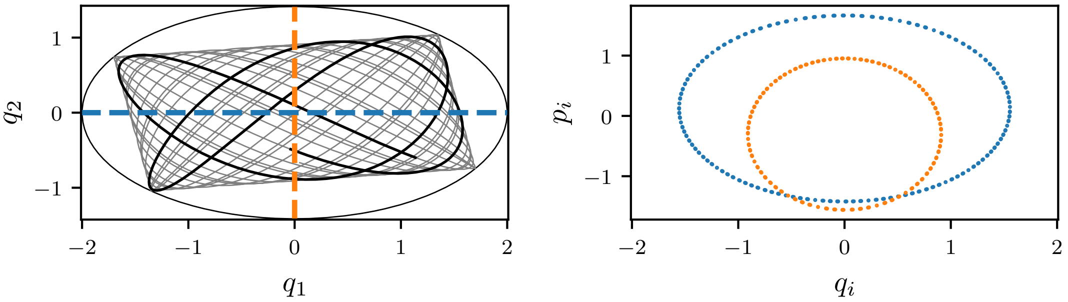

Figure 4.2: Left: A quasiperiodic trajectory is shown in gray for a limited number of cycles. The thick black line follows the trajectory for roughly one loop around its phase space torus and corresponds to the projection of the arrows in Fig. 4.1 onto position space. The boundary of the classically allowed region, $V(q_1, q_2) = E$, is depicted by the elliptical line. **Right:** The two Poincaré SOS corresponding to the two cuts (dashed lines) in the left panel are depicted in the respective colors. Each scatter point represents the position at which the trajectory pierces through the dashed line in the left panel in one direction and the corresponding momentum tangential to the dashed line.

ends along the torus. This method was termed "TRAJECTORY-CLOSE" method and it has been used to apply the EBK quantization condition to resonant trajectories in two degrees of freedom. [78]

For a quasiperiodic trajectory with an $\alpha : \beta$ resonance, the EBK quantization condition for a closed trajectory becomes

$$J_{\text{close}} = \int_{\mathbf{q}_{\text{ini}}}^{\mathbf{q}_{\text{fin}}} \mathbf{p} \cdot \mathrm{d}\mathbf{q} + \Delta J = \alpha \cdot 2\pi\hbar \left(n_1 + \tfrac{1}{2}\right) + \beta \cdot 2\pi\hbar \left(n_2 + \tfrac{1}{2}\right), \tag{4.5}$$

where ΔJ is the action coming from closing the two endpoints, $\mathbf{q}_{\text{ini}}$ and $\mathbf{q}_{\text{fin}}$, along the torus, α is the number of cycles along one coordinate, e.g., q_1, and β the number of cycles along the other coordinate, e.g., q_2. The Maslov indices are set to $\nu = 2$ since in each cycle the trajectory touches two pairs of opposing caustics. Eq. (4.5) can be further simplified,

$$J_{\text{close}} = 2\pi\hbar \left(\underbrace{\alpha n_1 + \beta n_2}_{m} + \frac{\alpha + \beta}{2} \right) = 2\pi\hbar \left(m + \frac{\alpha + \beta}{2} \right). \tag{4.6}$$

If one chooses $J_1 = 2\pi\hbar \left(n_1 + \frac{1}{2}\right)$ as one quantization condition, m can only take up values of $m = \alpha n_1 + \beta n_2$ where n_2 are non-negative integers. Therefore one must keep in mind that m cannot assume all integer values.

The Poincaré SOS method has been extended to three degrees of freedom. [79,80] However, generating proper torus cuts becomes more involved in higher dimensions. More sophisticated methods for the computation of action variables of quasiperiodic orbits involve the representation of the coordinates and momenta as Fourier series. [81–83] The action variables are then computed using the Fourier coefficients obtained via fast Fourier transform. Clearly, the advantage of this method is the liberation from the construction of SOS and therefore allows to treat systems with more degrees of freedom, $D \geq 3$. In any case, however, the runtime of each trajectory necessary for convergence grows exponentially with the dimensionality.

4.4 EBK quantization via GPR

Once the action variables of a quasiperiodic trajectory are extracted, one can check whether they fulfill the EBK quantization condition, i.e., all quantum numbers, n_i, in Eq. (4.3) being integers. For arbitrary quasiperiodic trajectories, however, the values of the n_i will in general not be integers but real numbers. In order to find all trajectories that can be quantized one would have to scan a multidimensional range of initial conditions and conduct a root-search for every combination of quantum numbers, requiring the propagation of prohibitively many trajectories.

To circumvent root-searching each trajectory individually, Ref. [84] suggests a method that applies interpolation between the action variables of arbitrary trajectories to points where the EBK condition is satisfied. The energies at those points then approximate the semiclassical eigenenergies of the Hamiltonian. Inter- and extrapolation compose the general task of regression. While Ref. [84] employs primitive interpolation methods limited to two or three dimensions, modern regression methods, such as GPR, are capable of constructing high-dimensional manifolds given relatively few training data. Additionally, GPR allows for pattern extrapolation in a controlled matter. For these reasons, we apply GPR to the task of finding semiclassical eigenenergies. GP models will be trained by sets of action variables of various quasiperiodic trajectories which need not fulfill any quantization condition. On top of the D actions, perturbation parameters of the Hamiltonian, λ_j, can be included into the input dimensions of the GP, yielding energy spectra as functions of the perturbations, i.e., $E_i(\lambda_1, \lambda_2, \dots)$. The resulting regression provides accurate interpolation to extract semiclassical eigenenergies but also enables extrapolation into regions where no quasiperiodic orbits are given.

4.5 Semiclassical diabatization in one dimension

Our goal is the construction of diabatic energy spectra, i.e., energy manifolds, $E_j(\lambda)$, that cross at adiabatically avoided crossings when the perturbation parameter, λ, is varied. As we have learned, primitive semiclassical eigenenergies constructed from single trajectories that fulfill the EBK quantization condition, Eq. (4.3), are extracted in isolated fashion and one expects coupling effects between individual states not to take place. [85]

The consequence of this semiclassical treatment shall be portrayed in a comparison of the EBK spectra of a one-dimensional bound potential with the exact solutions of the TISE. As a demonstrative potential we choose a concatenation of two harmonic wells with varying width ratio as the perturbation, $\lambda = \omega_2/\omega_1$. Fig. 4.3 compares the eigenstates given by an exact solution of the TISE in the left column to the semiclassically quantized trajectories in the right column. The crucial difference between classical trajectories and quantum wave functions is the ability of the latter to TUNNEL through the potential barrier. A primitive classical trajectory is confined within its energetic boundaries. In order for two states in separate wells to couple, tunneling is necessary. The bottom two panels show the energy spectra under variation of the width of the right well, ω_2, while fixing the width of the left well. The true eigenenergies in the bottom left panel given by the TISE avoid all crossings.

This is to be expected, since there are no degeneracies in one-dimensional regular bound systems, as shown in Appendix A.3. For the semiclassical case in the bottom right panel no crossings are avoided. The semiclassical energies therefore correspond to the diabatic case discussed in Chapter 3.

We conclude that the ability of tunneling is crucial for classically separated states to interact. In one dimension tunneling always occurs through potential barriers. In two or more dimensions a different kind of tunneling comes about, which takes place between classically disconnected regions in phase space that are not necessarily separated by any potential barriers. This kind of tunneling is referred to as DYNAMICAL tunneling. [86] In order to treat tunneling effects semiclassically, complex-valued phase spaces have been employed, allowing trajectories to connect otherwise classically disconnected regions. [87] Such treatment yields a correction to the primitive semiclassical eigenenergies and can lead to splittings between degenerate energy levels. For our purpose, however, primitive, real-valued trajectories do the trick, providing diabatic spectra as requested. In the remainder of this work, we will employ GPR for the semiclassical EBK quantization method to construct complete diabatic spectra for Hamiltonian systems of two degrees of freedom plus a perturbation parameter.

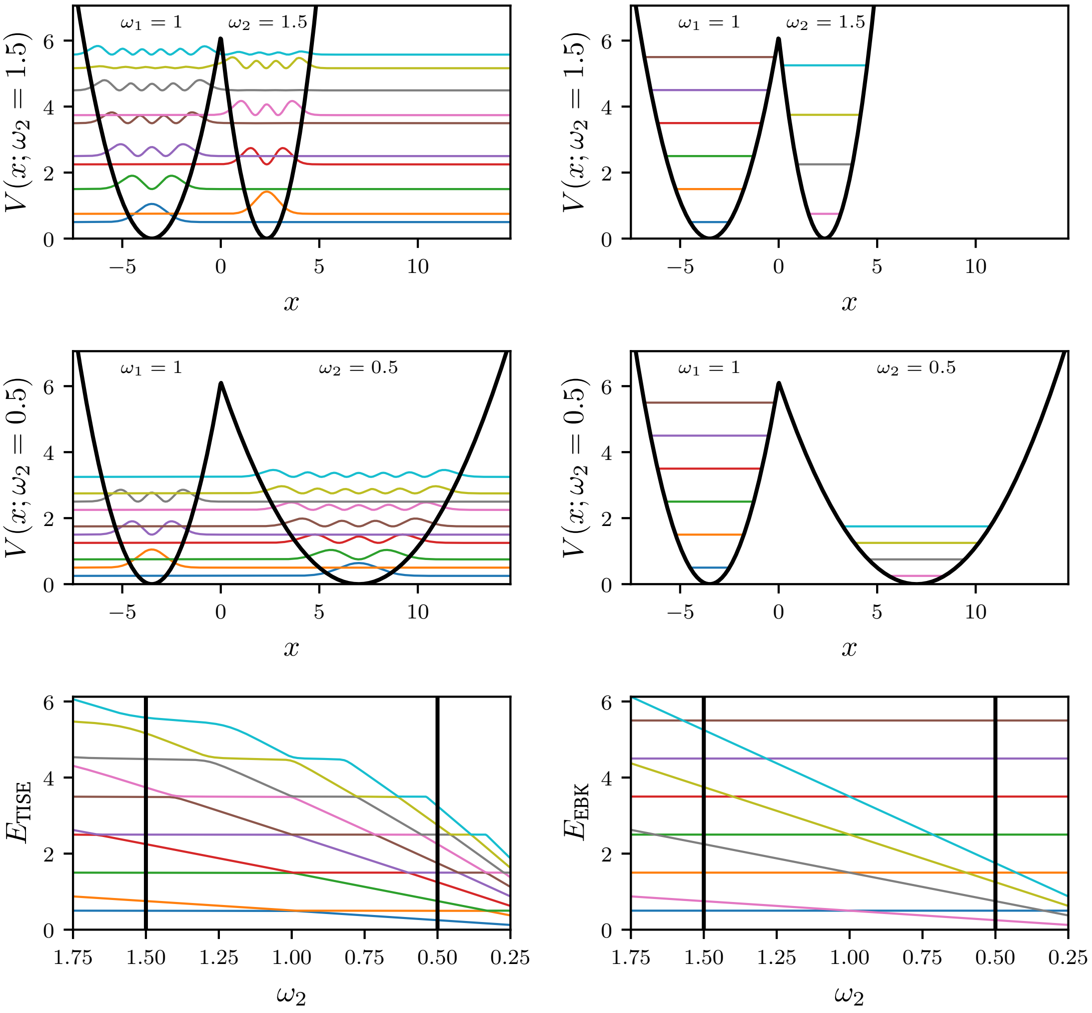

Figure 4.3: A visualization of the differences between the exact eigenstates given by the TISE (**left**) and the semiclassical EBK eigenstates (**right**) for a one-dimensional potential of two concatenated harmonic wells. The width ratio of the two wells is $\omega_2/\omega_1 = 1.5$ in the **top row** and $\omega_2/\omega_1 = 0.5$ in the **middle row**. A selection of eigenstates is drawn in colored lines in the top and middle left panels, whereas the right panels show the energies of the periodic orbits within their classically accessible range. The **bottom row** shows the energy spectra as functions of the width ω_2 in the color code of the eigenstates. The black vertical lines locate the ω_2-values used in the top and middle rows.

5 Semiclassical diabatization of a triatomic molecule

In this chapter we shall demonstrate the proposed method for semiclassical diabatization using GPR on the vibrational dynamics of a linearly aligned triatomic molecule, as depicted in Fig. 5.1, left. This system with two degrees of freedom can be described by two coupled MORSE oscillators. [88] The one-dimensional Morse potential is given by [89]

$$V_{\mathrm{Mor}}(x) = D_e \left(1 - e^{-ax}\right)^2 . \tag{5.1}$$

An example of the Morse potential is shown in the right panel of Fig. 5.1. The values of a and D_e define the stiffness and the dissociation energy as $x \to \infty$, respectively. The coordinate x represents the disposition of the interatomic distance from the equilibrium at $x = 0$. For our investigations we consider the special case of a symmetric triatomic molecule, i.e., the two outer atoms being identical. The Hamiltonian for the vibrational dynamics can then be represented by two identical Morse oscillators with a kinetic coupling term,

$$H_{\mathrm{Mor}}(\lambda) = \frac{p_1^2}{2\mu_r} + \frac{p_2^2}{2\mu_r} + V_{\mathrm{Mor}}(q_1) + V_{\mathrm{Mor}}(q_2) - \lambda p_1 p_2. \tag{5.2}$$

The derivation for this Hamiltonian is shown in Appendix B. The reduced mass, $\mu_r = \frac{mM}{m+M}$, shall be set to $\mu_r = 1$. The mass of the outer atoms is m and the mass of the center atom is M. The coupling/perturbation parameter is the inverse of the mass of the center atom, $\lambda = 1/M$. Due to the reduced mass being set to one, the perturbation parameter takes up values in the range $0 \leq \lambda \leq 1$. [90] The value of $\lambda = 0$ represents the case of a center atom of infinite mass, decoupling both vibrations completely and leaving $\mu_r = m$.

The opposite case of $\lambda = 1$ corresponds to outer atoms of infinite masses and $\mu_r = M$. This renders the two degrees of freedom linearly dependent,

$$H_{\mathrm{Mor}}(\lambda = 1) = \tfrac{1}{2} \underbrace{\left(p_1^2 + p_2^2 - 2p_1p_2\right)}_{(p_1-p_2)^2} + V_{\mathrm{Mor}}(q_1) + V_{\mathrm{Mor}}(q_2). \tag{5.3}$$

As a consequence, the dynamics becomes one-dimensional. The sole motion happens along the asymmetric stretch, i.e., the center atom oscillates between the two outer atoms which are static due to their infinite masses.

Appendix B.1 demonstrates a rotation of coordinates in which the coupling term, $\sim p_1p_2$, vanishes. In that coordinate system the coupling comes solely from the potential shape. The two corresponding coordinates are the symmetric, $q_s \sim q_1 + q_2$, and asymmetric stretches, $q_a \sim q_2 - q_1$. However, the coordinates of Eq. (5.2) are convenient due to the separate form of the Hamiltonian at $\lambda = 0$.

Being a simple model for triatomic molecules, the two coupled Morse oscillators have been extensively studied in the literature. A summary of early investigations is given in Ref. [91].

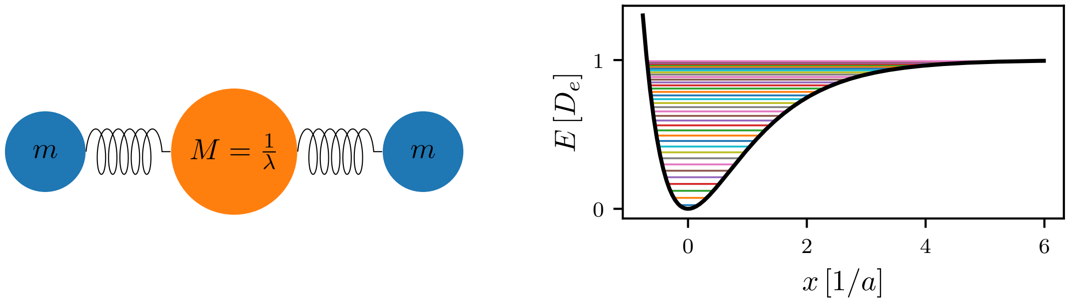

Figure 5.1: Left: An illustration of a linear, symmetric, triatomic molecule with two vibrational degrees of freedom. **Right:** The Morse potential (black) with a number of bound eigenenergies (colored). The parameters were set to $D_e = 32$ and $a = 1/5$, while using units of $\hbar = \mu_r = 1$.

5.1 Solving the TISE

The TISE for the Hamiltonian (5.2) can be efficiently solved if it is represented in a basis of bound, orthogonal Morse oscillator eigenstates, [92]

$$\psi_n(x) = N_n z^{k-n-\frac{1}{2}} e^{-\frac{1}{2}z} L_n^{(2k-2n-1)}(z), \tag{5.4}$$

with eigenenergies

$$e_n = \hbar\omega\left(n + \frac{1}{2}\right) - \frac{\hbar^2\omega^2}{4D_e}\left(n + \frac{1}{2}\right)^2, \tag{5.5}$$

where

$$z(x) = 2ke^{-ax} \quad (0 < z < \infty), \tag{5.6}$$

$$\omega = a\sqrt{\frac{2D_e}{\mu_r}}, \tag{5.7}$$

$$k = \frac{\sqrt{2\mu_r D_e}}{a\hbar}, \tag{5.8}$$

$$n = 0, 1, \ldots, \left[k - \frac{1}{2}\right], \tag{5.9}$$

with $[\xi]$ denoting the maximum integer quantum number $\leq \xi$ and $L_n^{(\alpha)}(z)$ are the associated (or generalized) Laguerre polynomials. The normalization factor is

$$N_n = \sqrt{\frac{a(2k - 2n - 1)n!}{\Gamma(2k - n)}}, \tag{5.10}$$

where $\Gamma(\cdot)$ denotes the Γ-function. Appendix B derives the matrix elements of Hamiltonian (5.2) in the basis of bound Morse oscillator eigenstates. The only non-trivial term is the coupling term, involving integrals of the form $\int \psi_m \frac{\mathrm{d}}{\mathrm{d}x} \psi_n \,\mathrm{d}x$. The results for the matrix elements are

$$\langle m_1, m_2| H_{\mathrm{Mor}} |n_1, n_2\rangle = \delta_{m_1 n_1}\delta_{m_2 n_2}(e_{n_1} + e_{n_2}) + \lambda\hbar^2 \langle m_1| \frac{\mathrm{d}}{\mathrm{d}x_1} |n_1\rangle \langle m_2| \frac{\mathrm{d}}{\mathrm{d}x_2} |n_2\rangle\,, \quad (5.11)$$

where

$$\langle m| \frac{\mathrm{d}}{\mathrm{d}x} |n\rangle = \int \psi_m(x) \frac{\mathrm{d}}{\mathrm{d}x} \psi_n(x)\,\mathrm{d}x = \begin{cases} N_m N_n (-1)^{m-n} \frac{\Gamma(2k-m)}{2n!} & \text{if } m > n, \\ -N_m N_n (-1)^{n-m} \frac{\Gamma(2k-n)}{2m!} & \text{if } m < n, \\ 0 & \text{if } m = n, \end{cases} \quad (5.12)$$

satisfying the property $\langle m| \frac{\mathrm{d}}{\mathrm{d}x} |n\rangle = - \langle n| \frac{\mathrm{d}}{\mathrm{d}x} |m\rangle$.

If the Morse parameters, D_e and a, are chosen such that k gets large, the number of bound states increases and the Γ-function in Eq. (5.12) can no longer be evaluated, as is quickly diverges. To circumvent this problem, we choose Morse parameters such that k becomes an integer. This allows us to rewrite the Γ-function as

$$\Gamma(2k - n) = (2k - n - 1)! \quad (5.13)$$

and lets us to treat systems with more eigenstates, since we will be dealing with a fraction of integers in Eq. (5.12). Throughout this chapter we set $\hbar = \mu_r = 1$ and choose $D_e = 32$ and $a = 1/5$, yielding $k = 40$.

Still, the basis of bound Morse eigenstates comes with one drawback, namely the fact that the size of the basis is limited by the value of k, see Eq. (5.9). In order to investigate convergence we choose, in addition, to represent the system Hamiltonian on a finite grid, which provides an alternative basis of arbitrary size. This method is also known as the FINITE DIFFERENCES method [93] and yields for the kinetic energy terms of the Hamiltonian sparse, off-diagonal matrices and for the potential terms diagonal matrices. Using such a basis one finds that with increasing energy towards the dissociation energy, D_e, eigenstates can no longer be properly converged using the finite Morse basis, as one would trivially expect. Additionally, with increasing perturbation parameter, λ, already some low-energetic states cannot be converged using the finite Morse basis. Such states correspond to eigenfunctions, $\psi(q_1, q_2)$, with large numbers of nodes along the symmetric stretch, $q_s \sim q_1 + q_2$. The failure in the convergence is caused by the fact that strongly coupled systems with $\lambda \to 1$ become more and more linearly dependent. At $\lambda = 1$ the dynamics occurs along the asymmetric stretch only, leaving only one degree of freedom. All eigenfunctions of the system are therefore compressed as λ grows and become more narrow in their symmetric spread, Δq_s. Eigenstates with a large number of nodes along q_s quickly lose energy and at the same time strongly oscillate due to the compression along q_s. Such strong oscillations require highly oscillating basis functions, which the finite Morse basis no longer includes. Instead, a grid basis with a large number of grid points can be used to generate a spectrum of desired convergence. We have used 10^6 grid points, but stopped converging the bound spectra in the range of $\lambda \leq 0.9$ at a relative error of 10^{-3} for a gain in computing speed.

At $\lambda = 1$ all states are fully “flattened” to one-dimensional wave functions and all nodes along the symmetric stretch have vanished. The former degree of freedom, q_s, i.e. the distance between the two infinitely heavy outer atoms, takes the role of a parameter of the Hamiltonian: $H_{\mathrm{Mor}}(q_a; \lambda = 1, q_s)$.

5.1.1 Symmetric and anti-symmetric eigenstates

For a Hamiltonian which is invariant under exchange of the coordinates, $q_1 \leftrightarrow q_2$, all eigenfunctions are either symmetric or anti-symmetric under such exchange. This is exactly our case for the two identical Morse oscillators with the coupling term $\sim p_1 p_2$. In order to extract all (anti-)symmetric eigenfunctions, one can (anti-)symmetrize the basis, in which the Hamiltonian matrix is expressed,

$$|n_1, n_2\rangle \to \left|n_1, n_2^{\pm}\right\rangle := \begin{cases} \frac{1}{\sqrt{2}} \left(|n_1, n_2\rangle \pm |n_2, n_1\rangle\right) & \text{if } n_1 \neq n_2, \\ |n_1, n_2\rangle & \text{if } n_1 = n_2 \text{ and symmetric.} \end{cases} \tag{5.14}$$

Here the '$+$' represents symmetric and the '$-$'-symbol anti-symmetric states. For the matrix elements we get for symmetric states

$$\begin{aligned} &\left\langle m_1, m_2^{+}\right| H \left|n_1, n_2^{+}\right\rangle \\ &= \begin{cases} \langle m_1, m_2| H |n_1, n_2\rangle & \text{if } m_1 = m_2 \text{ and } n_1 = n_2, \\ \frac{1}{\sqrt{2}}\big(\langle m_1, m_2| H |n_1, n_2\rangle + \langle m_1, m_2| H |n_2, n_1\rangle \big) & \text{if } m_1 = m_2 \text{ or } n_1 = n_2, \\ \langle m_1, m_2| H |n_1, n_2\rangle + \langle m_1, m_2| H |n_2, n_1\rangle & \text{else,} \end{cases} \end{aligned} \tag{5.15}$$

and for anti-symmetric states

$$\left\langle m_1, m_2^{-}\right| H \left|n_1, n_2^{-}\right\rangle = \langle m_1, m_2| H |n_1, n_2\rangle - \langle m_1, m_2| H |n_2, n_1\rangle \tag{5.16}$$

if $m_1 \neq m_2$ and $n_1 \neq n_2$. Here we have used the property

$$\langle m_1, m_2| H |n_1, n_2\rangle = \langle m_2, m_1| H |n_2, n_1\rangle \tag{5.17}$$

following from the symmetry of our Hamiltonian.

In order to extract the symmetric and anti-symmetric eigenstates in a grid basis, we make use of the rotation from Appendix B.1. In this coordinate system, the symmetric and anti-symmetric states correspond to the property $\psi(x, y) = \pm\psi(x, -y)$, respectively. The (anti-) symmetrized basis is then given by

$$\left\langle x, y \middle| x_j, y_j^{\pm} \right\rangle = \langle x | x_j\rangle \left\langle y \middle| y_j^{\pm}\right\rangle \tag{5.18}$$

$$= \delta(x - x_j)\, \frac{1}{\sqrt{2}} \left[\delta(y - y_j) \pm \delta(y + y_j)\right] \ , \quad y_j > 0. \tag{5.19}$$

The resulting spectrum is shown in Fig. 3.1 and Fig. 5.5, left, for $D_e = 32$ and $a = 1/5$ using $\mu_r = \hbar = 1$.

The discussed symmetry of above Hamiltonian appears whenever identical particles are considered. Then, the symmetric states correspond to bosonic particles and the anti-symmetric states correspond to fermionic particles. Chapter 3 discussed the fact that for Hamiltonians with symmetries, the eigenenergies do not cross the energies of other eigenstates of the same symmetry when varying a perturbation parameter. On the other hand, the energies of eigenstates with different symmetries do cross.

5.2 Classical dynamics

The classical dynamics of two coupled Morse oscillators has been extensively studied. [94–101] A classical trajectory is numerically propagated by solving Hamilton's equation of motion,

$$\frac{\mathrm{d}\mathbf{q}}{\mathrm{d}t} = \frac{\partial H}{\partial \mathbf{p}}, \qquad \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = -\frac{\partial H}{\partial \mathbf{q}}. \tag{5.20}$$

These equations can be numerically integrated using the LEAP FROG method, [37]

$$\mathbf{q}^{i+\frac{1}{2}} = \mathbf{q}^{i} + \frac{\Delta t}{2}\frac{\partial H}{\partial \mathbf{p}}\bigg|_{\mathbf{p}=\mathbf{p}^{i}}, \tag{5.21}$$

$$\mathbf{p}^{i+1} = \mathbf{p}^{i} - \Delta t\frac{\partial H}{\partial \mathbf{q}}\bigg|_{\mathbf{q}=\mathbf{q}^{i+\frac{1}{2}}}, \tag{5.22}$$

$$\mathbf{q}^{i+1} = \mathbf{q}^{i+\frac{1}{2}} + \frac{\Delta t}{2}\frac{\partial H}{\partial \mathbf{p}}\bigg|_{\mathbf{p}=\mathbf{p}^{i+1}}, \tag{5.23}$$

where $i = 0, 1, 2, \ldots$ is the index for each time step and $(\mathbf{q}^0, \mathbf{p}^0)$ are initial conditions. Depending on the initial conditions, the Hamiltonian (5.2) yields three different types of trajectories: non-resonant, resonant, and irregular trajectories. An example for each type is depicted in the top row of Fig. 5.2 together with a SOS along the axis of symmetry, $q_1 = q_2$, in the bottom row. Any non-resonant trajectory, like the one in the top left panel, will inherit the symmetry of the Hamiltonian, i.e., an exchange of $q_1 \leftrightarrow q_2$ yields the same trajectory. Resonant trajectories may inherit the same symmetry, otherwise they appear in degenerate pairs. One such pair is the resonant trajectory in the middle row and its counterpart, which is the mirror image with respect to the axis of symmetry. Many kinds of resonant trajectories exist; the one in the middle of Fig. 5.2 is a $2:1$ resonance since it oscillates twice along the asymmetric stretch and once along the symmetric stretch before coming back near its starting point. Both non-resonant and resonant trajectories are quasiperiodic and can be quantized using the EBK method. For the $2:1$ resonance the areas inside both loops on the SOS are equal and determine one action variable. The trajectory in the right panels of Fig. 5.2 is irregular. It is not confined to a torus but instead fills a portion in phase space. The SOS reveals this by showing densely filled areas instead of lines. Consequently, such trajectories cannot be quantized.

Figures 5.3 and 5.4 exhibit the dynamics of system (5.2) for different energies and coupling strengths. Both figures show Poincaré SOS located along the axis of symmetry, $q_1 = q_2$. At each energy trajectories were run for various initial conditions, sampling the full energetically accessible phase space. In two degrees of freedom the full phase space of an autonomous system can be explored by using initial conditions along a suited one-dimensional curve of constant energy. [102] We chose initial conditions along the edge of the energetically allowed region of the potential, i.e., $q_1, q_2 : V(q_1, q_2) = E$ and $p_1 = p_2 = 0$.

Fig. 5.3 explores the dynamics at a fixed coupling strength of $\lambda = 0.6$. The three types of trajectories from Fig. 5.2 can be found here. At low energies of $E < 0.5\,D_e$, quasiperiodic trajectories fill the phase space. The outer loops, as well as the island in the center correspond to non-resonant trajectories. The four smaller islands surrounding the center island correspond to the pairs of $2:1$ resonant trajectories. At about half of the dissociation

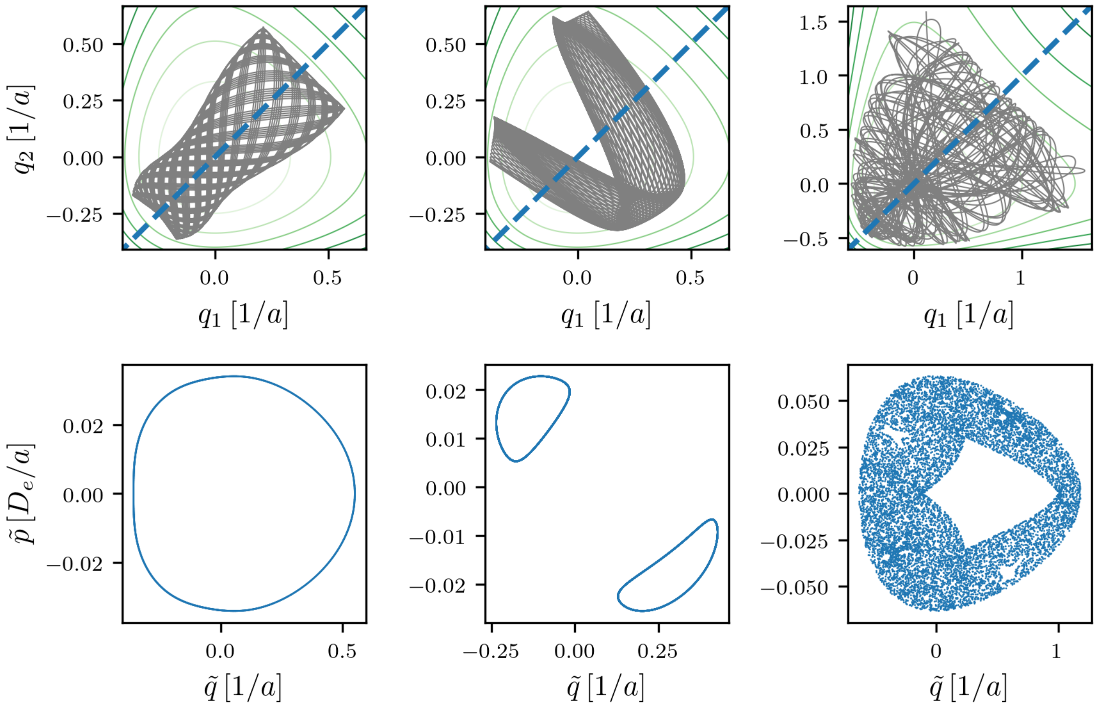

Figure 5.2: The **top row** depicts three types of bound trajectories in coordinate representation for the Hamiltonian of two kinetically coupled, identical Morse oscillators, Eq. (5.2). The **left** trajectory represents non-resonant motion, the **middle** one represents a quasiperiodic $2:1$ resonance, and the **right** one shows irregular behavior. The Morse potential is shown as green contour lines. The Morse parameters were set to $D_e = 32$ and $a = 1/5$ and the coupling strength is $\lambda = 0.6$. The **bottom row** shows the corresponding SOS along the blue dashed line in the top panels. The values $\tilde{q}$ denote the positions along the blue dashed line and $\tilde{p}$ denote the projections of the momenta tangential to that line.

energy, $E \approx 0.5\,D_e$, irregular motion appears near the separatrices between the regular islands. With increasing energy chaotic dynamics spreads through phase space, swallowing the outer non-resonant trajectories, as well as the resonant islands, which disappear around $E \approx 0.7\,D_e$. Only the non-resonant island in the center persists up to (and beyond) the dissociation threshold.

Fig. 5.4 explores the system at fixed $E = 0.5\,D_e$ for various coupling strengths. At $\lambda = 0$ the system is fully separated and (almost) all trajectories come in degenerate pairs of $1:1$ resonances, forming two distinct degenerate islands. At finite coupling, $\lambda > 0$, non-resonant trajectories create the center island and the outer rings, which exist throughout the whole range of $0 < \lambda < 1$. The two islands belonging to the $1:1$ resonances [103] decrease in size with increasing λ. As the coupling varies, resonant islands appear and disappear. Low order resonances are more prominent in phase space than higher order resonances. Chaotic motion is apparent for the whole range of $0 < \lambda < 1$, but least prominent at large λ. Note that for both $\lambda = 0$ (separable) and $\lambda = 1$ (one-dimensional) the dynamics is fully regular. In Ref. [98] the approximate relative area of chaotic and non-chaotic phase space regions is provided for the whole λ-range at energies of $0.8\,D_e$ and $1.0\,D_e$.

The origin of the resonant quasiperiodic motion can be directly seen when comparing the two coupled Morse oscillators, Eq. (5.2), to the dynamics of two coupled harmonic oscillators,

$$H_{\mathrm{HO}}(\lambda) = \frac{p_1^2}{2\mu_r} + \frac{p_2^2}{2\mu_r} + V_{\mathrm{HO}}(q_1) + V_{\mathrm{HO}}(q_2) - \lambda p_1 p_2, \tag{5.24}$$

where $V_{\mathrm{HO}}(x) = \frac{1}{2}\mu_r \omega^2 x^2$ and $\omega = a\sqrt{2D_e/\mu_r}$ (Eq. (5.7)). Using the same rotation in coordinates as in Appendix B.1, system (5.24) can be completely decoupled into two separate harmonic oscillators with different effective masses, [104]

$$\tilde{H}_{\mathrm{HO}}(\lambda) = \frac{p_1'^2}{2\mu_r}(1 - \mu_r\lambda) + V_{\mathrm{HO}}(q_1') + \frac{p_2'^2}{2\mu_r}(1 + \mu_r\lambda) + V_{\mathrm{HO}}(q_2') \tag{5.25}$$

with energies

$$E_{\mathrm{HO}}(n_1, n_2; \lambda) = \hbar\underbrace{\omega\sqrt{1 - \mu_r\lambda}}_{\omega_1}\left(n_1 + \tfrac{1}{2}\right) + \hbar\underbrace{\omega\sqrt{1 + \mu_r\lambda}}_{\omega_2}\left(n_2 + \tfrac{1}{2}\right). \tag{5.26}$$

The energy spectra of both systems, Eq. (5.2) and Eq. (5.24), are depicted in Fig. 5.5, left and right, respectively. Close examination reveals a resemblance between the two spectra. The Morse potential is an anharmonic distortion of the harmonic potential with a dissociation threshold and consequently one expects the energetically lower-lying states to behave nearly harmonic. The major difference between the two systems is the separability of the coupled harmonic oscillators. Hence, the dynamics is a superposition of two independent one-dimensional motions and no irregularity occurs. Additionally, due to the separability no crossings are avoided.

The degeneracies in the energy spectra in Fig. 5.5, right, occur whenever the two harmonic oscillators fulfill the resonance condition, i.e., whenever the ratio of their frequencies becomes a rational number,

$$\frac{\omega_2}{\omega_1} = \sqrt{\frac{1 + \mu_r\lambda}{1 - \mu_r\lambda}} = \frac{\alpha}{\beta}\,, \quad \alpha, \beta \in \mathbb{Z}. \tag{5.27}$$

Whenever this condition is fulfilled, we refer to the resonance as "$\alpha : \beta$ resonance". Due to the positivity of λ we have $\omega_2 \geq \omega_1$. The most prominent resonances are the $1 : 1$ resonance at $\lambda = 0$ and the $2 : 1$ resonance at $\lambda = 0.6$, being the two resonances of lowest orders. For $\alpha : \beta$ resonant harmonic oscillators all trajectories become periodic and wind α times around one topologically distinct loop around their phase space torus and β times around the other loop before connecting back into themselves. Such a resonant torus is densely filled by a family of periodic resonant trajectories. The top row of Fig. 5.6 depicts three resonant trajectories for the Hamiltonian (5.24).

If the harmonic system is perturbed into a system of anharmonic Morse oscillators, the resonant tori break up into alternating pairs of elliptic (stable) and hyperbolic (unstable) periodic orbits, creating the island chains surrounding the non-resonant center island, as seen in Figs. 5.3 and 5.4. This is a consequence of the POINCARÉ-BIRKHOFF theorem. [105] These island chains correspond to the quasiperiodic resonances mentioned before. The bottom row of Fig. 5.6 depicts three quasiperiodic resonances of the anharmonic system, Eq. (5.2), corresponding to the periodic resonances of the harmonic system, Eq. (5.24), in the top row. As we will see in the upcoming section, such quasiperiodic resonances provide "corrected" EBK energies to the non-resonant EBK spectra.

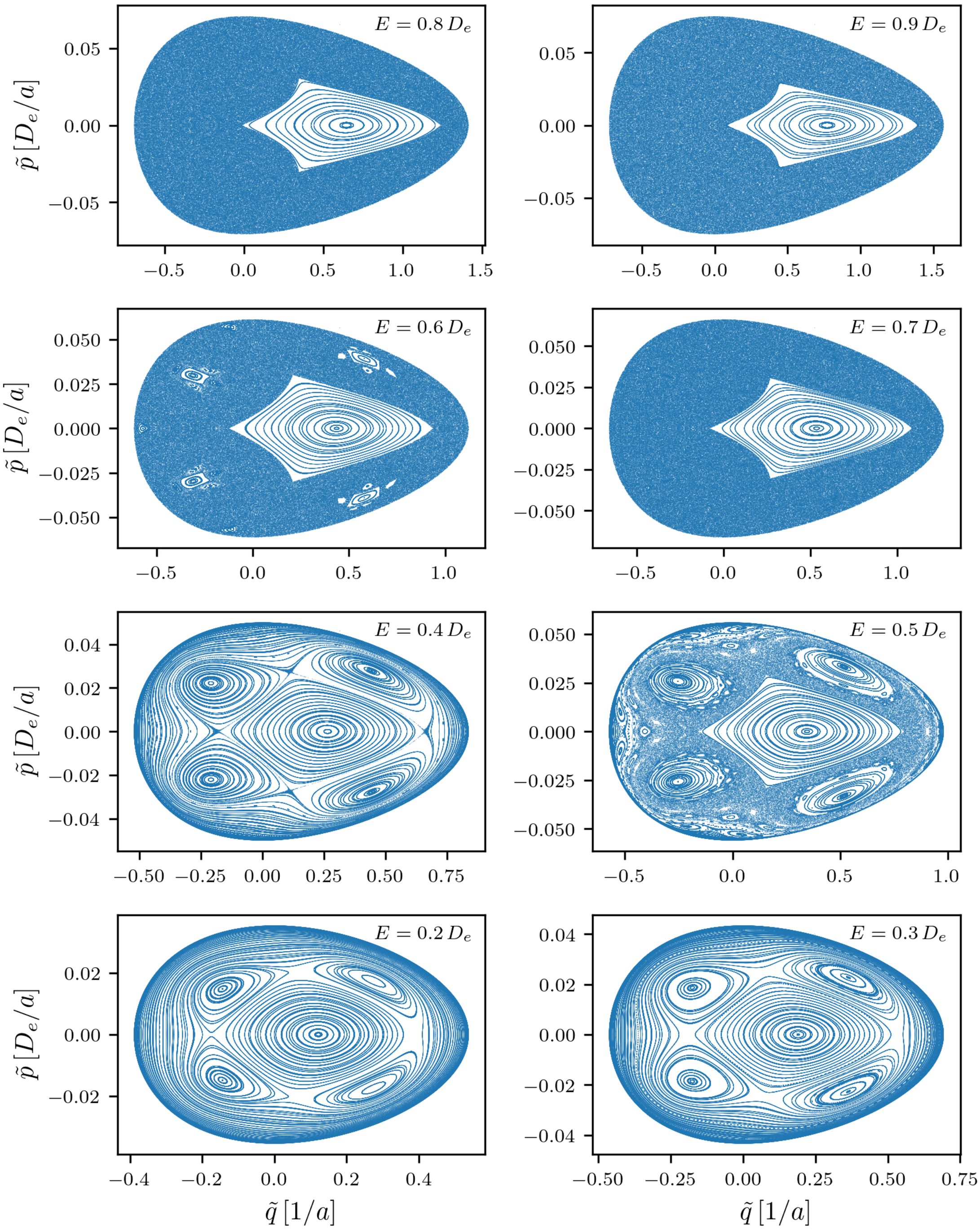

Figure 5.3: Poincaré SOS are given for the Hamiltonian (5.2) for various energies and $\lambda = 0.6$, $D_e = 32$ and $a = 1/5$. The section plane of Fig. 5.2 is used, i.e., along the axis of symmetry, $q_1 = q_2$. The values $\tilde{q}$ denote the positions along the section plane and $\tilde{p}$ denote the projections of the momenta tangential to the section plane.

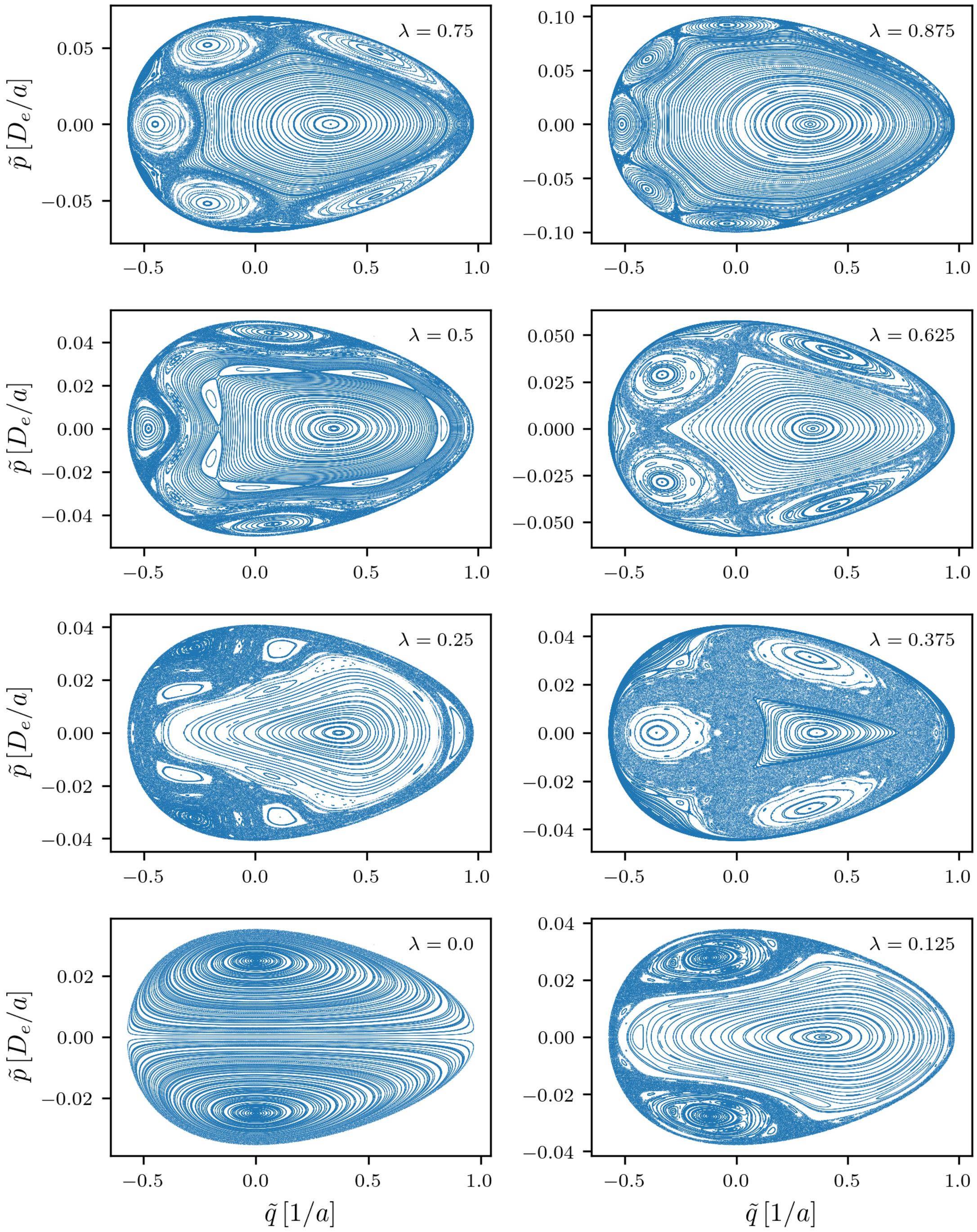

Figure 5.4: Same as Fig. 5.3, but for various coupling strengths and $E = 0.5\,D_e$.

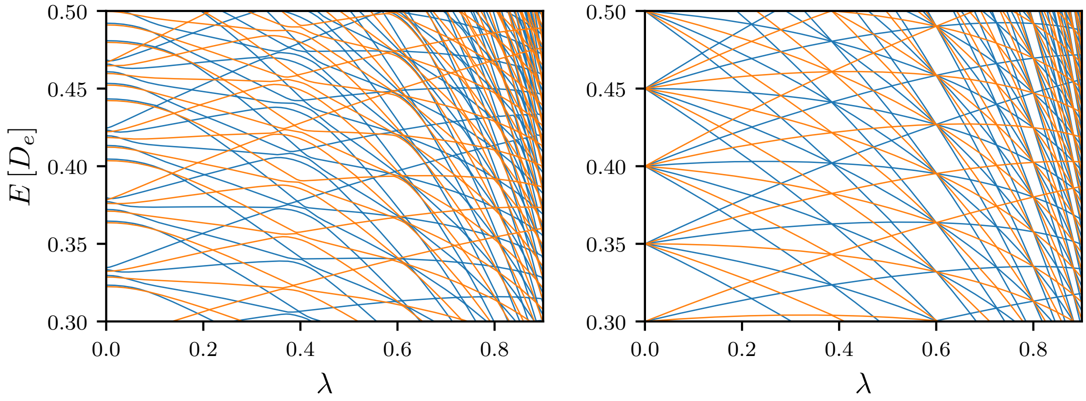

Figure 5.5: The energy spectra of two kinetically coupled Morse oscillators (**left**), Eq. (5.2), and two kinetically coupled harmonic oscillators (**right**), Eq. (5.24), are shown. The blue eigenenergies correspond to symmetric and the orange eigenenergies to antisymmetric eigenstates.

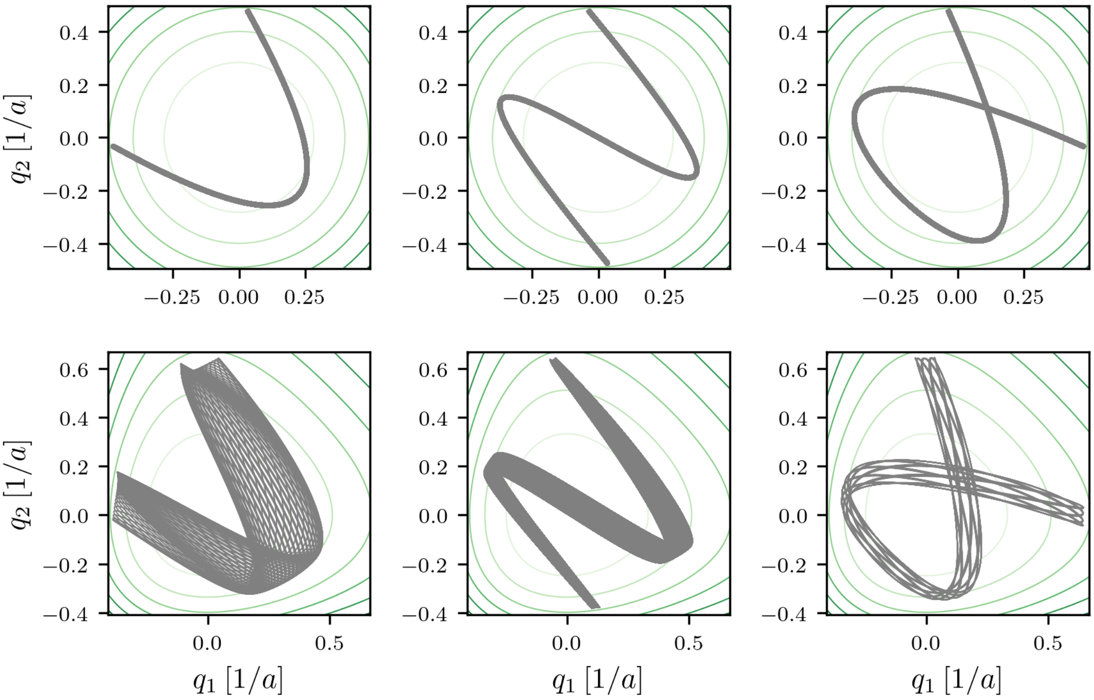

Figure 5.6: Three resonances of the coupled harmonic oscillators, Eq. (5.24), are depicted in the **top row**, together with their quasiperiodic resonant counterparts of the coupled Morse oscillators, Eq. (5.2), in the **bottom row**. The **left column** shows a $2:1$, the **middle column** a $3:1$, and the **right column** a $3:2$ resonance. These resonances are located at $\lambda = 0.6$, $\lambda = 0.8$, and $\lambda = 5/13 \approx 0.38$, respectively. The potentials are shown as green contour lines.

5.3 EBK quantization via GPR

Using quasiperiodic trajectories, semiclassical EBK spectra can be computed. Therefor we compute trajectories for various initial conditions throughout a range of energies and perturbations and compute their two actions, S_1 and S_2, as described in Chapter 4. Time steps for the numerical propagation are chosen such that the relative error of the trajectory energy remained below 10^{-5}. All irregular trajectories as well as those whose actions are not properly converged are discarded. The two (non-integer) quantum numbers, relating to the actions by

$$S_i = 2\pi\hbar \left(n_i + \tfrac{1}{2}\right) \ , \quad i = 1, 2, \tag{5.28}$$

together with the perturbation parameter and the energy of each trajectory are used as training data for a GPR. The values S_1, S_2 and λ are used as components of a three-dimensional input and the energy is the corresponding output value.

Fig. 5.7 illustrates the method at a single fixed value of $\lambda = 0.3$. The left panel shows the energy surface trained by a number of quasiperiodic trajectories. The energy landscape appears to be very flat and smooth with varying n_1 and n_2. The GPR registers this flatness by assigning large length-scales of $l_1, l_2 \approx 12$ via the MLE method, which are much larger than the average distance between nearest trajectories, depicted as circles. At the same time the noise hyperparameter is very small, $\sigma_n \approx 10^{-4}$, implying that all training points lie precisely on the trained energy surface. The very large length-scales allow for extrapolation to regions where no training data is provided, i.e., energies above $E \geq 0.5\,D_e$. The right panel of Fig. 5.7 marks all energies at points where n_1 and n_2 are integers and hence fulfill the EBK quantization condition. Naturally, with growing distance to the training data the uncertainty grows. However, as the learned length-scales are very large, the uncertainty grows very slowly with distance, allowing still for precise extrapolation. The depicted uncertainties in the right panel of Fig. 5.7 are highly upscaled solely for the purpose of visibility.

The method illustrated in Fig. 5.7 at a single value for λ is conducted using variable λ as a third input dimension. Trajectories are propagated using distributed initial conditions throughout the full energy range $0 < E < D_e$ and various perturbation strengths $0 \leq \lambda < 1$. Of all the propagated trajectories, only those quasiperiodic ones are used whose phase space torus is properly densely filled, i.e., their Poincaré SOS depict closed loops without large gaps between neighboring points. Due to these criteria the number of useful trajectories decreases with increasing energies above $E \gtrsim 0.5\,D_e$, as chaotic dynamics takes over phase space. Within the full batch of quasiperiodic trajectories, non-resonant, as well as trajectories with different resonances are treated separately. While non-resonant trajectories are found throughout the whole range of perturbation strengths, different resonances occur within narrow ranges of λ. For this reason, we generate one full λ-range of semiclassical energy spectra using only non-resonant trajectories, whereas resonant trajectories are only interpolated within the small region where they really occur. The resonant EBK energies then yield only components of spectra within narrow λ-ranges.

The result for the semiclassical eigenenergies is shown in Fig. 5.8. In the left panel the EBK spectra arising from GPR using non-resonant trajectories are partially depicted in colored dashed lines. The first section plane for the Poincaré SOS was chosen along $q_1 = q_2$, as depicted in Fig. 5.2. The second section was located perpendicularly such that it properly dissects the phase space torus. Consequently, the first SOS, yielding the values for S_1, are

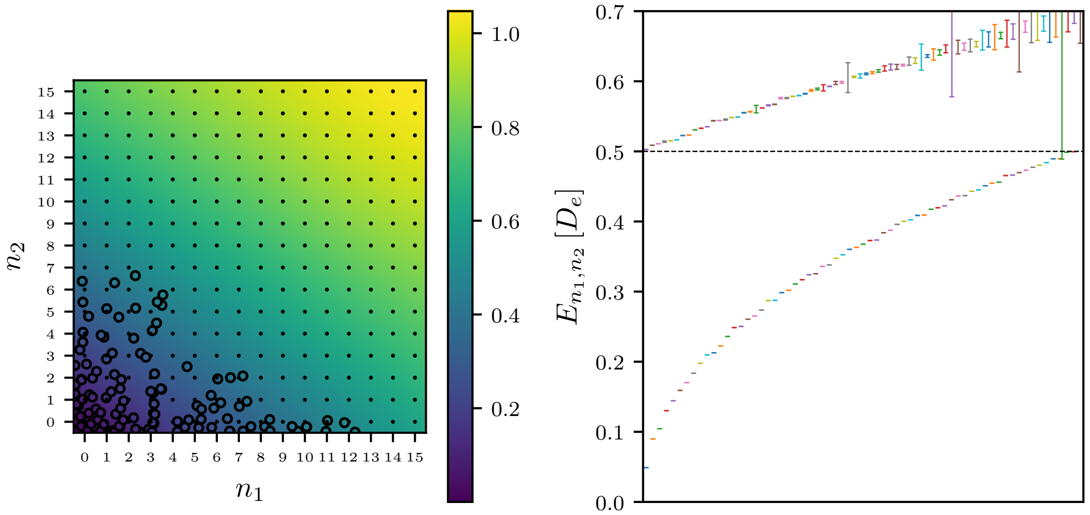

Figure 5.7: The **left** panel illustrates the GPR for the energy (colormap) as a function of the two (non-integer) quantum numbers of various quasiperiodic trajectories (circles) at fixed $\lambda = 0.3$. The black dots locate the positions where the EBK quantization condition is fulfilled. The length-scales and noise hyperparameters of the GPR are $l_1, l_2 \approx 12$ and $\sigma_n \approx 10^{-4}$. The **right** panel shows the EBK energies corresponding to dots in the left panel. The dashed line shows the energy up to which trajectories were computed in the left panel. The errorbars are highly upscaled uncertainties of the GPR.

the ones shown in Figs. 5.2 – 5.4. The inter- and extrapolated integer quantum numbers n_1 are associated with the number of nodes of the eigenfunctions along the symmetric stretch, $q_s \sim q_1 + q_2$. Accordingly, the perpendicular SOS yield the values for S_2 and n_2, the latter of which corresponds to the number of nodes of the eigenfunctions along the asymmetric stretch, $q_a \sim q_2 - q_1$. A GP is trained on the n_1, n_2, λ and E-values of 500 non-resonant trajectories. The MLE method yields optimal length-scales of $(l_1, l_2, l_3) \approx (22, 22, 0.12)$ and a noise hyperparameter of $\sigma_n \approx 10^{-3}$. The first two length-scales correspond to the length-scales along the n_1 and n_2-directions. Evidently, the energy landscape is extremely flat and smooth in those directions, allowing for extrapolation to higher energies, where chaotic dynamics takes over. The small noise hyperparameter implies that all training points lie precisely on the trained regression surface.

The blue, orange, green and red lines in the left panel of Fig. 5.8 represent energies of eigenstates with quantum numbers satisfying $n_1 + n_2 = 6$, 7, 8 and 9, respectively. For visualization purposes only states with even n_2 are shown, i.e., with an even number of nodes along the asymmetric stretch. This condition selects only those states which are symmetric under coordinate exchange, $q_1 \leftrightarrow q_2$. In gray, the symmetric full quantum spectra are shown for comparison. One can immediately see that the EBK energies show the diabatic spectra of the system, ignoring all avoided crossings. The result shares strong resemblance with the spectra of the harmonic system shown in Fig. 5.5, right. One convenience of this semiclassical method is the persistence of the sets of quantum numbers throughout the parameter λ, allowing for immediate identification of the states belonging to one and the same diabatic manifold. Note that the restrictions in the quantum numbers and symmetry of the shown states was done for visualization improvements only. The described method always supplies

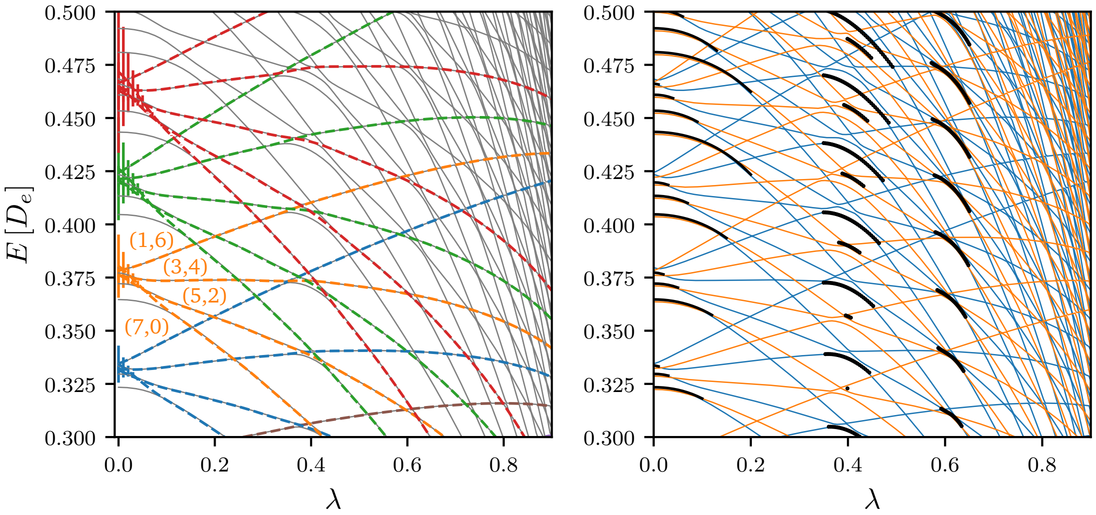

Figure 5.8: The **left** panel shows the full quantum (gray) and semiclassical EBK (dashed colored) spectra for symmetric states. For selected states the EBK quantum numbers, (n_1, n_2), are given. The errorbars show an upscaled GPR uncertainty. The length-scales and noise hyperparameters of the GPR are $(l_1, l_2, l_3) \approx (22, 22, 0.12)$ and $\sigma_n \approx 10^{-3}$. The **right** panel shows the full quantum spectra for both symmetric (blue) and anti-symmetric (orange) states, as well as energies of resonant quasiperiodic trajectories that fulfill the EBK condition in black.

the full spectrum.

We have thus shown that the full diabatic spectrum of the system can be extracted with one single GPR using only a few hundred regular, non-resonant trajectories. The errorbars at small λ-values are the ten times upscaled uncertainties of the posterior GP. These uncertainties appear due to the quasiperiodic $1:1$ resonances taking up a large portion of phase space, leaving little space for non-resonant trajectories. This effect can be seen in the two bottom panels of Fig. 5.4. Since the optimal length-scale for the λ-direction is rather small, $l_3 \approx 0.12$, extrapolation in λ-direction is less meaningful. If one were still to conduct extrapolation, the respective length-scale should be forced to take up larger values, which can be achieved by setting a lower bound for the MLE optimizer, e.g., $l_3 \geq 1$. On the other hand, the very large length-scales l_1 and l_2 suggest that extrapolation to higher energies is justified. Indeed, diabatic spectra can be continued into regions where chaotic dynamics already takes place. This can be seen in Fig. 5.9, where full diabats are shown for the energy region above $E > 0.5\,D_e$, where chaotic dynamics takes over the majority of phase space, as can be seen in Fig. 5.3. This method thus serves as a tool to generate semiclassical EBK spectra where classical dynamics is irregular. The generated diabatic energy spectra correspond to a purely regular approximation to the system. One could say that an additional, missing constant of motion is assumed, treating the system as if it were integrable. This aspect will be further investigated in the next chapter.

As discussed in Sec. 5.1, the higher the quantum number n_1, the stronger the eigenstate loses energy with increasing coupling, λ. This is due to the effect of compression of the eigenfunction along the direction of the symmetric stretch.

In the right panel of Fig. 5.8 the full quantum spectra for symmetric and anti-symmetric states are depicted in blue and orange, respectively. The semiclassical EBK eigenenergies of resonant trajectories are marked in black. The energies at small $\lambda < 0.2$ correspond to pairs of degenerate $1:1$ resonances and the energies around $\lambda = 0.6$ correspond to degenerate pairs of $2:1$ resonances, such as the one shown in Fig. 5.6, bottom left. The energies around $\lambda = 5/13 \approx 0.38$ correspond non-degenerate $3:2$ resonances, such as the one shown in Fig. 5.6, bottom right. All resonant trajectories are located in a small portion in phase space in a narrow λ-region. Most resonant trajectories, especially higher-order resonances, do not fulfill the EBK quantization condition as their phase space torus is too narrow, i.e., $n < 0$ for at least one quantum number (or $S < \pi$ for at least one action). Quasiperiodic resonances merge into non-resonant trajectories with changing initial conditions or perturbation strengths, and with increasing energies they vanish into the irregular sea, as seen in Fig. 5.3. The resonant EBK energies are results of GP interpolation between resonant trajectories. The narrow λ-regions in which these resonances exist cause extrapolation beyond those regions to be somewhat unphysical. However, extrapolation to higher energies into the irregular regime should be permitted, which is confirmed by the optimal GP length-scales along the n_1 and n_2-directions. These values lie at $l_1, l_2 \approx 6$ and are quite larger than the nearest distance between quantum numbers, $\Delta n_i = 1$. The non-resonant EBK spectra in the left panel are the result of an interpolation ignoring the small regions where resonant trajectories take place.

The partial spectra coming from resonant trajectories are qualitatively different from the non-resonant spectra. Fig. 5.8, right, shows that eigenenergies coming from resonances tend to lie on those states that are most strongly repelled at the avoided crossings. Hence, they “correct” the interpolated non-resonant spectra at avoided crossings. The $1:1$ and $2:1$ resonances come in degenerate pairs and therefore lie close to pairs of almost degenerate states with opposite symmetries, i.e., where one blue and one orange state nearly overlap. This is most visible for the $1:1$ resonance at $\lambda < 0.2$, but similar for the $2:1$ resonance. The $3:2$ resonance is located where symmetric and anti-symmetric states are far apart from each other. This is in agreement with the fact that those trajectories do not come in degenerate pairs, since they inherit the symmetry of the Hamiltonian, i.e., exchange of $q_1 \leftrightarrow q_2$ yields the same trajectory. For these resonances it is also most visible that the lower-energetic anti-symmetric states vanish. This is caused by the mentioned effect that resonant quasiperiodic trajectories exist in a confined region of phase space and fulfill the EBK quantization condition only above a certain energy threshold. From a physical perspective one can observe that with increasingly avoided crossings resonant EBK energies gain larger contributions. If one were to extrapolate to larger quantum numbers, more states, including those missing anti-symmetric ones, would emerge. The connection between classical quasiperiodic resonances and avoided crossings has been investigated, e.g., in Ref. [106].

It should be noted that the quantum numbers, n_1 and n_2, corresponding to resonant trajectories take different values than those of non-resonant tori. More precisely, if the first quantum number of a quasiperiodic resonance is taken to be the area inside one loop such as the ones in the bottom middle panel of Fig. 5.2, then the second quantum number will be much larger due to the elongated torus, $n_2 > n_1$. This effect increases with the order of the resonance. For the case of the $3:2$ resonance in Fig 5.8, right, all visible symmetric EBK states correspond to $n_1 = 0$ and all anti-symmetric ones to $n_1 = 1$.

The primitive trajectories considered here live entirely in real-valued phase space and therefore do not account for effects due to dynamic tunneling, which would lead to splittings in the energies between pairs of degenerate, mirrored trajectories. Such effects are not present at $\lambda = 0$, due to the separability of the system. With increasing coupling strength, $\lambda > 0$, the system loses its separable character and tunneling effects grow in strength. Such effects are likely to lead to the energy splittings that are visible at the right ends of the black lines depicting the $1 : 1$ resonances at $\lambda < 0.2$. [107] However, this was not investigated within the present work. Contributions of other resonances besides the three depicted were not included. But according to Fig. 5.4, other resonances might also yield spectral fragments within the energy range of Fig. 5.8.

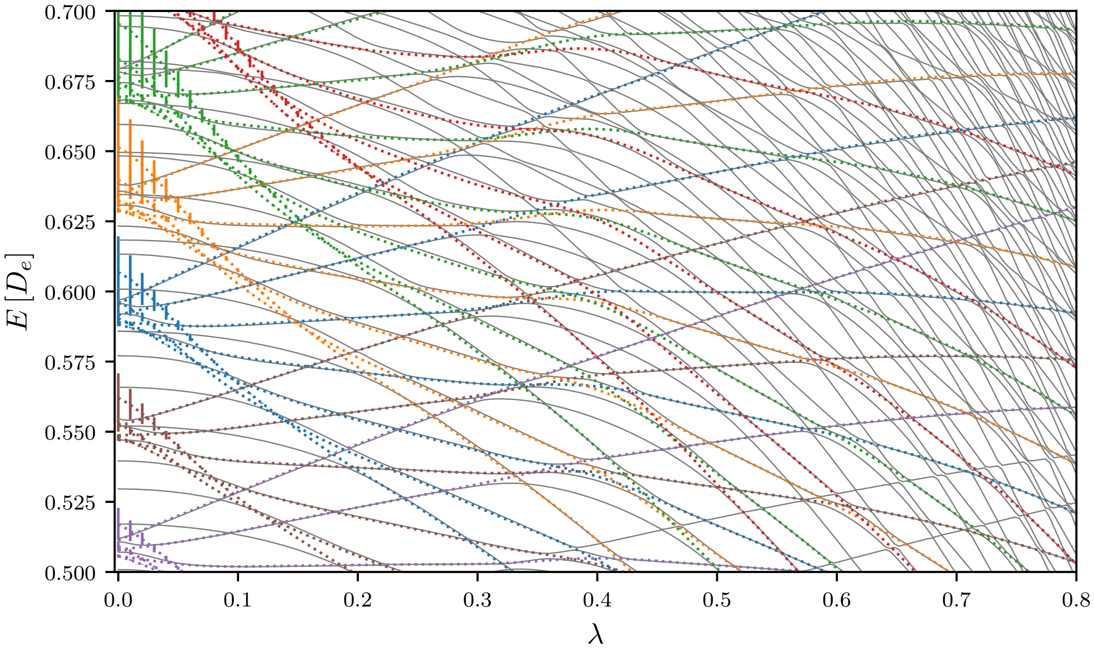

Figure 5.9: Similar to the left panel of Fig. 5.8, symmetric states of the the full quantum (gray) and semiclassical EBK (dotted colored) spectra are shown. The errorbars show the GPR uncertainty.

6 Semiclassical diabatization of hydrogen in a magnetic field

In this chapter we will demonstrate the method for semiclassical diabatization using GPR on the energy spectra of a single hydrogen atom exposed to an external homogeneous magnetic field. This simple but prominent system exhibits both regular and irregular classical dynamics and allows for a useful scaling transformation. [108]

The electronic TISE with minimal coupling to a magnetic field with vector potential $\mathbf{A}(\mathbf{x}, t)$ is found by substituting the momentum operator in the following way: $\mathbf{p} \to \mathbf{p} - q\mathbf{A}(\mathbf{x}, t)$, where $q = -e$. We will be using atomic units throughout the whole chapter: $\hbar = m_e = e = a_0 = 1$. For a homogeneous magnetic field parallel to the z-axis, $\mathbf{B} = (0, 0, B)^\top$, we use a vector potential in the symmetric gauge,

$$\mathbf{A} = \frac{B}{2}\left(-y, x, 0\right)^\top . \tag{6.1}$$

The Hamiltonian then writes

$$\begin{aligned} H &= \frac{1}{2}\big(\mathbf{p} + \mathbf{A}(\mathbf{x}, t)\big)^2 - \frac{1}{r} && (6.2)\\ &= \frac{\mathbf{p}^2}{2} + \frac{B}{2}\underbrace{(xp_y - yp_x)}_{L_z} + \frac{B^2}{8}\left(x^2 + y^2\right) - \frac{1}{r}. && (6.3) \end{aligned}$$

Due to the cylindrical symmetry of the system, one can perform a coordinate transformation to cylindrical coordinates, $(x, y, z) \to (\rho, \varphi, z)$, where $\rho = \sqrt{x^2 + y^2}$ and $\varphi = \arctan\left(\frac{y}{x}\right)$, yielding for the Laplace operator

$$\begin{aligned} \mathbf{\Delta} &= \frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho\frac{\partial}{\partial \rho}\right) + \frac{1}{\rho^2}\frac{\partial^2}{\partial \varphi^2} + \frac{\partial^2}{\partial z^2} && (6.4)\\ &= \frac{1}{\rho}\frac{\partial}{\partial \rho} + \frac{\partial^2}{\partial \rho^2} + \frac{1}{\rho^2}\frac{\partial^2}{\partial \varphi^2} + \frac{\partial^2}{\partial z^2}. && (6.5) \end{aligned}$$

The z-component of the orbital angular momentum, L_z, is a constant of motion and its corresponding magnetic quantum number, m, is therefore a good quantum number of the system. The full single-electron wave function is separable in the azimuthal angle, φ, [109]

$$\psi(\rho, \varphi, z) = \psi(\rho, z)\frac{e^{im\varphi}}{\sqrt{2\pi}}. \tag{6.6}$$

Here the φ-dependence follows from the fact that the SPHERICAL HARMONICS, $Y_{lm}(\theta, \varphi)$, are eigenfunctions of the orbital angular momentum operator, L_z. What is left is the nonseparable part of the TISE for the coordinates (ρ, z),

$$\left[\frac{1}{2}\left(-\frac{1}{\rho}\frac{\partial}{\partial\rho}-\frac{\partial^2}{\partial\rho^2}-\frac{\partial^2}{\partial z^2}+\frac{m^2}{\rho^2}\right)+\frac{B^2}{8}\rho^2-\frac{1}{\sqrt{\rho^2+z^2}}\right]\psi(\rho,z)=E\psi(\rho,z), \tag{6.7}$$

where we neglect the linear Zeeman term, $\frac{1}{2}BL_z$, which just adds a constant contribution to the energy. This system corresponds to that of a particle moving in an effective two-dimensional potential,

$$V_{\text{eff}}(\rho,z)=\frac{m^2}{2\rho^2}+\frac{B^2\rho^2}{8}-\frac{1}{\sqrt{\rho^2+z^2}}. \tag{6.8}$$

In order to compute the eigenenergies and eigenstates of the system, one represents the Hamiltonian in a specified basis and diagonalizes the corresponding matrix. The generic method of finite differences, which discretizes the eigenstates on a grid and has been used in the previous chapter, is presented in Appendix C.1. This method is unsuited to treat the singularity in the center of the atom properly and the number of necessary grid points per dimension becomes prohibitively large.

A more suitable basis needs to quickly converge to the true eigenenergies. For the hydrogen atom in a magnetic field the orthogonal STURMIAN functions pose as a sensible basis. Non-orthogonal Sturmian functions have been first applied to the hydrogen in a magnetic field in Refs. [110, 111]. Later, Ref. [112] introduced an orthogonalized version of the Sturmian functions, which will be applied in the present chapter.

6.1 Orthogonal Sturmian basis

A suitable basis for the hydrogen atom in a magnetic field can be formed using orthogonalized Sturmian functions, [112–114]

$$S_{nl}^{(\zeta)}(r)=\zeta^{\frac{3}{2}}\sqrt{\frac{(n-l-1)!}{(n+l+1)!}}e^{-\frac{\zeta r}{2}}(\zeta r)^l L_{n-l-1}^{(2l+2)}(\zeta r). \tag{6.9}$$

The associated (or generalized) Laguerre polynomials, $L_\beta^{(\alpha)}(x)$, satisfy the relation

$$\int_0^\infty \mathrm{d}x\; e^{-x}x^\alpha L_\beta^{(\alpha)}(x)L_{\beta'}^{(\alpha)}(x)=\frac{(\alpha+\beta)!}{\beta!}\delta_{\beta'\beta}, \tag{6.10}$$

from which the orthogonality of the Sturmian functions follows. The parameter ζ specifies around which principal quantum number, n, the eigenenergies converge fastest. If one chooses $\zeta = 2/n^\star$, the eigenvalues around $n = n^\star$ experience the highest accuracy. Using the basis $\langle r,\theta,\varphi|n,l,m\rangle = \langle r|n,l\rangle\,\langle\theta,\varphi|l,m\rangle \equiv S_{nl}(r)\,Y_{lm}(\theta,\varphi)$, where Y_{lm} are the spherical harmonics, the matrix elements of the Hamiltonian, given in spherical coordinates as (neglecting the constant linear Zeeman term)

$$H(r,\theta,\varphi)=\underbrace{\frac{p^2}{2}-\frac{1}{r}}_{H_0}+\frac{B^2}{8}r^2\sin^2\theta, \tag{6.11}$$

are

$$\left\langle n',l',m'\right| H \left|n,l,m\right\rangle = \underbrace{\left\langle n',l',m'\right| H_0 \left|n,l,m\right\rangle}_{I_0} + \frac{B^2}{8} \underbrace{\left\langle n',l'\right| r^2 \left|n,l\right\rangle}_{I_1} \underbrace{\left\langle l',m'\right| \sin^2\theta \left|l,m\right\rangle}_{I_2}. \tag{6.12}$$

The expressions for the matrix elements are derived in Appendix C.2; the results are

$$\begin{aligned} I_0 = & -\frac{\zeta^2}{8}\delta_{n'n}\delta_{l'l}\delta_{m'm} + \frac{\zeta^2}{2}\delta_{l'l}\delta_{m'm}\sqrt{\frac{(n'-l'-1)!}{(n'+l'+1)!}}\sqrt{\frac{(n-l-1)!}{(n+l+1)!}} \\ & \times \sum_{\tau=0}^{\substack{\leq n'-l'-1 \\ n-l-1}} \frac{(2l+\tau)!}{\tau!}\Bigg[(n-l-1)(n-l-\tau)(n'-l-\tau) \\ & -(n+l+1)(n-l-1-\tau)(n'-l-\tau) + \left(n-\frac{2}{\zeta}\right)(2l+\tau+1)\Bigg], \end{aligned} \tag{6.13}$$

$$\begin{aligned} I_1 = & \, \zeta^{-2}\sqrt{\frac{(n'-l'-1)!}{(n'+l'+1)!}\frac{(n-l-1)!}{(n+l+1)!}}(4+l+l')! \\ & \times \sum_{\tau=0}^{\substack{\leq n'-l'-1 \\ n-l-1}} (-1)^{n'+l'+n+l+\tau}\binom{2-l'+l}{n'-l'-1-\tau}\binom{2+l'-l}{n-l-1-\tau}\binom{-5-l'-l}{\tau}, \end{aligned} \tag{6.14}$$

$$\begin{aligned} I_2 \overset{m'=m}{\equiv} & \, \frac{2\left(l^2+l+m^2-1\right)}{(2l+3)(2l-1)}\delta_{l'l} - \frac{\sqrt{(l+m)(l+m-1)(l-m)(l-m-1)}}{\sqrt{(2l+1)(2l-3)(2l-1)}}\delta_{l',l-2} \\ & - \frac{\sqrt{(l-m+1)(l-m+2)(l+m+1)(l+m+2)}}{\sqrt{(2l+1)(2l+5)(2l+3)}}\delta_{l',l+2}. \end{aligned} \tag{6.15}$$

Integral I_1 has non-vanishing terms only for $l' = l$ and $l' = l \pm 2$, which are listed explicitly in Refs. [113, 114]. To simplify integral I_2, equal magnetic quantum numbers were assumed, $m = m'$. Due to the neglection of the linear Zeeman term only positive magnetic quantum numbers need to be considered, $m \geq 0$, and hence eigenstates with different signs of m have degenerate energies.

For the range of spectra shown in Fig. 6.4 the parameter of the Sturmian basis has been set to $\zeta = 2/20$, ensuring convergence of states corresponding to principal quantum numbers $n \approx 20$.

6.2 Scaling and regularization

6.2.1 Scaling transformation

The classical dynamics given by the Hamiltonian in Eq. (6.3) is invariant under a scaling of the coordinates and momenta by the magnetic field strength. [108, 115] Ignoring the contribution of the constant linear Zeeman term, $BL_z/2$, the Hamiltonian in cylindrical coordinates is given by

$$H = \frac{1}{2}\left[p_\rho^2 + \frac{L_z^2}{\rho^2} + p_z^2\right] - \frac{1}{\sqrt{\rho^2+z^2}} + \frac{B^2}{8}\rho^2, \tag{6.16}$$

with $L_z = \rho^2 m_e \dot{\varphi}$. Now we introduce the following scaling transformation

$$\mathbf{r} = B^{-2/3}\tilde{\mathbf{r}}, \tag{6.17}$$

$$\mathbf{p} = B^{1/3}\tilde{\mathbf{p}}, \tag{6.18}$$

which yields the Hamiltonian in scaled coordinates,

$$\tilde{H} = B^{-2/3}H = \frac{\tilde{p}_\rho^2}{2} + \frac{\tilde{p}_z^2}{2} + \frac{\tilde{L}_z^2}{2\tilde{\rho}^2} - \frac{1}{\sqrt{\tilde{\rho}^2+\tilde{z}^2}} + \frac{1}{8}\tilde{\rho}^2 = B^{-2/3}E \equiv \epsilon. \tag{6.19}$$

The classical dynamics is now controlled by two scaled parameters: the scaled energy, $\epsilon = B^{-2/3}E$, and the scaled angular momentum, $\tilde{L}_z = B^{1/3}L_z$. Accordingly, the action along a classical trajectory is scaled by

$$S(E, L_z, B) = \oint \mathbf{p}(\mathbf{q})\cdot \mathrm{d}\mathbf{q} = B^{-1/3}\oint \tilde{\mathbf{p}}(\tilde{\mathbf{q}})\cdot \mathrm{d}\tilde{\mathbf{q}} \equiv B^{-1/3}\tilde{S}(\epsilon, \tilde{L}_z). \tag{6.20}$$

With this transformation we have effectively eliminated the explicit dependence on the magnetic field strength and hence reduced the number of parameters by one.

Another major advantage of this scaling transformation is the possibility of completely factoring out the magnetic field strength from the action, as seen on the right hand side of Eq. (6.20). As a consequence, one can cancel out the B-dependence by simply considering the ratio of action variables of a given quasiperiodic trajectory. The EBK quantization condition is then stated as

$$\frac{S_1}{S_2} \overset{(6.20)}{=} \frac{\tilde{S}_1}{\tilde{S}_2} \overset{(4.3)}{=} \frac{n_1 + \frac{\nu_1}{4}}{n_2 + \frac{\nu_2}{4}}, \tag{6.21}$$

with n_1, n_2 being integers. As was the case in the previous chapter, for two oscillators the Maslov indices count two encounters of caustics per torus cut: $\nu_1, \nu_2 = 2$. Then the EBK condition can be written as

$$\frac{\tilde{S}_1}{\tilde{S}_2} \overset{!}{=} \frac{n_1 + \frac{1}{2}}{n_2 + \frac{1}{2}} = \frac{2n_1+1}{2n_2+1}, \tag{6.22}$$

which is simply a fraction of two odd numbers. Hence, one single trajectory which fulfills above condition, namely their action ratio being a ratio of two odd numbers, yields semiclassical EBK eigenenergies for all quantum numbers which maintain the same ratio. For example, if one finds one quasiperiodic trajectory evolving according to the Hamiltonian (6.19) which fills a phase space torus with actions such that $\tilde{S}_1/\tilde{S}_2 = 1/3$, semiclassical eigenenergies are obtained for all quantum numbers with $(n_1, n_2) = (j, 3j+1)$, where $j = 0, 1, 2, \ldots$. A general relation for the compatible quantum numbers for an action ratio $\tilde{S}_1/\tilde{S}_2 = a/b$ of two odd numbers, a and b, is given as

$$(n_1, n_2) = \left(aj + \frac{a-1}{2}, bj + \frac{b-1}{2}\right) \ , \quad j = 0, 1, 2, \dots . \tag{6.23}$$

If one deals with action ratios with general (different) Maslov indices,

$$\frac{\tilde{S}_1}{\tilde{S}_2} = \frac{a}{b} \overset{!}{=} \frac{n_1 + \frac{\nu_1}{4}}{n_2 + \frac{\nu_2}{4}} = \frac{4n_1 + \nu_1}{4n_2 + \nu_2}, \tag{6.24}$$

all values for the compatible quantum numbers can be found once the first (lowest) pair of quantum numbers, $(n_1^\star, n_2^\star)$, corresponding to given action ratio is found. Assuming the valid reduced action ratio $\tilde{S}_1/\tilde{S}_2 = a/b$, one common integer factor k yields for the quantum numbers $(n_1^\star, n_2^\star)$

$$ka = 4n_1^\star + \nu_1, \tag{6.25}$$

$$kb = 4n_2^\star + \nu_2, \tag{6.26}$$

or

$$n_1^\star = \frac{ak - \nu_1}{4}, \tag{6.27}$$

$$n_2^\star = \frac{bk - \nu_2}{4}. \tag{6.28}$$

Then one needs to increase k by integer steps of Δk such that (n_1, n_2) remain integer values

$$n_1 = \frac{a(k + j \cdot \Delta k) - \nu_1}{4}, \tag{6.29}$$

$$n_2 = \frac{b(k + j \cdot \Delta k) - \nu_2}{4}. \tag{6.30}$$

Since we know that $ka = 4n_1^\star + \nu_1$, we can rewrite above equation as

$$n_1 = \frac{a(k + j \cdot \Delta k) - \nu_1}{4} \tag{6.31}$$

$$= \frac{ka - \nu_1 + aj\Delta k}{4} \tag{6.32}$$

$$= \frac{4n_1^\star + aj\Delta k}{4} \tag{6.33}$$

$$= n_1^\star + aj\frac{\Delta k}{4}, \tag{6.34}$$

and similarly $n_2 = n_2^\star + bj\frac{\Delta k}{4}$. Since a/b is already a reduced fraction, we find that Δk must equal to 4, yielding the general generators

$$(n_1, n_2) = (aj + n_1^\star, bj + n_2^\star) \ , \quad j = 0, 1, 2, \dots . \tag{6.35}$$

In summary, the scaling transformation, Eqs. (6.17) and (6.18), allows for factoring the magnetic field strength out of the expression for the scaled action, Eq. (6.20). By simply considering the ratio of the two actions one obtains a whole series of quantum numbers for which the EBK condition is simultaneously fulfilled, Eq. (6.21). In order to generate a full spectrum, regular trajectories for various values of the action ratio are needed. The resulting values for E and B can be extracted from the relations

$$E = B^{-2/3}\epsilon, \tag{6.36}$$

$$S_1 = \tilde{S}_1 B^{-1/3} = 2\pi\hbar\left(n_1 + \frac{\nu_1}{4}\right). \tag{6.37}$$

Note that the scaled action remains dependent on the two parameters ϵ and $\tilde{L}_z$.

6.2.2 Regularization

The singularity of the potential at $\mathbf{r} = 0$ forbids efficient propagation of classical trajectories. Every time a trajectory approaches the singularity, the time step size must be decreased accordingly to avoid extreme kicks from the steep potential, see Eq. (5.22). To circumvent this problem, we introduce a regularization using semiparabolic coordinates, (μ, ν), [109]

$$\nu^2 = \tilde{r} - \tilde{z}, \tag{6.38}$$

$$\mu^2 = \tilde{r} + \tilde{z}, \tag{6.39}$$

$$p_\nu = \frac{\mathrm{d}\nu}{\mathrm{d}\tau}, \tag{6.40}$$

$$p_\mu = \frac{\mathrm{d}\mu}{\mathrm{d}\tau}, \tag{6.41}$$

with a scaled time

$$\mathrm{d}t = 2\tilde{r}\,\mathrm{d}\tau = \left(\nu^2 + \mu^2\right)\mathrm{d}\tau\,, \tag{6.42}$$

and the relations

$$\tilde{\rho} = \nu\mu, \tag{6.43}$$

$$\tilde{z} = \tfrac{1}{2}\left(\mu^2 - \nu^2\right), \tag{6.44}$$

$$\frac{\mathrm{d}\tilde{\rho}}{\mathrm{d}t} = \mu\frac{\mathrm{d}\nu}{\mathrm{d}t} + \nu\frac{\mathrm{d}\mu}{\mathrm{d}t}, \tag{6.45}$$

$$\frac{\mathrm{d}\tilde{z}}{\mathrm{d}t} = \mu\frac{\mathrm{d}\mu}{\mathrm{d}t} - \nu\frac{\mathrm{d}\nu}{\mathrm{d}t}. \tag{6.46}$$

The Hamiltonian (6.19) can then be rewritten as

$$h = \frac{p_\nu^2}{2} + \frac{p_\mu^2}{2} - \epsilon\left(\nu^2 + \mu^2\right) + \frac{1}{8}\nu^2\mu^2\left(\nu^2 + \mu^2\right) + \frac{\tilde{L}_z^2}{2\mu^2} + \frac{\tilde{L}_z^2}{2\nu^2} = 2. \tag{6.47}$$

Classical trajectories propagated using the above equation are completely equivalent to those of Hamiltonian (6.19). The left hand side of Eq. (6.47) takes the form of the Hamiltonian of two identical harmonic oscillators with frequency $\omega = \sqrt{-2\epsilon}$ ($\epsilon < 0$ for bound states), coupled by a term $\sim\nu^2\mu^2\left(\nu^2 + \mu^2\right)$ plus the two terms $\sim\mu^{-2}$ and $\sim\nu^{-2}$. The energy of this system is then fixed at 2.

It is important to keep in mind that the last two terms in the Hamiltonian (6.47) diverge to $+\infty$ as $\mu, \nu \to 0$. This corresponds to the barrier at $\tilde{\rho} = 0$, caused by the term $\sim \tilde{\rho}^{-2}$

in Eq. (6.19). For this reason, these terms cannot be simply neglected by setting $\tilde{L}_z = 0$. The barrier forces trajectories to remain within their quadrant in semiparabolic coordinates and in cylindrical coordinates trajectories are forced to remain within $\tilde{\rho} \geq 0$. Hence, the semiparabolic coordinates are confined to one quadrant and we will choose $\mu, \nu \geq 0$. Note that these positive singularities do not cause problems when propagating trajectories like the Coulomb singularity does, as they lead to decrease of momenta, whereas negative singularities cause extremely large momenta.

6.3 Classical dynamics

The classical dynamics of the hydrogen atom in a homogeneous magnetic field has been extensively studied in the literature. [109] As shown in the previous section, the classical dynamics of system (6.19) is dependent on two parameters only, namely the scaled energy, ϵ, and the scaled angular momentum, $\tilde{L}_z$. Ref. [116] qualitatively explored the classical dynamics of the system for the full parameter region.

For our purposes, classical trajectories are propagated using the leap frog integration method, Eqs. (5.21) – (5.23). Fig. 6.1 shows the three types of emerging trajectories using the Hamiltonians (6.19) in the top row and (6.47) in the bottom row. Each trajectory in the top row is completely equivalent to its respective bottom one. The left and middle trajectories show regular, quasiperiodic behavior, while the right one is irregular. The left trajectory exhibits the same symmetry as the Hamiltonian (6.19), i.e., $z \leftrightarrow -z$. In semiparabolic coordinates this symmetry is given by $\mu \leftrightarrow \nu$. The middle trajectory does not have this symmetry and therefore comes in degenerate pairs, which are mirror images of each other with respect to the symmetry axis along $z = 0$. Comparing Hamiltonian (6.47) to (5.24) of the previous chapter, one can conclude that the middle type of trajectories represents a quasiperiodic resonance, likely a $1:1$ resonance according to its shape. Similarly to the system of two identical coupled Morse oscillators, these $1:1$ resonances appear in pairs and arise for two perturbed oscillators with frequencies $\omega_1 \simeq \omega_2$.

While both coordinate systems are equivalent, the Hamiltonian in semiparabolic coordinates has the major advantage of having no negative singularity. This allows for numerically stable integration of trajectories. Propagation of trajectories in cylindrical coordinates requires very small time steps whenever the trajectory comes close to the singularity. Especially for angular momenta of $\tilde{L}_z < 0.1$ numerical integration becomes unfeasible. For this reason, sample trajectories are shown for $\tilde{L}_z = 0.1$. In semiparabolic coordinates trajectories can be easily evolved for arbitrary values of $\tilde{L}_z$ and ϵ. However, we will restrict our investigations in this chapter to the case of $\tilde{L}_z = 0$.[1] In this case, the terms $\tilde{L}_z^2/(2\mu^2)$ and $\tilde{L}_z^2/(2\nu^2)$ in Eq. (6.47) become singular hard walls located along $\mu = 0$ and $\nu = 0$, respectively, on which a trajectory is reflected.

In Fig. 6.2 the classical dynamics is explored for $\tilde{L}_z = 0$ for increasing values of ϵ below zero. Above $\epsilon = 0$ the states are no longer necessarily bound, but trapped trajectories exist for $\tilde{L}_z > 0$ and are referred to as "resonant", which should not be confused with the kind of resonances we are discussing in this work. For the purpose of simplified treatment the hard boundaries at $\mu = 0$ and $\nu = 0$ were omitted and the Poincaré SOS was aligned along

---

[1] Vanishing angular momentum of $\tilde{L}_z = 0$ satisfies the semiclassical EBK quantization condition since the angular motion corresponds to a rotation, which does not encounter any caustics and the Maslov index is therefore zero, i.e., $\tilde{L}_z = 2\pi\hbar \cdot m$, $m = 0, 1, 2, \ldots$.

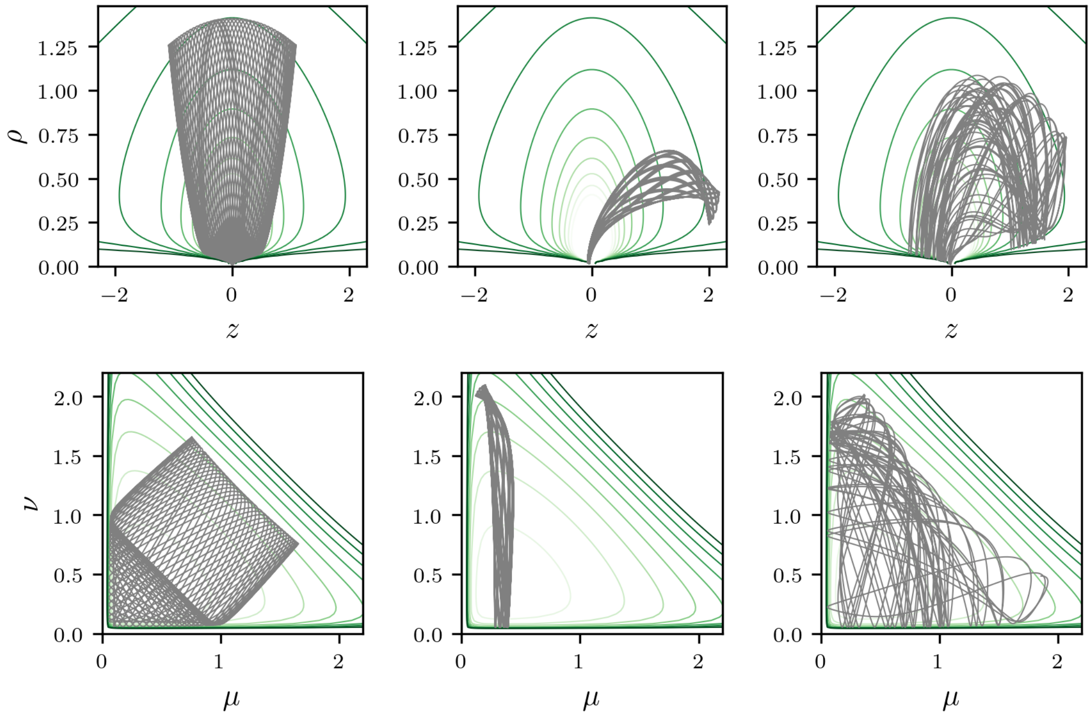

Figure 6.1: The **top row** depicts three types of bound trajectories for the electronic Hamiltonian of hydrogen in a homogeneous magnetic field in cylindrical coordinates, Eq. (6.19). The **left** trajectory represents “rotating” motion, the **middle** one “librating”, and the **right** one shows irregular behavior. The effective potentials are shown as green contour lines. The **bottom row** shows the corresponding trajectories in semiparabolic coordinates, Eqs. (6.38) – (6.41), using Hamiltonian (6.47). The scaled energy and angular momentum were set to $\epsilon = -0.4$ and $\tilde{L}_z = 0.1$, respectively, for all trajectories.

the axis of symmetry of the system, i.e., $\mu = \nu$. Under neglection of the boundaries the regular trajectories then have box-like shape in position space, such as the one depicted in Fig. 4.2. Fig. 6.2 shows purely regular motion up to about $\epsilon = -0.5$. The two types of regular quasiperiodic trajectories are represented as loops around the stable fixed points centered in the three visible islands. The middle island as well as the outer loops around the center fixed point correspond to the type of trajectory depicted in the left column in Fig. 6.1. Comparing to Figs. 5.3 and 5.4, these trajectories resemble the characteristics of non-resonant motion of the coupled Morse oscillators from the previous chapter. The two islands above and below the center one correspond to pairs of the middle type of trajectories of Fig. 6.1. These trajectories resemble the characteristics of the $1 : 1$ resonances of the coupled Morse oscillators. Irregular motion appears near the separatrices around $\epsilon \approx -0.5$ and spreads through phase space with increasing ϵ, swallowing the two resonant islands completely until $\epsilon \approx -0.2$. The non-resonant regular trajectories disappear slightly later between $-0.2 < \epsilon < -0.1$. The order of disappearance of regular motion matches that of the system of the coupled Morse oscillators, i.e., first the resonances, and finally non-resonant motion.

The two types of regular trajectories, non-resonant and $1 : 1$ resonant, have been labeled “rotating” and “librating” (or “vibrating”) trajectories, respectively. [117] These terms arise when representing the dynamics in the new coordinates $R^2 = p_\mu^2 - 2\epsilon\mu^2$, which corresponds to the pseudoenergy of one oscillator, and the angle $\phi = \arctan\bigl(p_\mu/(\sqrt{-2\epsilon}\mu)\bigr)$. Such a representation yields the same two regular modes as those of a pendulum, i.e., libration and rotation.

The neglection of the boundaries at $\mu = 0$ and $\nu = 0$ needs to be considered when evaluating the actions. For librational trajectories, whose SOS can be chosen to lie on the coordinate axes, $\mu, \nu = 0$, the two action variables are simply one half of those of a box-like trajectory moving without the limitations $\mu, \nu \geq 0$. For rotational trajectories, whose SOS can be chosen to lie along $\mu = \nu$ and perpendicular, the action corresponding to the latter remains unchanged, while the former action variable, corresponding to the SOS reaching across the regions $\mu, \nu > 0$ to $\mu, \nu < 0$, needs to be taken as one half of the value of a box-like trajectory also. The Maslov indices for both types of regular trajectories can be obtained by close examination of the top row of Fig. 6.1. Ref. [118] demonstrates that the Maslov indices for both types of trajectories are $\nu_1 = 2$ and $\nu_2 = 4$, according to the number of encounters of caustics along their respective torus paths. Hence, the EBK quantization condition, Eq. (6.21), for this system then becomes

$$\frac{\tilde{S}_1}{\tilde{S}_2} = \frac{n_1 + \frac{1}{2}}{n_2 + 1}\,, \quad n_1, n_2 = 0, 1, \ldots . \tag{6.48}$$

Ref. [118] has related the denominator, $n_2 + 1$, to the principal quantum number of the system, which takes positive integer values.

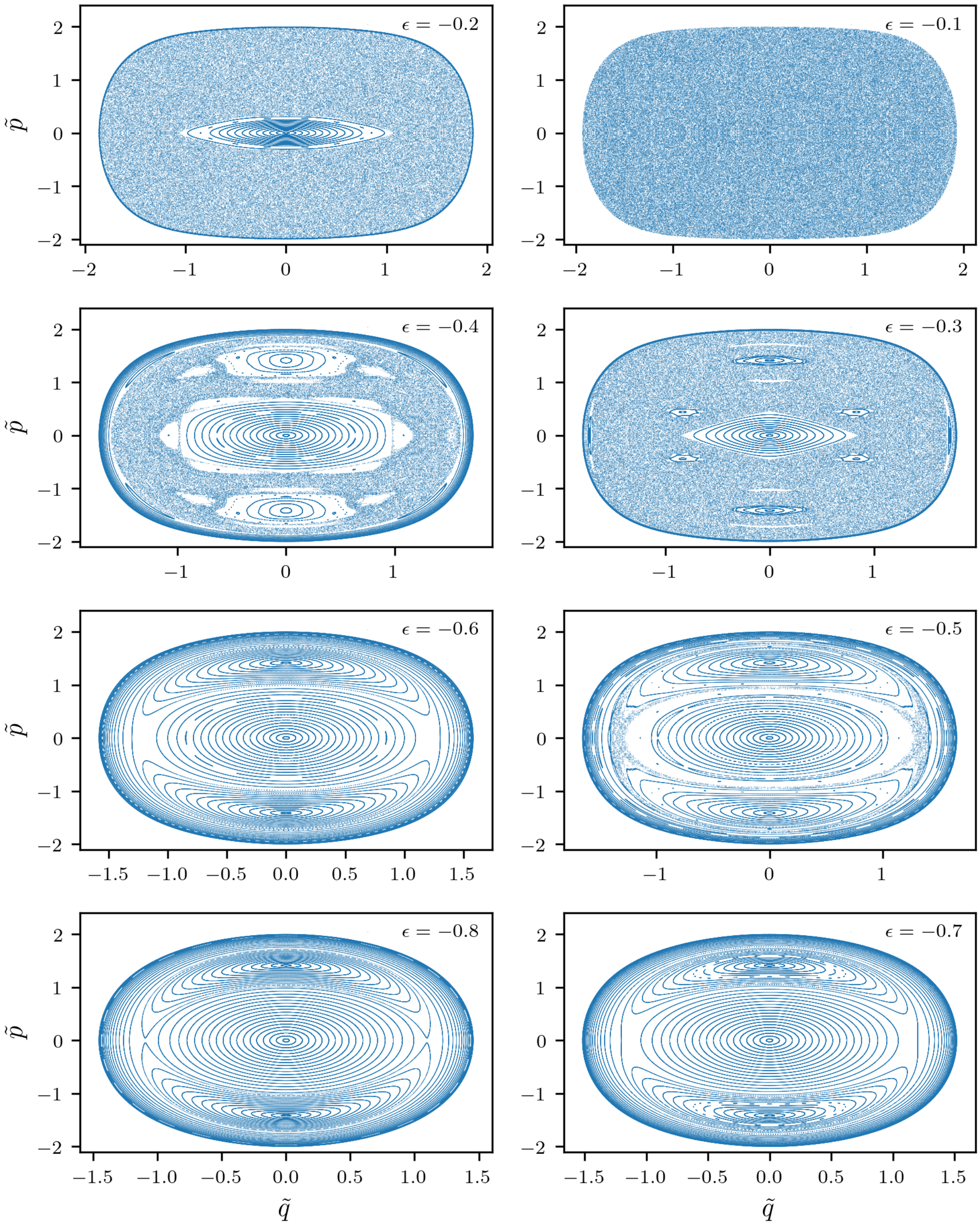

Figure 6.2: Poincaré SOS for the Hamiltonian (6.47) are depicted for various scaled energies, ϵ, and $\tilde{L}_z = 0$. The SOS was chosen along $\mu = \nu$. The values $\tilde{q}$ denote the positions along the section plane and $\tilde{p}$ denote the projections of the momenta tangential to the plane. The hard barriers at $\mu = 0$ and $\nu = 0$ were ignored for the computed trajectories.

6.4 EBK quantization via GPR

The same procedure as in the previous chapter is used to generate the semiclassical EBK spectra, with the difference of the reduced input dimension by one due to the scaling of the coordinates by the magnetic field, Eqs. (6.17) – (6.20). Hence, a GP is trained by regular trajectories whose ratios of actions and scaled energies serve as input variables, $\mathbf{x} = (S_1/S_2, \epsilon)$, and one action serves as output variable, $f(\mathbf{x}) = S_1$. With this choice of input one can easily read off the points where the action ratios fulfill the EBK condition (6.21) or (6.48). Since the numbering of actions is arbitrary as long as it is consistent, we choose the numbering such that $S_1 < S_2$. This leads to action ratios approaching zero as $S_1 \to 0$, which is preferred over approaching infinity as $S_2 \to 0$. As a consequence, the regression manifold, $S_1(S_1/S_2, \epsilon)$, is zero at $S_1/S_2 = 0$. This prior knowledge can be implemented into the GPR by choosing a prior covariance function with the included boundary condition. Following the description in Sec. 2.3.1 and Eqs. (2.17) – (2.19), we apply the non-stationary function $B(\mathbf{x}) = S_1/S_2$ to the simple RBF covariance function, yielding the desired behavior along $S_1/S_2 = 0$. The resulting values for the energies and magnetic field strengths are obtained via Eqs. (6.36) and (6.37).

Fig. 6.3 shows the two GPRs for the librating and rotating motions. 500 quasiperiodic trajectories of each type whose phase space torus is properly densely filled have been used as training data. These training trajectories were chosen from a larger batch of computed trajectories in an AL manner, exploiting the boundary behavior at $S_1/S_2 = 0$. Higher-order quasiperiodic resonances that arise around regular islands due to the Poincaré-Birkhoff scenario are not included into the investigation, as their Poincaré sections are difficult to place and consequently their action variables cannot be easily extracted. Such resonances correspond to the small island chains in Fig. 6.2, such as in the panel for $\epsilon = -0.3$. The optimal length-scales for both GPs suggest that moderate extrapolation to higher scaled energies is justified, since $l_2 \approx 1.0$ and 0.9 are large compared to the increments in ϵ. However, since no more quasiperiodic trajectories are found in the region of large ϵ and S_1/S_2-values, the extrapolated values in that region do suffer from larger uncertainties. The two types of trajectories ought to be treated separately and no extrapolation beyond their respective S_1/S_2-boundaries is conducted. The two black lines, dashed and solid, mark the values of first appearance of irregularity and completely chaotic behavior, respectively, according to Ref. [115]. The very low noise hyperparameter implies that all training points lie precisely on the regression surface.

The resulting semiclassical energy spectra are depicted in Fig. 6.4 for selected symmetric states, while Fig. C.1 in Appendix C shows all states. The full quantum spectra are shown in gray for comparison. Fig. 6.4 shows only symmetric states, having even azimuthal quantum numbers, l, as well as the semiclassical EBK energies for quantum numbers $n_2 = 9$ and 11, corresponding to the principal quantum numbers 10 and 12. The blue energies correspond to librational trajectories and the orange energies to rotational ones. Both constitute the full spectra and follow diabatic patterns with increasing magnetic field strength, B^2, i.e., all avoided crossings are traversed. The blue and orange colored circles represent EBK energies resulting from interpolation between computed trajectories within the regular phase space, whereas crosses represent extrapolated energies. For both groups of n_2-quantum numbers the uppermost states with steepest slopes correspond to rotational trajectories with small action ratios, whose regular islands persist up to scaled energies of $\epsilon \approx -0.15$, as seen in Fig. 6.3. Rotational states with smaller slopes correspond to larger action ratios and

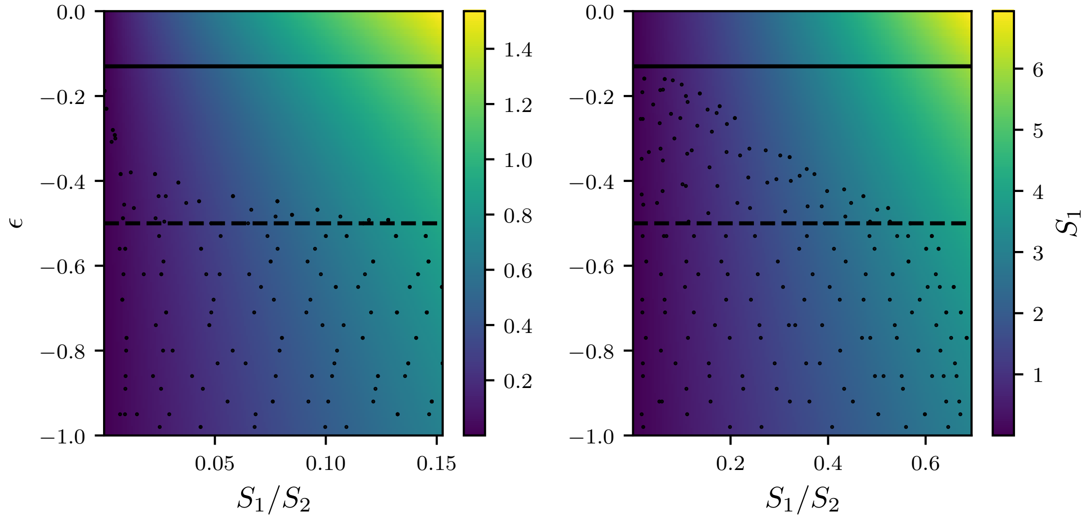

Figure 6.3: The GPRs for the actions S_1 (colormap) as functions of the action ratios and scaled energies, ϵ, trained by various quasiperiodic trajectories (dots). The **left** panel corresponds to librating and the **right** panel to rotating trajectories. The optimized hyperparameters are $(\sigma_f, l_1, l_2, \sigma_n) \approx \left(32, 0.04, 1.0, 10^{-5}\right)$ for the left GP and $\left(7, 0.87, 0.9, 10^{-4}\right)$ for the right GP. The black dashed and solid lines mark the region in which both regular and irregular dynamics coexist, $-0.5 \leq \epsilon \leq -0.13$. [115]

extrapolate into the irregular regime earlier. On the contrary, blue librational states with smaller slopes correspond to smaller action ratios, which persist up to larger scaled energies.

Comparing to the system of two coupled Morse oscillators from the previous chapter, we have again computed the full diabatic energy spectrum, this time with two two-dimensional GPRs instead of one three-dimensional GPR, each using merely 500 regular trajectories. Again, the system is treated as an integrable approximation, since regular dynamics is extrapolated.

One can observe a strong connection between irregular classical dynamics and the repelling strength at avoided energy level crossings. [119, 120] Each semiclassical diabat corresponds to a set of EBK quantum numbers, (n_1, n_2), and thus to a ratio of actions, S_1/S_2. Those diabats that correspond to trajectories with low action ratio, such as the uppermost states of the two bunches in Fig. 6.4, are more easily revealed by the naked eye since they jump only weakly avoided crossings up to large values of ϵ. Diabats corresponding to larger action ratios jump over much stronger avoided crossings, such as the lowest rotational and highest librational states of the two bunches in Fig. 6.4. These states exhibit chaotic classical dynamics already at low scaled energies, $\epsilon \approx -0.5$. Regarding the full spectra, as irregular motion takes over with increasing magnetic field strength or scaled energy, diabatic patterns are less easily revealed due to the increasing strength of avoided crossings. These observations relate to the known Poisson or Wigner distributions of nearest neighbor level spacings of systems with classically regular or ergodic dynamics, respectively. [109, 121]

As it was noticed in the previous chapter, the energy landscape as function of the classical action variables shows rather flat and smooth characteristics that allow for controlled extrapolation into the classically irregular regime. This extrapolation works better the larger the length-scales of the trained GPs are. Otherwise, higher uncertainties will occur. In Fig. 6.4

the $3\hat{\sigma}$-uncertainties of the posterior GPs are shown as errorbars. These are aligned with lines of constant ϵ.

When looking at the full spectra in Fig. C.1 in Appendix C, one can see that not all semiclassical EBK energies correctly describe the true quantum results, apart from the disagreements at the avoided crossings. This is especially the case for rotational and librational states at higher action ratios due to their early vanishing into the irregular regime. The predictive accuracies of the two extrapolating GPs are weaker than that of the GP in the previous chapter. The smaller length-scales do not permit extrapolation far beyond the training data, suggesting that the GPs fall short of discovering a long-ranging trend of the latent function, which seems to strongly increase with larger S_1/S_2 and ϵ-values. This problem could be solved by considering alternative covariance functions to the simple RBF covariance used here for the prior GP, such as the more sophisticated SM covariance function discussed in Sec. 2.3. Further shall be noted that even at very low scaled energies not all eigenstates are found using the semiclassical EBK method. These missing states lie very close to the separatrix between rotational and librational motion. Same states, however, appear if the angular momentum, $\tilde{L}_z$, is increased, causing librational motion to disappear. [116] This effect is discussed in Ref. [122].

6.4.1 Scaling to large data sets in higher dimensions

Finally, we would like to address the task of upscaling our presented method to problems of higher dimensions. As was extensively discussed in Sec. 2.5, the number of training points states a bottleneck for GP inference and data sets containing more than roughly 1000 points become prohibitive. Hence, scaling methods, such as the ones discussed in Sec. 2.5, ought to be employed.

The exact scaling methods presented in Sec. 2.5.1, i.e., methods exploiting Toeplitz and Kronecker structure, are the most desirable methods for upscaling. Though in numerical experiments training data can often be generated on grids, the extracted actions necessary for the EBK quantization will not exhibit lattice structure, and hence, such methods cannot be directly employed. Consequently, approximate inference methods, described in Sec. 2.5.2, are the sensible choice for the objective of upscaling. However, it was demonstrated in Sec. 2.5.3 that length-scales larger than average distances between training points can lead to very slow convergence of the CG method, which is employed in all highly scalable inducing point methods.

For this reason we suggest a simple “thinning” of any dense data set, i.e., removing surplus data points from the training set, with the goal of generating a smaller, more dilute data set without losing information about the latent input-output relationship. Dilute data sets can be efficiently gathered using an AL method (see Sec. 2.6), especially when non-stationary features, such as the boundary behavior used for the GPs in this chapter, are incorporated. Such thinning should be justified whenever a posterior GP observes length-scales much larger than the average distances between training points, while having a very low noise hyperparameter. This leads to faster inference and allows for the use of upscaling methods in higher dimensions. In fact, we have performed such thinning on both training data sets used for both systems, the coupled Morse oscillators (Chap. 5) and the hydrogen atom in a magnetic field (this chapter). We experienced that due to the smoothness and flatness of the regression landscapes, diluted training data sets suffice and allow for larger length-scales without the need of adding diagonal noise for numerical stability.

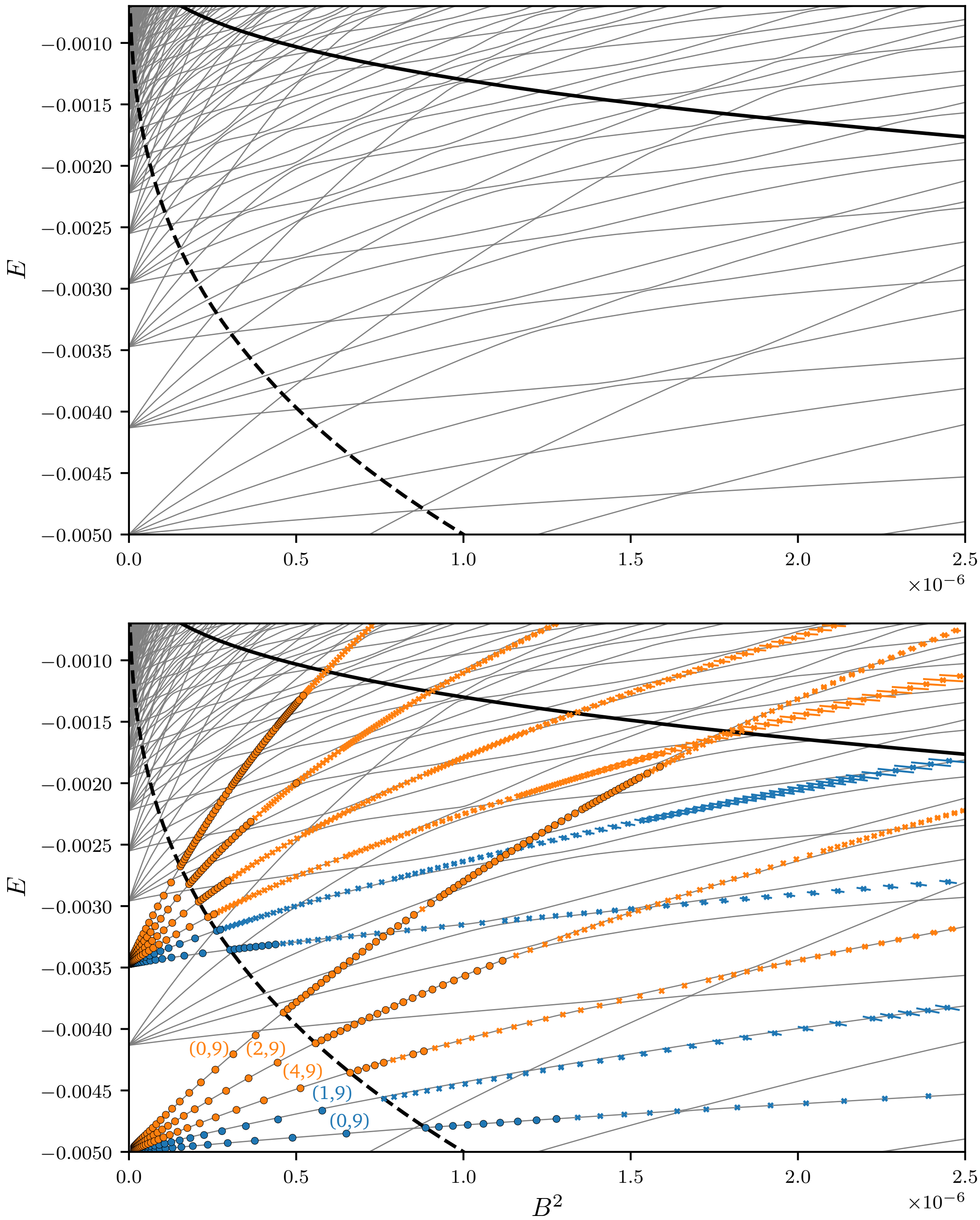

Figure 6.4: The energy spectra of symmetric states (even azimuthal quantum number, l) of hydrogen in a homogeneous magnetic field are shown. The exact quantum spectra are depicted as gray lines. The semiclassical EBK spectra are partly illustrated as blue markers for librational and orange for rotational trajectories. The EBK quantum numbers, (n_1, n_2), for the lower bunch of states are given. Circles depict energies which are interpolated between computed trajectories within regular phase space, while crosses depict extrapolated energy values. The $3\hat{\sigma}$-uncertainty range is shown as errorbars. The black dashed and solid lines show scaled energy values of $\epsilon = -0.5$ and -0.13, respectively, marking the region of mixed phase space. Both E and B are given in atomic units. In SI units B is given in units of $B_0 = 2.35 \cdot 10^5 T$.

7 Conclusion

In this work we have employed the state of the art supervised ML technique of GPR to efficiently extract diabatic energy manifolds from spectra whose underlying physical systems possess classical phase spaces with both regular and chaotic dynamics. For this, GPs have been extensively introduced. Their particular strengths were thoroughly discussed, which include most of all the interpretability of the complete model with its few hyperparameters. They can be tuned in a principled manner using the maximum likelihood estimation method and adjusted to meet specific demands. Their optimal values indicate precisely the properties of the regression manifold and to what extend it can be extrapolated into unknown regions. A unique feature of GPs are the naturally provided reliable uncertainty measures for all predictions, which are essential for extrapolation. We claim that extrapolations should always be provided with predictive probabilities, otherwise they would be meaningless. Hence, GPs are uniquely suited for the task.

The interpretability of GPs facilitates implementations of non-stationary features into the models, such as boundary behaviors. We have constructed modifications to a commonly used covariance function to match non-stationary properties in order to enhance predictions as well as active learning procedures. The effects were illustrated on simple physical systems.

For the purpose of tackling higher-dimensional problems procedures for upscaling have been discussed. It was shown that upscaling methods involving the CG method can suffer from large length-scales in the covariance functions. This has to be accounted for when training data sets are generated for extrapolation purposes.

GPs have previously been used for pattern discovery and extrapolation. [8, 9] Inspired by this finding, we used GPR for the isolation of diabatic energy manifolds, which emerge as patterns in generic energy spectra when a parameter of the Hamiltonian is varied. Such diabatic manifolds are continuations of energy levels which pass through avoided crossings and their correspondingly constructed wave functions conserve their characteristics. While this procedure has shown to be capable of extracting diabatic patterns, a collection of fully transformed diabatic spectra ought to be generated in a semiclassical way using the EBK quantization condition.

The EBK method of isolating regular trajectories whose phase space torus fulfills certain conditions in order to find semiclassical eigenenergies of a given Hamiltonian system serves as an approximate tool to generate energy spectra. This semiclassical treatment ignores coupling effects between individual states which lead to avoided crossings, e.g., due to tunneling, and hence yields diabatic manifolds. Instead of root finding individual trajectories which fulfill the quantization conditions, we took up the idea of Ref. [84] and applied GPR for the inter- and extrapolation of classical action variables, as well as other parameters of the Hamiltonian, using training data from arbitrary regular trajectories. This allows for the extraction of all energies satisfying the EBK conditions with one single regression. The optimal length-scales along the directions of the action variables suggest smooth and flat energy surfaces with long-ranged correlations and therefore allow for controlled extrapolation beyond the region of regular dynamics. Hence, we produced extended diabatic spectra for two systems, namely two coupled Morse oscillators and the hydrogen atom in a magnetic field.

The first of these two systems has been classically explored throughout the whole range of mass ratios of the outer two identical atoms and the center atom. Analogously to its harmonic equivalent the system exhibits various resonances at different regions of the mass ratio parameter. It was shown that these resonances appear most prominently where the true eigenspectra undergo strongly avoided crossings. Their contributions to the semiclassical spectra yield "corrections" for the repelled energy levels, i.e., their energies lie along those states which are most strongly repelled. This phenomenon has already been investigated in Ref. [106]. Ignoring the resonant trajectories, bound diabatic spectra have been generated using only a few hundred non-resonant trajectories. These spectra range far into the classically irregular regime of the system, where they still succeed in producing precise diabatic manifolds.

For the system of hydrogen in a magnetic field the scaling of coordinates by the magnetic field strength allows for the consideration of ratios of action variables, canceling out the direct dependence on the magnetic field strength while providing a series of EBK energies for each single trajectory that fulfills the quantization condition. This reduces the number of classical parameters effectively by one. The two regular types of trajectories, which can be identified to be non-resonant and 1 : 1 resonant, produce complement parts of the full diabatic spectra with increasing magnetic field strength. Extrapolation into irregular regions reveal continuations of diabats where they are no longer well-defined.

The presented procedure for extrapolation bypasses the shortcomings of the EBK method to some extend, namely the limitation to regular dynamics, while exploiting the advantage that regular, quasiperiodic trajectories are easily found and can be efficiently and parallely propagated. Earlier methods have been brought forward to extend the EBK method into moderately irregular regions. [64, 123] On the other hand, semiclassical methods relying on summation over periodic orbits [65, 66] are extendable to fully chaotic regions, though the isolation of periodic orbits is a demanding task and the method does not assign quantum numbers to individual states, since it computes densities of states. A sophisticated method is presented in Ref. [102], which constructs approximate integrable Hamiltonians for generic systems with mixed phase spaces. The combination with our method promises improvements in convergence and accessibility to higher dimensions.

Our extrapolation method enables the generation of diabatic spectra in irregular regions of non-integrable systems. Hence, one may consider the diabatic spectra constructed from our GPR method as an integrable approximation to the system with implicit additional constants of motion.

7.1 Outlook

Looking forward, the application of GPR for the computation of semiclassical EBK energy spectra as functions of one or more variable parameters of the Hamiltonian can be further improved. In this work the Poincaré SOS method for the extraction of action variables is used, which is limited to low-dimensional systems with two to three degrees of freedom. Additionally, the treatment of resonant trajectories with this method demands considerable effort. To alleviate both limitations, other methods, such as the Fourier transform approach from Refs. [64,83,124], ought to be employed. While allowing for more degrees of freedom, the treatment of resonant trajectories becomes straightforward. [124] Further, the extracted approximate action variables coming from irregular “vague tori” from Ref. [64] can be added to the training data set with increased individual variable noise, which GPR allows for. Using the Fourier transform approach the Gaussian noise assumption on the training data, Eq. (2.26), becomes much more suited, whereas for the Poincaré SOS method the action variables always converge from below.

Regarding resonant trajectories, it was shown that their contributions to the EBK spectra lie on the strongly repelled energy surfaces. Although those trajectories only fulfill the quantization condition in a small, confined area, their extrapolation could allow for completely corrected spectra, exhibiting more avoided crossings. In fact, the resonant character can still be seen for many periods in close-lying non-resonant trajectories.

When applying upscaling methods, GPR can be used for high-dimensional regression. Since the time for convergence of trajectories grows exponentially with the degrees of freedom of the system, the propagation of trajectories can become a bottleneck for semiclassical methods. As a consequence, fewer trajectories may need to suffice, which is one of the strong points of GPR. In addition to the action variables of the trajectories, more than one parameter of the Hamiltonian can be added to the input space of the GP, resulting in energy surfaces as functions of all parameters.

A Derivations

A.1 The arithmetic mean as a maximum likelihood estimator

One well-known maximum likelihood estimator is the one for the mean of a univariate Gaussian PDF, Eq. (2.6). The logarithm of the likelihood, Eq. (2.42), is given as

$$\mathcal{L}(\mu; y_1, \ldots, y_K) \equiv \log L(\mu; y_1, \ldots, y_K) = -\frac{1}{2\sigma^2} \sum_{i=1}^{K} (y_i - \mu)^2 + c' \tag{A.1}$$

and its maximum is found at

$$\left.\frac{\mathrm{d}\mathcal{L}(\mu)}{\mathrm{d}\mu}\right|_{\hat{\mu}} = \frac{2}{2\sigma^2} \sum_{i=1}^{K} (y_i - \hat{\mu}) = \frac{1}{\sigma^2} \left(\sum_{i=1}^{K} y_i - K\hat{\mu} \right) \overset{!}{=} 0, \tag{A.2}$$

yielding for the maximum likelihood estimator

$$\hat{\mu} = \frac{1}{K} \sum_{i=1}^{K} y_i \equiv \bar{y}, \tag{A.3}$$

which is the arithmetic mean.

A.2 Matrix-vector product for Kronecker matrices

The product of a Kronecker matrix, Eq. (2.47), and a vector can be efficiently computed in $\mathcal{O}\big(DN^{\frac{D+1}{D}}\big)$ operations instead of $\mathcal{O}(N^2)$ for the generic case. [27, 28] Using the property

$$(\mathbf{A} \otimes \mathbf{B}) \operatorname{vec}(\mathbf{V}) = \operatorname{vec}\Big(\mathbf{B}\mathbf{V}\mathbf{A}^\top\Big), \tag{A.4}$$

where $\operatorname{vec}(\mathbf{V})$ is the column-wise concatenation that reshapes the matrix $\mathbf{V}$ into a vector, one can rewrite the matrix-vector product. The product of a Kronecker matrix, $\mathbf{K} \in \mathbb{R}^{N \times N}$

and a vector, $\mathbf{b} \in \mathbb{R}^N$, is then given by

$$\mathbf{Kb} = \left(\bigotimes_{d=1}^{D} \mathbf{A}_d\right) \mathrm{vec}(\mathbf{B}) \tag{A.5}$$

$$= \mathrm{vec}\left(\mathbf{A}_D \mathbf{B} \left(\bigotimes_{d=1}^{D-1} \mathbf{A}_d\right)^{\top}\right) \tag{A.6}$$

$$= \mathrm{vec}\left(\left[\mathrm{vec}^{-1}\left(\mathrm{vec}\left(\left(\bigotimes_{d=1}^{D-1} \mathbf{A}_d\right)(\mathbf{A}_D\mathbf{B})^{\top}\right)\right)\right]^{\top}\right) \tag{A.7}$$

$$= \mathrm{vec}\left(\left[\mathrm{vec}^{-1}\left(\mathrm{vec}\left(\left(\bigotimes_{d=1}^{D-1} \mathbf{A}_d\right)(\mathbf{A}_D\mathbf{B})^{\top}\mathbf{I}_{N^{1/D}}\right)\right)\right]^{\top}\right) \tag{A.8}$$

$$= \mathrm{vec}\Bigg(\Bigg[\mathrm{vec}^{-1}\underbrace{\left(\mathbf{I}_{N^{1/D}} \otimes \left(\bigotimes_{d=1}^{D-1} \mathbf{A}_d\right)\mathrm{vec}\left((\mathbf{A}_D\mathbf{B})^{\top}\right)\right)}_{\text{repeat from line (A.6)}}\Bigg]^{\top}\Bigg) \tag{A.9}$$

$$= \dots \tag{A.10}$$

$$= \mathrm{vec}\Big([\mathbf{A}_1 \dots [\mathbf{A}_{D-1}[\mathbf{A}_D\mathbf{B}]^{\top}]^{\top}]^{\top}\Big), \tag{A.11}$$

where $\mathbf{B}$ is the matrix of dimensions $N^{\frac{1}{D}} \times N^{\frac{D-1}{D}}$ formed from $\mathbf{b}$, and $\mathrm{vec}^{-1}(\mathrm{vec}(\mathbf{A})) = \mathbf{A}$. In lines (A.6) and (A.9) we used property (A.4). Repeating at line (A.9) over all D dimensions and noting that $(\bigotimes_{d=1}^{D} \mathbf{I})\mathbf{b} = \mathbf{b}$ yields above simplified expression, where the bracket notation denotes transpose and reshape, i.e. $[\mathbf{AB}]^{\top} = \mathrm{reshape}((\mathbf{AB})^{\top})$. This reshaping occurs due to the generation of matrix $[\mathbf{AB}]^{\top}$ out of $\mathrm{vec}([\mathbf{AB}]^{\top})$ to serve equation (A.4), where the matrix $\mathbf{AB}^{\top}$ lost its original shape.

A.3 No degenerate bound eigenstates in one dimension

Considering a regular (or non-singular) [48] one-dimensional bound system, we start with the assumption that there are two linearly independent eigenstates of a one-dimensional Hamiltonian with the same energy,

$$\left[-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V(x)\right]\psi_1 = E\psi_1, \tag{A.12}$$

$$\left[-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V(x)\right]\psi_2 = E\psi_2. \tag{A.13}$$

Multiplying the first equation by ψ_2 and the second by ψ_1 and subtracting one from the other one gets

$$\psi_1\frac{\mathrm{d}^2}{\mathrm{d}x^2}\psi_2 - \psi_2\frac{\mathrm{d}^2}{\mathrm{d}x^2}\psi_1 = 0. \tag{A.14}$$

Close observation reveals that one derivative can be factored out,

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\psi_1\frac{\mathrm{d}}{\mathrm{d}x}\psi_2-\psi_2\frac{\mathrm{d}}{\mathrm{d}x}\psi_1\right)=0. \tag{A.15}$$

Integrating both sides of above equation yields

$$\psi_1\frac{\mathrm{d}}{\mathrm{d}x}\psi_2-\psi_2\frac{\mathrm{d}}{\mathrm{d}x}\psi_1=C \tag{A.16}$$

for any value of x. For bound states, the boundary condition for $x\to\pm\infty$ is $\psi_{1,2}\to 0$. Therefore the constant on the right hand side must be zero, $C=0$. Consequently one gets

$$\begin{aligned}
\psi_1\frac{\mathrm{d}}{\mathrm{d}x}\psi_2&=\psi_2\frac{\mathrm{d}}{\mathrm{d}x}\psi_1 && \text{(A.17)}\\
\Rightarrow\frac{\frac{\mathrm{d}}{\mathrm{d}x}\psi_2}{\psi_1}&=\frac{\psi_2\frac{\mathrm{d}}{\mathrm{d}x}}{\psi_2} && \text{(A.18)}\\
\Rightarrow\frac{\mathrm{d}}{\mathrm{d}x}\log\psi_1&=\frac{\mathrm{d}}{\mathrm{d}x}\log\psi_2 && \text{(A.19)}\\
\Rightarrow\log\psi_1&=\log\psi_2+\tilde{C} && \text{(A.20)}\\
\Rightarrow\psi_1&=e^{\tilde{C}}\psi_2, && \text{(A.21)}
\end{aligned}$$

where $\psi_1,\psi_2\neq 0$. Evidently, since $\psi_1\sim\psi_2$, both eigenstates are not linearly independent, which contradicts the assumption.

A.4 Von Neumann-Wigner theorem

The von Neumann-Wigner theorem [49] states that for general perturbations, described by a Hermitian matrix depending on a number of parameters, $W(\lambda_1,\ldots,\lambda_{N_\lambda})$, one needs at least three parameters, $N_\lambda=3$, to be able to locate a crossing of two eigenenergies. The effective Hamiltonian matrix in the subspace of the two considered eigenstates can be written as the sum of the unperturbed Hamiltonian plus the perturbation in the basis of the unperturbed eigenstates,

$$H_0+W(\lambda_1,\ldots,\lambda_{N_\lambda})=\begin{pmatrix}E_1&0\\0&E_2\end{pmatrix}+\begin{pmatrix}W_{11}&W_{12}\\W_{12}^\dagger&W_{22}\end{pmatrix}. \tag{A.22}$$

Expressing the off-diagonal elements as $W_{12}=x+iy$, the eigenenergies of above matrix are

$$E_\pm=\frac{1}{2}\left(E_1+W_{11}+E_2+W_{22}\pm\sqrt{(E_1+W_{11}-E_2-W_{22})^2+4(x^2+y^2)}\right). \tag{A.23}$$

In order for both energies to be degenerate, the square root must vanish, i.e.,

$$\begin{aligned}
E_1+W_{11}(\lambda_1,\ldots,\lambda_{N_\lambda})&\overset{!}{=}E_2+W_{22}(\lambda_1,\ldots,\lambda_{N_\lambda}), && \text{(A.24)}\\
x(\lambda_1,\ldots,\lambda_{N_\lambda})&\overset{!}{=}0, && \text{(A.25)}\\
y(\lambda_1,\ldots,\lambda_{N_\lambda})&\overset{!}{=}0. && \text{(A.26)}
\end{aligned}$$

For all three equations to hold, at least three independent parameters must be tuned. If only three parameters exist, i.e., $N_\lambda = 3$, all three equations can be satisfied only at isolated parameter sets. Such points in parameter space lead to conical intersections. If four parameters are available, a one-dimensional manifold in parameter space will satisfy above conditions.

For the special case of real-valued perturbations, i.e., $W_{12} = x$, the above conditional equations reduce to two. In this case, only two parameters are required to locate degeneracies.

Another special case appears for systems that possess symmetries. If the two considered eigenenergies correspond to eigenstates of different symmetry, the off-diagonal matrix elements vanish identically, $W_{12} = 0$. [125] In this case, only one conditional equation remains and crossings can appear by varying a single parameter.

A.5 Poincaré surface of section method

We use a linear interpolation between two neighboring points, i and $i+1$, if they cross the chosen SOS. For SOS along $q_2 \equiv y = y_0,\ p_2 > 0$ and $q_1 \equiv x = x_0,\ p_1 > 0$ we find the $\tilde{q}_1$ and $\tilde{q}_2$ locations to be

$$\tilde{q}_1 = x_i + (y_0 - y_i)\frac{x_{i+1} - x_i}{y_{i+1} - y_i}, \tag{A.27}$$

$$\tilde{q}_2 = y_i + (x_0 - x_i)\frac{y_{i+1} - y_i}{x_{i+1} - x_i}. \tag{A.28}$$

For the momenta tangential to the SOS we assume $p(t) = mt + n$, where $p(t_i) = p_i$, to yield

$$p(t) = p_i + (p_{i+1} - p_i)\frac{t - t_i}{t_{i+1} - t_i}, \tag{A.29}$$

and for the values in the SOS

$$\tilde{p}_1 = p_i + (p_{i+1} - p_i)\frac{\tilde{q}_1 - x_i}{x_{i+1} - x_i} = p_i + (p_{i+1} - p_i)\frac{y_0 - y_i}{y_{i+1} - y_i}, \tag{A.30}$$

$$\tilde{p}_2 = p_i + (p_{i+1} - p_i)\frac{\tilde{q}_2 - y_i}{y_{i+1} - y_i} = p_i + (p_{i+1} - p_i)\frac{x_0 - x_i}{x_{i+1} - x_i}. \tag{A.31}$$

For a general linear section planes of the form $y = mx + n$ we find the values for the positions of the SOS points to be

$$\tilde{q} = \sqrt{\tilde{x}^2 + \tilde{y}^2}, \tag{A.32}$$

$$\tilde{x} = \frac{x_i \frac{y_{i+1} - y_i}{x_{i+1} - x_i} - y_i + n}{\frac{y_{i+1} - y_i}{x_{i+1} - x_i} - m}, \tag{A.33}$$

$$\tilde{y} = m\tilde{x} + n, \tag{A.34}$$

and the momenta

$$\tilde{p} = \frac{1}{\sqrt{1 + m^2}} \begin{pmatrix} p_{\tilde{x}} \\ p_{\tilde{y}} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ m \end{pmatrix} = \frac{p_{\tilde{x}} + m\, p_{\tilde{y}}}{\sqrt{1 + m^2}}. \tag{A.35}$$

B Two kinetically coupled Morse oscillators

The Morse oscillator

A linearly aligned, triatomic molecule possesses two vibrational degrees of freedom, namely the two distances between the center atom and the outer two atoms, $x_{1,2}$. The two corresponding potentials can be approximated via the anharmonic Morse potential,

$$V_{\mathrm{Mor}}(x) = D_e \left(1 - e^{-ax}\right)^2 , \tag{B.1}$$

where D_e is the dissociation energy and a the stiffness parameter of the potential. The eigenfunctions of the single Morse oscillator with Hamiltonian

$$H = \frac{p^2}{2\mu} + V_{\mathrm{Mor}}(x) \tag{B.2}$$

are known to be [92]

$$\psi_n(x) = N_n z^{k-n-\frac{1}{2}} e^{-\frac{1}{2}z} L_n^{(2k-2n-1)}(z), \tag{B.3}$$

where

$$n = 0, 1, \ldots, \left[k - \frac{1}{2}\right], \tag{B.4}$$

with $[\xi]$ denoting the maximum integer $\leq \xi$, $k = \sqrt{2\mu_r D_e}/a$, $z(x) = 2ke^{-ax}$ $(0 < z < \infty)$, and $L_n^{(\alpha)}$ are associated/generalized Laguerre polynomials. To show orthogonality we compute the overlap

$$\Omega_{m,n} \equiv \int_{-\infty}^{\infty} \mathrm{d}x\, \psi_m(x)\psi_n(x) \tag{B.5}$$

$$= N_m N_n \int_{-\infty}^{\infty} \mathrm{d}x\, z^{2k-m-n-1} e^{-z} L_m^{(2k-2m-1)}(z) L_n^{(2k-2n-1)}(z) \tag{B.6}$$

$$= \frac{N_m N_n}{a} \int_{0}^{\infty} \mathrm{d}z\, z^{2k-m-n-2} e^{-z} L_m^{(2k-2m-1)}(z) L_n^{(2k-2n-1)}(z). \tag{B.7}$$

Using the property of the associated Laguerre polynomials,

$$L_n^{(\beta)}(z) = \sum_{j=0}^{n} \binom{\beta - \alpha + j - 1}{j} L_{n-j}^{(\alpha)}(z), \tag{B.8}$$

one can express

$$L_m^{(2k-2m-1)}(z) = \sum_{i=0}^{m} \binom{-m+n+i}{i} L_{m-i}^{(2k-m-n-2)}(z), \tag{B.9}$$

$$L_n^{(2k-2n-1)}(z) = \sum_{j=0}^{n} \binom{m-n+j}{j} L_{n-j}^{(2k-m-n-2)}(z). \tag{B.10}$$

Here, the only terms that do not vanish are those whose binomial coefficients are non-zero. (Note that binomial coefficients, $\binom{n}{k}$, are also non-zero for negative n and positive k.) Then, the overlap simplifies:

$$\begin{aligned}
\Omega_{m,n} = {} & \frac{N_m N_n}{a} \sum_{i=0}^{m} \binom{-m+n+i}{i} \sum_{j=0}^{n} \binom{m-n+j}{j} \\
& \times \int_0^\infty \mathrm{d}z\, z^{2k-m-n-2} e^{-z} L_{m-i}^{(2k-m-n-2)}(z) L_{n-j}^{(2k-m-n-2)}(z) && \text{(B.11)} \\
= {} & \frac{N_m N_n}{a} \sum_{i=0}^{m} \binom{-m+n+i}{i} \sum_{j=0}^{n} \binom{m-n+j}{j} \frac{\Gamma(2k-m-1-j)}{(n-j)!} \delta_{m-i,n-j} && \text{(B.12)} \\
= {} & \frac{N_m N_n}{a} \sum_{i=0}^{m} \binom{-m+n+i}{i} \sum_{\tilde{j}=0}^{n} \binom{m-\tilde{j}}{n-\tilde{j}} \frac{\Gamma(2k-m-n-1+\tilde{j})}{\tilde{j}!} \delta_{m-i,\tilde{j}} && \text{(B.13)} \\
= {} & \frac{N_m N_n}{a} \sum_{i=0}^{m} \binom{n-m+i}{i} \binom{i}{n-m+i} \frac{\Gamma(2k-n-1-i)}{(m-i)!}. && \text{(B.14)}
\end{aligned}$$

Here we used the orthogonality of the associated Laguerre polynomials, [126]

$$\int_0^\infty \mathrm{d}z\, z^\alpha e^{-z} L_n^{(\alpha)}(z) L_m^{(\alpha)}(z) = \frac{\Gamma(n+\alpha+1)}{n!} \delta_{n,m}. \tag{B.15}$$

For $m \neq n$ all of the above terms vanish, yielding 0. For $m = n$ we get

$$\begin{aligned}
\Omega_{n,n} &= \int_{-\infty}^{\infty} \mathrm{d}x\, \psi_n^2(x) && \text{(B.16)} \\
&= \frac{N_n^2}{a} \sum_{i=0}^{n} \frac{\Gamma(2k-n-1-i)}{(n-i)!} && \text{(B.17)} \\
&= \frac{N_n^2}{a} \sum_{\tilde{i}=0}^{n} \frac{\Gamma(2k-2n-1+\tilde{i})}{\tilde{i}!} && \text{(B.18)} \\
&= \frac{N_n^2}{a} \Gamma(2k-2n-1) \sum_{i=0}^{n} \binom{2(k-n-1)+i}{i} && \text{(B.19)} \\
&= \frac{N_n^2}{a} \frac{\Gamma(2k-n)}{n!(2k-2n-1)} \overset{!}{=} 1, && \text{(B.20)}
\end{aligned}$$

where the two last steps follow Ref. [90]. Hence, we find

$$N_n = \sqrt{\frac{a(2k-2n-1)n!}{\Gamma(2k-n)}}. \tag{B.21}$$

Two kinetically coupled Morse oscillators

The potential energies of the two vibrational degrees of freedom of a symmetric, linearly aligned, triatomic molecule, as shown in Fig. 5.1, left, can be described by two Morse oscillators, Eq. (B.1), where the arguments, x, correspond to the dispositions from the equilibrium distances between two neighboring atoms. If the positions of the three atoms are x_1, x_2 and x_3 in a laboratory frame of reference, then the arguments of the Morse potentials are $q_1 = x_2 - x_1$ and $q_2 = x_3 - x_2$. The kinetic energy of the molecule is

$$T = \frac{m_1}{2}\dot{x}_1^2 + \frac{m_2}{2}\dot{x}_2^2 + \frac{m_3}{2}\dot{x}_3^2, \tag{B.22}$$

where $m_1 = m_3 = m$ is the mass of the outer atoms and $m_2 = M$ is the mass of the central atom. In order to express the kinetic energy in terms of the internal dispositions, we employ

$$\dot{q}_1 = \dot{x}_2 - \dot{x}_1, \tag{B.23}$$

$$\dot{q}_2 = \dot{x}_3 - \dot{x}_2. \tag{B.24}$$

For further convenience, we transform into a coordinate system in which the total momentum vanishes, i.e.,

$$0 = m\dot{x}_1 + M\dot{x}_2 + m\dot{x}_3. \tag{B.25}$$

Combined we have

$$\begin{pmatrix} \dot{q}_1 \\ \dot{q}_2 \\ 0 \end{pmatrix} = \underbrace{\begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ m & M & m \end{pmatrix}}_{U} \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{pmatrix}, \tag{B.26}$$

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{pmatrix} = U^{-1} \begin{pmatrix} \dot{q}_1 \\ \dot{q}_2 \\ 0 \end{pmatrix}. \tag{B.27}$$

Inserting these expressions into Eq. (B.22) we get [127]

$$T = \frac{1}{2}\left(g\dot{q}_1^2 + 2h\dot{q}_1\dot{q}_2 + g\dot{q}_2^2\right), \tag{B.28}$$

where

$$g = \frac{m(m+M)}{2m+M}, \tag{B.29}$$

$$h = \frac{m^2}{2m+M}. \tag{B.30}$$

Finally, we would like to express the kinetic energy in terms of the momenta canonically conjugate to the coordinates q_1 and q_2. They are defined by the partial derivatives of the Lagrangian $\mathcal{L}(q_1, q_2, \dot{q}_1, \dot{q}_2)$,

$$p_i = \frac{\partial \mathcal{L}}{\partial \dot{q}_i} = \frac{\partial}{\partial \dot{q}_i}(T - V), \tag{B.31}$$

yielding

$$p_1 = g\dot{q}_1 + h\dot{q}_2, \tag{B.32}$$
$$p_2 = g\dot{q}_2 + h\dot{q}_1. \tag{B.33}$$

Using these momenta the kinetic energy term takes the form

$$T = \frac{1}{2}\left(\frac{p_1^2}{\mu_r} - \frac{2}{M}p_1p_2 + \frac{p_2^2}{\mu_r}\right), \tag{B.34}$$

where $\mu_r = \frac{mM}{m+M}$ is the reduced mass of one oscillator. The full Hamiltonian then reads

$$H = \frac{p_1^2}{2\mu_r} + V_{\text{Mor}}(q_1) + \frac{p_2^2}{2\mu_r} + V_{\text{Mor}}(q_2) - \frac{1}{M}p_1p_2. \tag{B.35}$$

The corresponding Hamiltonian matrix elements can be analytically computed in the basis of the Morse oscillator eigenfunctions. We choose $\mu_r = \hbar = 1$, therefore the coupling parameter corresponds to the mass ratio $\lambda \equiv \frac{1}{M} = \frac{m}{m+M}$, which is in the range $0 \leq \lambda \leq 1$. [91] The coupling term, $H_{12} \sim p_1p_2$, can be expressed analytically in a basis of the eigenfunctions of the two one-dimensional Morse oscillators, $|n_1, n_2\rangle$, [90]

$$\langle m_1, m_2| \, p_1p_2 \, |n_1, n_2\rangle = \langle m_1| \, p_1 \, |n_1\rangle \, \langle m_2| \, p_2 \, |n_2\rangle \tag{B.36}$$
$$= -\int_{-\infty}^{\infty} \mathrm{d}x_1 \, \psi_{m_1}(x_1)\frac{\mathrm{d}}{\mathrm{d}x_1}\psi_{n_1}(x_1) \int_{-\infty}^{\infty} \mathrm{d}x_2 \, \psi_{m_2}(x_2)\frac{\mathrm{d}}{\mathrm{d}x_2}\psi_{n_2}(x_2). \tag{B.37}$$

Here we use the relationship

$$\frac{\mathrm{d}}{\mathrm{d}x}L_n^{(\alpha)}(x) = \begin{cases} -L_{n-1}^{(\alpha+1)}(x) & \text{if } n > 0, \\ 0 & \text{else.} \end{cases} \tag{B.38}$$

We insert

$$\frac{\mathrm{d}}{\mathrm{d}x}\psi_n(x) = -aN_n e^{-\frac{1}{2}z}\Big[\left(k - n - \tfrac{1}{2}\right) z^{k-n-\frac{1}{2}} L_n^{(2k-2n-1)}(z) - \tfrac{1}{2}z^{k-n+\frac{1}{2}} L_n^{(2k-2n-1)}(z) - z^{k-n+\frac{1}{2}} L_{n-1}^{(2k-2n)}(z)\Big] \tag{B.39}$$

to get

$$
\begin{aligned}
I_{m_1,n_1} &\equiv \int_{-\infty}^{\infty} \mathrm{d}x\, \psi_{m_1}(x) \frac{\mathrm{d}}{\mathrm{d}x} \psi_{n_1}(x) && \text{(B.40)} \\
&= -\, a N_{m_1} N_{n_1} \int \mathrm{d}x\, e^{-z} z^{k-m_1-\frac{1}{2}} L_{m_1}^{(2k-2m_1-1)}(z) \\
&\quad \times \Big[\left(k - n_1 - \tfrac{1}{2}\right) z^{k-n_1-\frac{1}{2}} L_{n_1}^{(2k-2n_1-1)}(z) - \tfrac{1}{2} z^{k-n_1+\frac{1}{2}} L_{n_1}^{(2k-2n_1-1)}(z) \\
&\qquad - z^{k-n_1+\frac{1}{2}} L_{n_1-1}^{(2k-2n_1)}(z) \Big] && \text{(B.41)} \\
&= -\, a N_{m_1} N_{n_1} \int \mathrm{d}x\, e^{-z} \Big[\left(k - n_1 - \tfrac{1}{2}\right) z^{2k-m_1-n_1-1} L_{m_1}^{(2k-2m_1-1)}(z) L_{n_1}^{(2k-2n_1-1)}(z) \\
&\qquad - \tfrac{1}{2} z^{2k-m_1-n_1} L_{m_1}^{(2k-2m_1-1)}(z) L_{n_1}^{(2k-2n_1-1)}(z) \\
&\qquad - z^{2k-m_1-n_1} L_{m_1}^{(2k-2m_1-1)}(z) L_{n_1-1}^{(2k-2n_1)}(z) \Big]. && \text{(B.42)}
\end{aligned}
$$

The first integral we have already computed when finding the normalization factor N_n,

$$
\begin{aligned}
I_1 &\equiv \int \mathrm{d}x\, e^{-z} z^{2k-m_1-n_1-1} L_{m_1}^{(2k-2m_1-1)}(z) L_{n_1}^{(2k-2n_1-1)}(z) && \text{(B.43)} \\
&= \frac{1}{a} \frac{\Gamma(2k-n_1)}{n_1!(2k-2n_1-1)} \delta_{m_1,n_1}. && \text{(B.44)}
\end{aligned}
$$

The second integral yields

$$I_2 \equiv \int \mathrm{d}x\, e^{-z} z^{2k-m_1-n_1} L_{m_1}^{(2k-2m_1-1)}(z) L_{n_1}^{(2k-2n_1-1)}(z) \tag{B.45}$$

$$= \frac{1}{a} \int_0^\infty \mathrm{d}z\, e^{-z} z^{2k-m_1-n_1-1} L_{m_1}^{(2k-2m_1-1)}(z) L_{n_1}^{(2k-2n_1-1)}(z) \tag{B.46}$$

$$= \frac{1}{a} \sum_{i=0}^{m_1} \binom{-m_1+n_1+i-1}{i} \sum_{j=0}^{n_1} \binom{m_1-n_1+j-1}{j} \times \int_0^\infty \mathrm{d}z\, e^{-z} z^{2k-m_1-n_1-1} L_{m_1-i}^{(2k-m_1-n_1-1)}(z) L_{n_1-j}^{(2k-m_1-n_1-1)}(z) \tag{B.47}$$

$$= \frac{1}{a} \sum_{i=0}^{m_1} \binom{-m_1+n_1+i-1}{i} \sum_{j=0}^{n_1} \binom{m_1-n_1+j-1}{j} \frac{\Gamma(2k-m_1-j)}{(n_1-j)!} \delta_{m_1-i,n_1-j} \tag{B.48}$$

$$= \frac{1}{a} \sum_{i=0}^{m_1} \binom{-m_1+n_1+i-1}{i} \sum_{\tilde{j}=0}^{n_1} \binom{m_1-\tilde{j}-1}{n_1-\tilde{j}} \frac{\Gamma(2k-m_1-n_1+\tilde{j})}{\tilde{j}!} \delta_{m_1-i,\tilde{j}} \tag{B.49}$$

$$= \frac{1}{a} \sum_{i=0}^{m_1} \binom{n_1-m_1+i-1}{i} \binom{i-1}{n_1-m_1+i} \frac{\Gamma(2k-n_1-i)}{(m_1-i)!} \tag{B.50}$$

$$= \frac{1}{a} \sum_{\tilde{i}=0}^{m_1} \binom{n_1-\tilde{i}-1}{m_1-\tilde{i}} \binom{m_1-\tilde{i}-1}{n_1-\tilde{i}} \frac{\Gamma(2k-m_1-n_1+\tilde{i})}{\tilde{i}!}, \tag{B.51}$$

where we used the property in Eq. (B.8). Now we use the identity

$$\binom{n-1}{k} = \frac{n-k}{k} \binom{n}{k} \tag{B.52}$$

to write

$$I_2 = \frac{1}{a} \sum_{i=0}^{m_1} \binom{n_1-i-1}{m_1-i} \binom{m_1-i-1}{n_1-i} \frac{\Gamma(2k-m_1-n_1+i)}{i!} \tag{B.53}$$

$$= \frac{1}{a} \sum_{i=0}^{m_1} \frac{n_1-m_1}{m_1-i} \frac{m_1-n_1}{n_1-i} \binom{n_1-i}{m_1-i} \binom{m_1-i}{n_1-i} \frac{\Gamma(2k-m_1-n_1+i)}{i!}. \tag{B.54}$$

For $m_1 \neq n_1$, all binomial coefficients vanish, except for the terms $i = m_1$ if $m_1 < n_1$, or $i = n_1$ if $m_1 > n_1$, giving

$$I_2 = \frac{1}{a} (-1)^{m_1-n_1} \frac{\Gamma(2k-m_1)}{n_1!} \quad \text{for } m_1 > n_1 \text{ and reversed otherwise.} \tag{B.55}$$

For $m_1 = n_1$, all terms but one vanish as well due to the fractions. However, in this case the initial integral simply satisfies the orthogonality relation and yields

$$I_2 = \frac{1}{a} \frac{\Gamma(2k-n_1)}{n_1!} \quad \text{for } m_1 = n_1. \tag{B.56}$$

The remaining third integral gives

$$
\begin{aligned}
I_3 \equiv & \int \mathrm{d}x\, e^{-z} z^{2k-m_1-n_1} L_{m_1}^{(2k-2m_1-1)}(z) L_{n_1-1}^{(2k-2n_1)}(z) && \text{(B.57)}\\
= & \frac{1}{a}\int_0^\infty \mathrm{d}z\, e^{-z} z^{2k-m_1-n_1-1} L_{m_1}^{(2k-2m_1-1)}(z) L_{n_1-1}^{(2k-2n_1)}(z) && \text{(B.58)}\\
= & \frac{1}{a}\sum_{i=0}^{m_1} \binom{-m_1+n_1+i-1}{i} \sum_{j=0}^{n_1-1} \binom{m_1-n_1+j}{j} \\
& \times \int_0^\infty \mathrm{d}z\, e^{-z} z^{2k-m_1-n_1-1} L_{m_1-i}^{(2k-m_1-n_1-1)}(z) L_{n_1-1-j}^{(2k-m_1-n_1-1)}(z) && \text{(B.59)}\\
= & \frac{1}{a}\sum_{i=0}^{m_1} \binom{-m_1+n_1+i-1}{i} \sum_{j=0}^{n_1-1} \binom{m_1-n_1+j}{j} \frac{\Gamma(2k-m_1-1-j)}{(n_1-1-j)!}\delta_{m_1-i,n_1-1-j} \\
& && \text{(B.60)}\\
= & \frac{1}{a}\sum_{i=0}^{m_1} \binom{-m_1+n_1+i-1}{i} \sum_{\tilde{j}=0}^{n_1-1} \binom{m_1-1-\tilde{j}}{n_1-1-\tilde{j}} \frac{\Gamma(2k-m_1-n_1+\tilde{j})}{\tilde{j}!}\delta_{m_1-i,\tilde{j}} \\
& && \text{(B.61)}\\
= & \frac{1}{a}\sum_{i=0}^{m_1} \binom{n_1-m_1+i-1}{i} \binom{i-1}{n_1-m_1+i-1} \frac{\Gamma(2k-n_1-i)}{(m_1-i)!} && \text{(B.62)}\\
= & \frac{1}{a}\sum_{i=0}^{m_1} \frac{m_1-n_1+1}{n_1-m_1+i-1} \binom{n_1-m_1+i-1}{i} \binom{i}{n_1-m_1+i-1} \frac{\Gamma(2k-n_1-i)}{(m_1-i)!}. \\
& && \text{(B.63)}
\end{aligned}
$$

The two last lines imply that not all binomial coefficients vanish. However, we could only consider the case of $m_1 \geq n_1$, for which all terms of this non-symmetric integral vanish. The two other integrals, I_1 and I_2, are invariant under exchange of $m_1 \leftrightarrow n_1$. For the opposite case of $m_1 < n_1$ we can simply use the property $\langle m| \frac{\mathrm{d}}{\mathrm{d}x} |n\rangle = -\langle n| \frac{\mathrm{d}}{\mathrm{d}x} |m\rangle$.

Then, we get for $m_1 \geq n_1$

$$
\begin{aligned}
I_{m_1,n_1} = & \int_{-\infty}^{\infty} \mathrm{d}x\, \psi_{m_1}(x) \frac{\mathrm{d}}{\mathrm{d}x} \psi_{n_1}(x) && \text{(B.64)}\\
= & -N_{m_1}N_{n_1}\left[\frac{1}{2}\frac{\Gamma(2k-n_1)}{n_1!}\delta_{m_1,n_1} - (-1)^{m_1-n_1}\frac{\Gamma(2k-m_1)}{2(n_1!)}\right], && \text{(B.65)}
\end{aligned}
$$

and finally

$$
I_{m_1,n_1} = \begin{cases} N_{m_1}N_{n_1}(-1)^{m_1-n_1}\frac{\Gamma(2k-m_1)}{2(n_1!)} & \text{if } m_1 > n_1, \\ -N_{m_1}N_{n_1}(-1)^{n_1-m_1}\frac{\Gamma(2k-n_1)}{2(m_1!)} & \text{if } m_1 < n_1, \\ 0 & \text{if } m_1 = n_1, \end{cases} \qquad \text{(B.66)}
$$

and similar for $\int \mathrm{d}x_2\, \psi_{m_2} \frac{\mathrm{d}}{\mathrm{d}x_2} \psi_{n_2}$.

B.1 Kinetic decoupling

It is possible to represent the problem of two kinetically coupled Morse oscillators, Eqs. (B.35) or (5.2), as a two-dimensional oscillator with potential coupling only. This can be done by a proper rotation of coordinates, such that the term $\sim p_1 p_2$ vanishes. To achieve this, we rotate our coordinates by an angle θ,

$$\begin{pmatrix} q_1' \\ q_2' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} \tag{B.67}$$

or

$$\begin{pmatrix} q_1 \\ q_2 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} q_1' \\ q_2' \end{pmatrix}, \tag{B.68}$$

and the momenta, p_1 and p_2, accordingly. This change of variables is a valid symplectic transformation and therefore leaves Hamilton's equations of motion invariant. [128] Then, we can rewrite the terms involving momenta,

$$\begin{aligned} \frac{p_1^2}{2\mu_r} + \frac{p_2^2}{2\mu_r} - \lambda p_1 p_2 = \frac{{p_1'}^2}{2\mu_r}\left(1 - 2\lambda\mu_r \cos\theta\sin\theta\right) + \frac{{p_2'}^2}{2\mu_r}\left(1 + 2\lambda\mu_r \cos\theta\sin\theta\right) \\ - \lambda {p_1'}^2 {p_2'}^2 \left(\cos^2\theta - \sin^2\theta\right). \end{aligned} \tag{B.69}$$

Now we choose $\theta = \frac{\pi}{4}$, which gives $\cos\theta = \sin\theta = \frac{1}{\sqrt{2}}$, for the coupling term to vanish,

$$\frac{p_1^2}{2\mu_r} + \frac{p_2^2}{2\mu_r} - \lambda p_1 p_2 = \frac{{p_1'}^2}{2\mu_r}\left(1 - \lambda\mu_r\right) + \frac{{p_2'}^2}{2\mu_r}\left(1 + \lambda\mu_r\right). \tag{B.70}$$

The potential part of Hamiltonian (B.35)/(5.2) then gives

$$V_{\mathrm{Mor}}(q_1) + V_{\mathrm{Mor}}(q_2) = D\left(1 - e^{-aq_1}\right)^2 + D\left(1 - e^{-aq_2}\right)^2 \tag{B.71}$$

$$= D\left(1 - e^{-\frac{a}{\sqrt{2}}(q_1' - q_2')}\right)^2 + D\left(1 - e^{-\frac{a}{\sqrt{2}}(q_1' + q_2')}\right)^2. \tag{B.72}$$

In this coordinate system the unperturbed Hamiltonian is no longer separate in its coordinates.

For the case of two coupled harmonic oscillators, Eq. (5.24), the motion can be completely decoupled, [104]

$$V_{\mathrm{HO}}(q_1) + V_{\mathrm{HO}}(q_2) = \tfrac{1}{2}\mu_r\omega^2 q_1^2 + \tfrac{1}{2}\mu_r\omega^2 q_2^2 \tag{B.73}$$

$$= \tfrac{1}{2}\mu_r\omega^2 {q_1'}^2 + \tfrac{1}{2}\mu_r\omega^2 {q_1'}^2. \tag{B.74}$$

C Hydrogen in a homogeneous magnetic field

C.1 Coordinate basis

The energy spectrum of a single hydrogen atom in an external homogeneous magnetic field, aligned along the z-axis, is given as the solutions of the TISE. Coupling the hydrogen atom to the external magnetic field, the system Hamiltonian in Cartesian coordinates writes

$$H = \frac{1}{2}\left(\mathbf{p} + \mathbf{A}(\mathbf{x}, t)\right)^2 - \frac{1}{r} \tag{C.1}$$

$$= \frac{\mathbf{p}^2}{2} + \frac{B}{2}\underbrace{(xp_y - yp_x)}_{L_z} + \frac{B^2}{8}\left(x^2 + y^2\right) - \frac{1}{r}. \tag{C.2}$$

Due to the cylindrical symmetry of the system, one can perform a coordinate transformation to cylindrical coordinates $(x, y, z) \to (\rho, \varphi, z)$, where $\rho = \sqrt{x^2 + y^2}$ and $\varphi = \arctan\left(\frac{y}{x}\right)$, yielding for the Laplace operator

$$\mathbf{\Delta} = \frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho \frac{\partial}{\partial \rho}\right) + \frac{1}{\rho^2}\frac{\partial^2}{\partial \varphi^2} + \frac{\partial^2}{\partial z^2} = \frac{1}{\rho}\frac{\partial}{\partial \rho} + \frac{\partial^2}{\partial \rho^2} + \frac{1}{\rho^2}\frac{\partial^2}{\partial \varphi^2} + \frac{\partial^2}{\partial z^2}. \tag{C.3}$$

The magnetic quantum number, m, as well as the z-parity, π_z, with respect to $z = 0$ are good quantum numbers of the system. The full single-electron wave function is separable in the azimuth angle, φ, [109]

$$\psi(\rho, \varphi, z) = \psi(\rho, z)\frac{e^{im\varphi}}{\sqrt{2\pi}}, \tag{C.4}$$

where the φ-dependence follows from the fact, that the eigenfunction of the z-component of the orbital angular momentum operator, L_z, are the spherical harmonics, $Y_{lm}(\vartheta, \varphi)$. What is left is the non-separable part of the TISE for the coordinates (ρ, z),

$$\left[\frac{1}{2}\left(-\frac{1}{\rho}\frac{\partial}{\partial \rho} - \frac{\partial^2}{\partial \rho^2} - \frac{\partial^2}{\partial z^2} + \frac{m^2}{\rho^2}\right) + \frac{B^2}{8}\rho^2 - \frac{1}{\sqrt{\rho^2 + z^2}}\right]\psi(\rho, z) = E\psi(\rho, z), \tag{C.5}$$

where we neglected the constant contribution from the linear Zeeman term, $\frac{1}{2}BL_z$. This system corresponds to that of a particle moving in an effective two-dimensional potential,

$$V_{\text{eff}}(\rho, z) = \frac{m^2}{2\rho^2} + \frac{B^2\rho^2}{8} - \frac{1}{\sqrt{\rho^2 + z^2}}. \tag{C.6}$$

We could numerically solve above eigenvalue problem by discretizing the wave function on a two-dimensional grid in the coordinates (ρ, z) and use the finite differences method for representing the derivatives. For the boundary at $\rho = 0$ we use the symmetric boundary condition, $\frac{\partial\psi}{\partial\rho}\big|_{\rho=0} = 0$. The singularity in the first term can then be resolved by using L'Hôpital's rule,

$$\left.\frac{\frac{\partial\psi}{\partial\rho}}{\rho}\right|_{\rho=0} = \left.\frac{\frac{\partial^2\psi}{\partial\rho^2}}{\frac{\partial\rho}{\partial\rho}}\right|_{\rho=0} = \left.\frac{\partial^2\psi}{\partial\rho^2}\right|_{\rho=0}, \tag{C.7}$$

which yields the governing equation for the Laplace operator at the boundary,

$$\frac{1}{\rho}\frac{\partial}{\partial\rho} + \frac{\partial^2}{\partial\rho^2} + \frac{\partial^2}{\partial z^2} \overset{\rho=0}{=} 2\frac{\partial^2}{\partial\rho^2} + \frac{\partial^2}{\partial z^2}. \tag{C.8}$$

The finite differences equation for the derivative $\frac{\partial^2\psi}{\partial\rho^2}$ with the boundary condition $\frac{\partial\psi}{\partial\rho} = 0$ at $\rho = 0$ are derived by Taylor expanding $\psi(\rho_2)$ and $\psi(\rho_3)$ around the boarder points $\rho_1 \equiv 0$, where ρ_1, ρ_2, ρ_3 are the first three grid sites $\rho_j \geq 0$,

$$\psi(\rho_2) = \psi(\rho_1) + \Delta\rho \underbrace{\psi'(\rho_1)}_{=0} + \frac{\Delta\rho^2}{2}\psi''(\rho_1) + \frac{\Delta\rho^3}{6}\psi'''(\rho_1) + \mathcal{O}(\Delta\rho^4), \tag{C.9}$$

$$\psi(\rho_3) = \psi(\rho_1) + 2\Delta\rho \underbrace{\psi'(\rho_1)}_{=0} + \frac{4\Delta\rho^2}{2}\psi''(\rho_1) + \frac{8\Delta\rho^3}{6}\psi'''(\rho_1) + \mathcal{O}(\Delta\rho^4), \tag{C.10}$$

and subtracting $\psi(\rho_3)$ by $\psi(\rho_2)$, such that the terms $\sim \mathcal{O}(\Delta\rho^3)$ vanish,

$$\psi(\rho_3) - 8\psi(\rho_2) = -7\psi(\rho_1) - 2\Delta\rho^2\psi''(\rho_1) + \mathcal{O}(\Delta\rho^4) \tag{C.11}$$

$$\longrightarrow \psi''(\rho_1) = \frac{1}{\Delta\rho^2}\left(-\frac{7}{2}\psi(\rho_1) + 4\psi(\rho_2) - \frac{1}{2}\psi(\rho_3)\right) + \mathcal{O}(\Delta\rho^2). \tag{C.12}$$

This expression is used in the governing equation for the Laplace operator above for all boundary points at $\rho = 0$.

Note that the corresponding Hamilton matrix is not symmetric. However, the Hamilton operator is still Hermitian with respect to the inner product in cylindrical coordinates,

$$\sum_{\rho_i} \Delta\rho\rho_i \sum_{z_j} \Delta z \left\langle \rho_i', z_j' \right| \hat{H} \left|\rho_i, z_j\right\rangle \psi(\rho_i, z_j) = E\psi(\rho_i, z_j). \tag{C.13}$$

To further simplify above Hamiltonian, we use the fact that the parity operator commutes with the Hamiltonian of our system, $\left[H, \hat{\pi}_z\right]$. Therefore we use an eigenbasis of $\hat{\pi}_z$ in our finite differences method, namely

$$\left\langle z \middle| b_i^{\pm} \right\rangle = b_i^{\pm}(z) = \frac{1}{\sqrt{2}}\left[\delta(z - z_i) \pm \delta(z + z_i)\right], \quad z_i > 0. \tag{C.14}$$

This then leads to the Hamiltonian matrix elements

$$\langle b_i|\, \hat{H}\, |b_j\rangle = \int \mathrm{d}x \int \mathrm{d}x' \, \langle b_i|x\rangle \, \langle x|\, \hat{H}\, |x'\rangle \, \langle x'|b_j\rangle \tag{C.15}$$

$$= \frac{1}{2} \int \mathrm{d}x \int \mathrm{d}x' \, [\delta(x - x_i) \pm \delta(x + x_i)] \, \langle x|\, \hat{H}\, |x'\rangle \, \left[\delta(x' - x_j) \pm \delta(x' + x_j)\right] \tag{C.16}$$

$$= \frac{1}{2} \left[\langle x_i|\, \hat{H}\, |x_j\rangle \pm \langle -x_i|\, \hat{H}\, |x_j\rangle \pm \langle x_i|\, \hat{H}\, |-x_j\rangle + \langle -x_i|\, \hat{H}\, |-x_j\rangle\right]. \tag{C.17}$$

In order to circumvent the singularity at the origin, $\mathbf{r} = 0$, we chose our z-grid such that points start at $\pm z_1 = \pm\frac{\Delta z}{2}$. Otherwise we ought to use $\langle z|b_0\rangle = \delta(z)$ with symmetric/anti-symmetric boundary conditions at $z = 0$ for even/odd parity.

C.2 Orthogonal Sturmian basis

A suitable orthogonal basis for the Hydrogen atom in a magnetic field involves orthogonalized Sturmian functions, [112–114]

$$S_{nl}^{(\zeta)}(r) = \zeta^{\frac{3}{2}} \sqrt{\frac{(n-l-1)!}{(n+l+1)!}} e^{-\frac{\zeta r}{2}} (\zeta r)^l L_{n-l-1}^{(2l+2)}(\zeta r). \tag{C.18}$$

The associated (or generalized) Laguerre polynomials, $L_n^{(\alpha)}(x)$, satisfy the relation

$$\int_0^\infty \mathrm{d}x \; e^{-x} x^\alpha L_n^{(\alpha)}(x) L_{n'}^{(\alpha)}(x) = \frac{(\alpha+n)!}{n!} \delta_{n'n}, \tag{C.19}$$

from which the orthonormality of the Sturmian functions follows. Using the basis $|n, l, m\rangle = |n, l\rangle \, |l, m\rangle \equiv S_{nl}(r)\, Y_{lm}(\theta, \varphi)$, where Y_{lm} are the spherical harmonics, the matrix elements of the Hamiltonian, given in spherical coordinates as

$$H(r, \theta, \varphi) = \underbrace{\frac{p^2}{2} - \frac{1}{r}}_{H_0} + \lambda \underbrace{\frac{1}{8} r^2 \sin^2\theta}_{H_1} \tag{C.20}$$

are

$$\langle n', l', m'|\, H\, |n, l, m\rangle = \langle n', l', m'|\, H_0\, |n, l, m\rangle + \lambda \, \langle n', l'|\, r^2\, |n, l\rangle \, \langle l', m'|\, \frac{\sin^2\theta}{8}\, |l, m\rangle \, . \tag{C.21}$$

For nonzero but equal azimuthal quantum numbers, $m' = m \geq 0$, the matrix elements in the basis of the spherical harmonics, $Y_{lm}(\theta,\varphi)$, yield

$$\begin{aligned}\langle l',m|\,\frac{\sin^2\theta}{8}\,|l,m\rangle &= \frac{1}{8}\int_{-\pi}^{\pi}\mathrm{d}\varphi\int_0^{\pi}\mathrm{d}\theta\sin\theta\,Y_{l'm}\sin^2\theta\,Y_{lm} \qquad \text{(C.22)}\\ &= \frac{2\pi}{8}\sqrt{\frac{2l'+1}{4\pi}\frac{(l'-m)!}{(l'+m)!}}\sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}\\ &\quad\times\int_0^{\pi}\mathrm{d}\theta\sin\theta\,P_{l'}^m(\cos\theta)\sin^2\theta\,P_l^m(\cos\theta) \qquad \text{(C.23)}\\ &= \frac{1}{16}\sqrt{(2l'+1)(2l+1)\frac{(l'-m)!}{(l'+m)!}\frac{(l-m)!}{(l+m)!}}\\ &\quad\times\int_{-1}^{1}\mathrm{d}u\,P_{l'}^m(u)\left(1-u^2\right)P_l^m(u). \qquad \text{(C.24)}\end{aligned}$$

The integral can be evaluated using the properties of the associated Legendre polynomials, P_l^m,

$$\int_{-1}^{1}\mathrm{d}u\,P_{l'}^m(u)\,P_l^m(u) = \frac{2\,(l+m)!}{(2l+1)\,(l-m)!}\delta_{l'l}\,, \qquad \text{(C.25)}$$

$$(l-m+1)P_{l+1}^m(u)+(l+m)P_{l-1}^m(u) = (2l+1)uP_l^m(u). \qquad \text{(C.26)}$$

Then, we arrive at the expression

$$\begin{aligned}\langle l',m|\,\frac{\sin^2\theta}{8}\,|l,m\rangle = \frac{1}{8}\Bigg[&\frac{2\left(l^2+l+m^2-1\right)}{(2l+3)(2l-1)}\delta_{l'l}\\ &-\frac{\sqrt{(l+m)(l+m-1)(l-m)(l-m-1)}}{\sqrt{(2l+1)(2l-3)(2l-1)}}\delta_{l',l-2}\\ &-\frac{\sqrt{(l-m+1)(l-m+2)(l+m+1)(l+m+2)}}{\sqrt{(2l+1)(2l+5)(2l+3)}}\delta_{l',l+2}\Bigg].\end{aligned} \qquad \text{(C.27)}$$

The matrix elements $\langle n',l'|\,r^2\,|n,l\rangle$ in the basis of the Sturmians can be calculated in closed form using the property from Refs. [113, 129],

$$\int_0^{\infty}\mathrm{d}x\,x^p e^{-x}L_m^{(\alpha)}(x)L_n^{(\beta)}(x) = (-1)^{m+n}p!\sum_{\tau=0}^{\min(m,n)}\binom{p-\alpha}{m-\tau}\binom{p-\beta}{n-\tau}\binom{p+\tau}{\tau}, \qquad \text{(C.28)}$$

yielding

$$I_{nl}^{n'l'} \equiv \left\langle n', l' \right| r^2 \left| n, l \right\rangle \tag{C.29}$$

$$= \zeta^{3+l'+l} \sqrt{\frac{(n'-l'-1)!}{(n'+l'+1)!} \frac{(n-l-1)!}{(n+l+1)!}} \int_0^\infty \mathrm{d}r \, r^{4+l'+l} e^{-\zeta r} L_{n'-l'-1}^{(2l'+2)}(\zeta r) L_{n-l-1}^{(2l+2)}(\zeta r) \tag{C.30}$$

$$= \zeta^{-2} \sqrt{\frac{(n'-l'-1)!}{(n'+l'+1)!} \frac{(n-l-1)!}{(n+l+1)!}} (4+l+l')! \times \sum_{\tau=0}^{\le \substack{n'-l'-1 \\ n-l-1}} (-1)^{n'+l'+n+l+\tau} \binom{2-l'+l}{n'-l'-1-\tau} \binom{2+l'-l}{n-l-1-\tau} \binom{-5-l'-l}{\tau}. \tag{C.31}$$

All non-vanishing terms for $l' = l$, $l' = l+2$ and $l' = l-2$ are listed explicitly in Refs. [113, 114].

The matrix elements for the unperturbed hydrogen atom,

$$\left\langle n', l', m' \right| H_0 \left| n, l, m \right\rangle = \left\langle n', l' \right| H_0 \left| n, l \right\rangle \delta_{l'l} \delta_{m'm}, \tag{C.32}$$

where

$$\left\langle n', l' \right| H_0 \left| n, l \right\rangle = \int_0^\infty \mathrm{d}r \, r^2 S_{n'l'}(r) H_0 S_{nl}(r) \tag{C.33}$$

and

$$H_0(r, \theta, \varphi) = \frac{p^2}{2} - \frac{1}{r} = -\frac{1}{2} \left(\frac{\mathrm{d}^2}{\mathrm{d}r^2} + \frac{2}{r} \frac{\mathrm{d}}{\mathrm{d}r} - \frac{l(l+1)}{r^2} \right) - \frac{1}{r}, \tag{C.34}$$

are found using the properties of the associated Laguerre polynomials,

$$\left(r \frac{\mathrm{d}^2}{\mathrm{d}r^2} + (k+1-r) \frac{\mathrm{d}}{\mathrm{d}r} + n \right) L_n^{(k)}(r) = 0, \tag{C.35}$$

$$r \frac{\mathrm{d}}{\mathrm{d}r} L_n^{(k)}(r) = n L_n^{(k)}(r) - (n+k) L_{n-1}^{(k)}(r), \tag{C.36}$$

and the chain-rule of differentiation giving

$$\left(r \frac{\mathrm{d}^2}{\mathrm{d}r^2} + (2l+3-\zeta r) \frac{\mathrm{d}}{\mathrm{d}r} + \zeta \left(n-l-1 \right) \right) L_{n-l-1}^{(2l+2)}(\zeta r) = 0, \tag{C.37}$$

$$r \frac{\mathrm{d}}{\mathrm{d}r} L_\beta^{(\alpha)}(\zeta r) = \beta L_\beta^{(\alpha)}(\zeta r) - (\beta+\alpha) L_{\beta-1}^{(\alpha)}(\zeta r). \tag{C.38}$$

We get, using above relations

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}r^2}+\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\right)S_{nl}(r)=\zeta^{\frac{3}{2}+l}\sqrt{\frac{(n-l-1)!}{(n+l+1)!}}e^{-\frac{\zeta r}{2}}r^l \times\left[\frac{l^2}{r^2}+\frac{\zeta^2}{4}-\frac{\zeta}{r}(n-l-1)-\frac{\zeta}{r}\left(l+\frac{1}{2}\right)-\frac{2}{r}\frac{\mathrm{d}}{\mathrm{d}r}\right]L_{n-l-1}^{(2l+2)}(\zeta r) \tag{C.39}$$

or

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}r^2}+\frac{2}{r}\frac{\mathrm{d}}{\mathrm{d}r}\right)S_{nl}(r)=\zeta^{\frac{3}{2}+l}\sqrt{\frac{(n-l-1)!}{(n+l+1)!}}e^{-\frac{\zeta r}{2}}r^l \times\left[\frac{l(l+1)}{r^2}+\frac{\zeta^2}{4}-\frac{\zeta n}{r}-\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\right]L_{n-l-1}^{(2l+2)}(\zeta r) \tag{C.40}$$

$$=\zeta^{\frac{3}{2}+l}\sqrt{\frac{(n-l-1)!}{(n+l+1)!}}e^{-\frac{\zeta r}{2}}r^l \tag{C.41}$$

$$\times\left\{\left[\frac{l(l+1)}{r^2}+\frac{\zeta^2}{4}-\frac{\zeta n}{r}-\frac{n-l-1}{r^2}\right]L_{n-l-1}^{(2l+2)}(\zeta r) +\frac{n+l+1}{r^2}L_{n-l-2}^{(2l+2)}\right\} \tag{C.42}$$

$$=\left[\frac{l(l+1)}{r^2}+\frac{\zeta^2}{4}-\frac{\zeta n}{r}-\frac{n-l-1}{r^2}\right]S_{nl}(r) +\frac{\sqrt{(n-l-1)(n+l+1)}}{r^2}S_{n-1,l}(r). \tag{C.43}$$

Then,

$$H_0S_{nl}(r)=\left[-\frac{1}{2}\left(\frac{\mathrm{d}^2}{\mathrm{d}r^2}+\frac{2}{r}\frac{\mathrm{d}}{\mathrm{d}r}-\frac{l(l+1)}{r^2}\right)-\frac{1}{r}\right]S_{nl}(r) \tag{C.44}$$

$$=\left[\frac{n-l-1}{2r^2}+\frac{\zeta n}{2r}-\frac{\zeta^2}{8}-\frac{1}{r}\right]S_{nl}(r)-\frac{\sqrt{(n-l-1)(n+l+1)}}{2r^2}S_{n-1,l}(r). \tag{C.45}$$

The matrix elements then become

$$\begin{aligned}
\langle n', l', m' | \, H_0 \, | n, l, m \rangle &= \delta_{l'l}\delta_{m'm} \int_0^\infty \mathrm{d}r \, r^2 S_{n'l'}(r) H_0 S_{nl}(r) \qquad \text{(C.46)} \\
&= \frac{1}{2}\delta_{l'l}\delta_{m'm} \Bigg\{ (n-l-1) \int_0^\infty \mathrm{d}r \, S_{n'l'}(r) S_{nl}(r) \\
&\qquad + (\zeta n - 2) \int_0^\infty \mathrm{d}r \, r S_{n'l'}(r) S_{nl}(r) \\
&\qquad - \frac{\zeta^2}{4} \underbrace{\int_0^\infty dr \, r^2 S_{n'l'}(r) S_{nl}(r)}_{=\delta_{n'n}} \\
&\quad - \sqrt{(n-l-1)(n+l+1)} \int_0^\infty \mathrm{d}r \, S_{n'l'}(r) S_{n-1,l}(r) \Bigg\}. \qquad \text{(C.47)}
\end{aligned}$$

Using above integral relation we find

$$\begin{aligned}
\int_0^\infty \mathrm{d}r \, S_{n'l'}(r) S_{nl}(r) &= \zeta^2 \sqrt{\frac{(n'-l'-1)!(n-l-1)!}{(n'+l'+1)!(n+l+1)!}} (-1)^{n'+n-l'-l} (l'+l)! \\
&\quad \times \sum_{\tau=0}^{\leq \substack{n'-l'-1 \\ n-l-1}} \binom{l-l'-2}{n'-l'-1-\tau} \binom{l'-l-2}{n-l-1-\tau} \binom{l'+l+\tau}{\tau},
\end{aligned} \qquad \text{(C.48)}$$

$$\begin{aligned}
\int_0^\infty \mathrm{d}r \, r \, S_{n'l'}(r) S_{nl}(r) &= \zeta \sqrt{\frac{(n'-l'-1)!(n-l-1)!}{(n'+l'+1)!(n+l+1)!}} (-1)^{n'+n-l'-l} (l'+l+1)! \\
&\quad \times \sum_{\tau=0}^{\leq \substack{n'-l'-1 \\ n-l-1}} \binom{l-l'-1}{n'-l'-1-\tau} \binom{l'-l-1}{n-l-1-\tau} \binom{l'+l+1+\tau}{\tau}.
\end{aligned} \qquad \text{(C.49)}$$

Substituting the integrals and simplifying the sums, we finally arrive at the expressions for the matrix elements for the unperturbed hydrogen,

$$\begin{aligned}
\langle n', l', m' | \, H_0 \, | n, l, m \rangle = & -\frac{\zeta^2}{8}\delta_{n'n}\delta_{l'l}\delta_{m'm} + \frac{\zeta^2}{2} \sqrt{\frac{(n'-l'-1)!}{(n'+l'+1)!}} \sqrt{\frac{(n-l-1)!}{(n+l+1)!}} \\
& \times \sum_{\tau=0}^{\leq \substack{n'-l'-1 \\ n-l-1}} \frac{(2l+\tau)!}{\tau!} \Big[(n-l-1)(n-l-\tau)(n'-l-\tau) \\
& - (n+l+1)(n-l-1-\tau)(n'-l-\tau) \\
& + \left(n - \frac{2}{\zeta}\right)(2l+\tau+1) \Big] \delta_{l'l}\delta_{m'm}.
\end{aligned} \qquad \text{(C.50)}$$

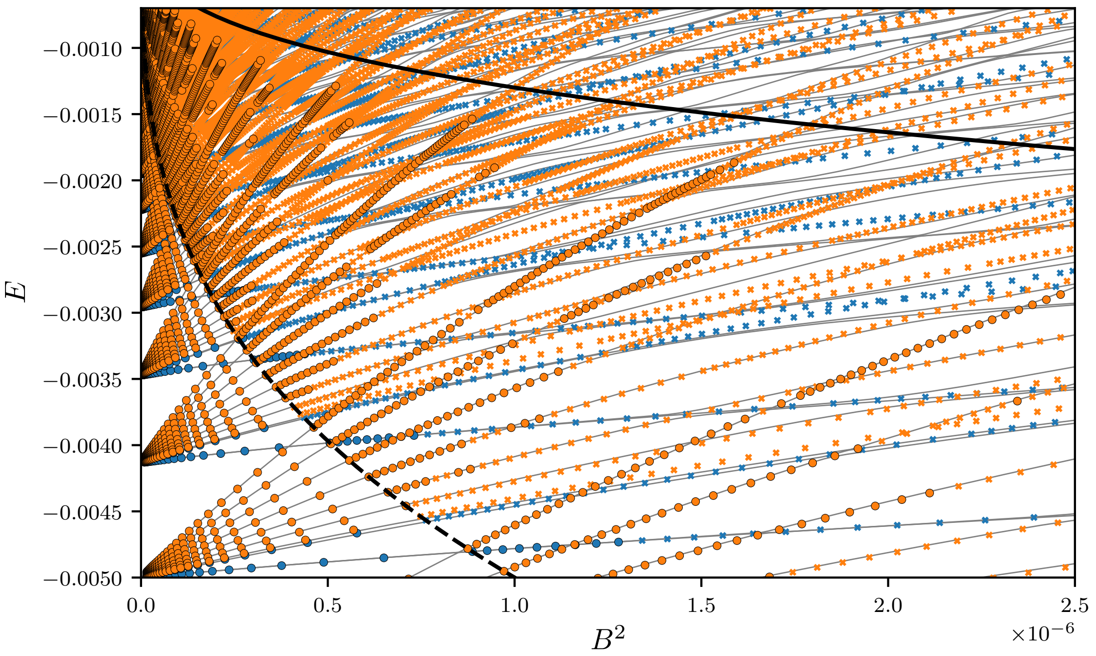

Figure C.1: Corresponding to Fig. 6.4, the energy spectra of hydrogen in a homogeneous magnetic field are shown. The exact quantum spectra are depicted as gray lines. The semiclassical EBK spectra are illustrated as blue markers for librational and orange for rotational trajectories. Circles depict energies which are interpolated between computed trajectories within regular phase space, while crosses depict extrapolated energy values. The black dashed and solid lines show scaled energy values of $\epsilon = -0.5$ and -0.13, respectively, marking the region of mixed phase space. The errorbars of the GP uncertainty are omitted for the sake of visibility.

Abbreviations

a.u.	atomic units 12, 31
AL	active learning 28–32, 75, 77
ANN	artificial neural network 2–6, 9, 10, 13, 14, 18, 26, 32
BNN	Bayesian neural network 5, 6, 14, 18
BO	Bayesian optimization 31, 32
CG	conjugate gradient 22, 24–27, 77, 79
DKL	deep kernel learning 10, 26
EBK	Einstein-Brillouin-Keller 2, 42, 45–48, 53, 55, 59–63, 68, 70, 71, 73, 75–81, 102
GP	Gaussian process 3, 5–22, 26, 28, 29, 31, 32, 36, 46, 60–62, 75–77, 79, 81, 102
GPR	Gaussian process regression 2, 3, 15, 18–21, 25, 26, 28, 35–38, 41, 46, 47, 49, 59–61, 63, 65, 75, 76, 79–81
LCAO	linear combination of atomic orbitals 30, 31
ML	machine learning 1–8, 13–16, 18, 21, 36, 79
MLE	maximum likelihood estimation 19, 20, 27, 36, 37, 59–61
MLP	multilayer perceptron 4–6
PDF	probability density function 7, 19, 83
RBF	radial basis function 9–11, 16, 18–20, 23, 26, 27, 32, 36, 75, 77
RQ	rational quadratic 10
SKI	structured kernel interpolation 25, 26
SM	spectral mixture 10, 23, 36, 77
SOS	surfaces of section 44, 45, 53, 54, 56, 59, 60, 71, 73, 74, 80, 81, 86
TISE	time-independent Schrödinger equation 33, 41, 42, 46, 48, 50, 65, 66, 95

List of symbols

D	input dimension, $\mathbf{x}, \mathbf{q} \in \mathbb{R}^D$. 4, 7, 9, 11, 13, 21–26, 41–46, 83, 84
N	number of training data. 7, 8, 13, 15, 16, 20–26, 29, 32, 83
$\mathbf{X}$	design matrix of training data, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N) \in \mathbb{R}^{D \times N}$. 13, 15–17, 20, 21, 25
$\mathbf{x}$	input variable for ML model. 4–17, 19, 22–24, 28, 29, 31, 32, 75
$\mathbf{y}$	output values of training data, $\mathbf{y} = (y_1, \ldots, y_N)^\top$. 13–17, 19–27, 29

Σ	covariance of a multivariate probability distribution. 5, 7–12, 14–17, 20–26
$\mathcal{N}$	Gaussian probability distribution. 5–8, 14, 15
$\mathcal{GP}$	Gaussian process. 7, 8, 11, 12, 14, 15
μ	mean (or expected value) of a probability distribution. 5–8, 11, 12, 14–16, 31, 36, 83
$\mathcal{P}$	probability density function. 5–8, 14, 15, 19, 20
σ	standard deviation of a probability distribution. 5–7, 9, 12, 16, 17, 28, 31, 36, 77, 78, 83

$\boldsymbol{\Sigma}_n$	covariance matrix of training data with added diagonal noise, $\boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}(\mathbf{X}, \mathbf{X}) + {\sigma_n}^2\mathbf{I}$. 20–22
$\boldsymbol{\theta}$	vector of all hyperparameters. 18–21
l	length-scale. 9, 10, 12, 16, 18–20, 26, 27, 29, 36, 37, 59–62, 75, 76
$\mathcal{L}$	logarithm of the likelihood function. 19–21, 83
$\mathbf{D}_n$	diagonal matrix with variable noise, $D_{n,ij} = {\sigma_n}_{,j}^2\delta_{ij}$. 24, 25
σ_n	standard deviation of Gaussian noise on training data used for spherical noise. 14, 15, 17–26, 37, 59–61, 76
σ_f	output-scale. 9, 10, 12, 16–20, 26, 37, 76

D_e	dissoziation energy of Morse potential. 35, 49–57, 59, 61, 87
V_{HO}	harmonic potential. 55
V_{Mor}	Morse potential. 49, 87, 90
a	stiffness parameter of Morse potential. 49–52, 54–56, 87–93
S	action variable. 43, 59, 60, 62, 68–70, 73, 75–77

ϵ	energy scaled by the magnetic field. 70–78, 102
μ_r	reduced mass. 49–52, 55, 87, 90, 94
ν	Maslov index. 42, 45, 68–70, 73

Bibliography

[1] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, “Machine learning and the physical sciences”, Rev. Mod. Phys. **91**, 045002 (2019).

[2] R. V. Krems, “Bayesian machine learning for quantum molecular dynamics”, Phys. Chem. Chem. Phys. **21**, 13392 (2019).

[3] T. B. Blank, S. D. Brown, A. W. Calhoun, and D. J. Doren, “Neural network models of potential energy surfaces”, J. Chem. Phys. **103**, 4129 (1995).

[4] J. Behler, “Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations”, Phys. Chem. Chem. Phys. **13**, 17930 (2011).

[5] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, “Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons”, Phys. Rev. Lett. **104**, 136403 (2010).

[6] A. P. Bartók and G. Csányi, “Gaussian approximation potentials: A brief tutorial introduction”, Int. J. Quantum Chem. **115**, 1051 (2015).

[7] A. Kamath, R. A. Vargas-Hernández, R. V. Krems, T. Carrington, and S. Manzhos, “Neural networks vs Gaussian process regression for representing potential energy surfaces: A comparative study of fit quality and vibrational spectrum accuracy”, J. Chem. Phys. **148**, 241702 (2018).

[8] A. Wilson and R. Adams, “Gaussian process kernels for pattern discovery and extrapolation”, International Conference on Machine Learning, 1067, PMLR: vol. 28, Atlanta, Georgia, USA (2013).

[9] A. G. Wilson, *Covariance kernels for fast automatic pattern discovery and extrapolation with Gaussian processes*, Ph.D. thesis, University of Cambridge, UK (2014).

[10] A. Björck, *Numerical Methods for Least Squares Problems*, Society for Industrial and Applied Mathematics (1996).

[11] F. Rosenblatt, “The perceptron: A probabilistic model for information storage and organization in the brain.”, Psychol. Rev. **65**, 386 (1958).

[12] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press (2016).

[13] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press (2005).

[14] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors”, Nature **323**, 533 (1986).

[15] R. M. Neal, *Bayesian Learning for Neural Networks*, Springer New York (1996).

[16] K.-I. Funahashi, “On the approximate realization of continuous mappings by neural networks”, Neural Netw. **2**, 183 (1989).

[17] K. Hornik, M. Stinchcombe, and H. White, “Multilayer feedforward networks are universal approximators”, Neural Netw. **2**, 359 (1989).

[18] G. Cybenko, “Approximation by superpositions of a sigmoidal function”, Math. Control Signals Syst. **2**, 303 (1989).

[19] E. V. Bonilla, K. Chai, and C. Williams, “Multi-task gaussian process prediction”, Advances in Neural Information Processing Systems, vol. 20, Curran Associates, Inc., Vancouver, Canada (2007).

[20] A. G. Wilson, Z. Hu, R. Salakhutdinov, and E. P. Xing, “Deep kernel learning”, International Conference on Artificial Intelligence and Statistics, 370, PMLR: vol. 51, Cadiz, Spain (2016).

[21] M. Lange-Hegermann, “Linearly constrained Gaussian processes with boundary conditions”, International Conference on Artificial Intelligence and Statistics, 1090, PMLR: vol. 130, virtual (2021).

[22] A. Solin and M. Kok, “Know your boundaries: constraining Gaussian processes by variational harmonic features”, International Conference on Artificial Intelligence and Statistics, 2193, PMLR: vol. 89, Naha, Okinawa, Japan (2019).

[23] C. Cortes and V. Vapnik, “Support-vector networks”, Mach. Learn. **20**, 273 (1995).

[24] J. P. Cunningham, K. V. Shenoy, and M. Sahani, “Fast Gaussian process methods for point process intensity estimation”, International Conference on Machine Learning, 192, ACM Press, Helsinki, Finland (2008).

[25] G. H. Golub and C. F. V. Loan, *Matrix computations*, JHU press (2013).

[26] Y. Saatçi, *Scalable inference for structured Gaussian process models*, Ph.D. thesis, University of Cambridge, UK (2011).

[27] E. Gilboa, Y. Saatci, and J. P. Cunningham, “Scaling multidimensional inference for structured Gaussian processes”, IEEE Trans. Pattern Anal. Mach. Intell. **37**, 424 (2015).

[28] A. G. Wilson, E. Gilboa, A. Nehorai, and J. P. Cunningham, “Fast kernel learning for multidimensional pattern extrapolation”, Advances in Neural Information Processing Systems, vol. 27, Curran Associates, Inc. (2014).

[29] J. Quiñonero-Candela and C. E. Rasmussen, “A unifying view of sparse approximate Gaussian process regression”, J. Mach. Learn. Res. **6**, 1939 (2005).

[30] T. D. Bui, J. Yan, and R. E. Turner, “A unifying framework for Gaussian process pseudo-point approximations using power expectation propagation”, J. Mach. Learn. Res. **18**, 3649 (2017).

[31] A. Wilson and H. Nickisch, “Kernel interpolation for scalable structured Gaussian processes (KISS-GP)”, International Conference on Machine Learning, 1775, PMLR: vol. 37, Lille, France (2015).

[32] A. G. Wilson, C. Dann, and H. Nickisch, “Thoughts on massively scalable Gaussian processes”, (2015), `arXiv:1511.01870`.

[33] J. Gardner, G. Pleiss, R. Wu, K. Weinberger, and A. Wilson, “Product kernel interpolation for scalable Gaussian processes”, International Conference on Artificial Intelligence and Statistics, 1407, PMLR: vol. 84, Lanzarote, Spain (2018).

[34] G. Pleiss, J. Gardner, K. Weinberger, and A. G. Wilson, “Constant-time predictive distributions for Gaussian processes”, International Conference on Machine Learning, 4114, PMLR: vol. 80, Stockholm, Sweden (2018).

[35] S. Seo, M. Wallat, T. Graepel, and K. Obermayer, “Gaussian process regression: active

data selection and test point rejection", International Joint Conference on Neural Networks, vol. 3, 241, IEEE-INNS-ENNS (2000).

[36] M. H. Y. Tan, "Gaussian process modeling with boundary information", Stat. Sin. **28**, 621 (2018).

[37] F. Grossmann, *Theoretical femtosecond physics: atoms and molecules in strong laser fields*, Springer (2018).

[38] D. D. Cox and S. John, "A statistical method for global optimization", International Conference on Systems, Man, and Cybernetics, vol. 2, 1241, IEEE (1992).

[39] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, "Taking the human out of the loop: a review of Bayesian optimization", Proc. IEEE **104**, 148 (2016).

[40] M. U. Gutmann and J. Corander, "Bayesian optimization for likelihood-free inference of simulator-based statistical models", J. Mach. Learn. Res. **17**, 1 (2016).

[41] N. Giannakis, N. Gorgolis, and I. Hatzilygeroudis, "Bayesian optimization for the design of deep neural networks", International Conference on Information, Intelligence, Systems & Applications, 1, IEEE (2021).

[42] E. Schrödinger, "An undulatory theory of the mechanics of atoms and molecules", Phys. Rev. **28**, 1049 (1926).

[43] M. Born and V. Fock, "Beweis des Adiabatensatzes", Z. Phys. **51**, 165 (1928).

[44] L. Landau, "Zur Theorie der Energieubertragung. II", Phys. Z. Sowjetunion **2**, 46 (1932).

[45] C. Zener, "Non-adiabatic crossing of energy levels", Proc. R. Soc. Lond. A **137**, 696 (1932).

[46] E. C. G. Stueckelberg, "Theorie der unelastischen Stösse zwischen Atomen", Helv. Phys. Acta **5**, 369 (1932).

[47] E. Majorana, "Atomi orientati in campo magnetico variabile", Nuovo Cim. **9**, 43 (1932).

[48] R. Loudon, "One-dimensional hydrogen atom", Am. J. Phys. **27**, 649 (1959).

[49] J. von Neumann and E. P. Wigner, "Über das Verhalten von Eigenwerten bei adiabatischen Prozessen", Z. Phys. **30**, 467 (1929).

[50] D. R. Yarkony, "Diabolical conical intersections", Rev. Mod. Phys. **68**, 985 (1996).

[51] M. Baer, "Adiabatic and diabatic representations for atom-molecule collisions: Treatment of the collinear arrangement", Chem. Phys. Lett. **35**, 112 (1975).

[52] M. Baer, "Introduction to the theory of electronic non-adiabatic coupling terms in molecular systems", Phys. Rep. **358**, 75 (2002).

[53] T. Takami, "Semiclassical study of avoided crossings", Phys. Rev. E **52**, 2434 (1995).

[54] A. Sommerfeld, "Zur Quantentheorie der Spektrallinien", Ann. Phys. **356**, 1 (1916).

[55] A. Engel, "On the quantum theorem of Sommerfeld and Epstein", The Collected Papers of Albert Einstein **6**, 434 (1997).

[56] A. D. Stone, "Einstein's unknown insight and the problem of quantizing chaos", Phys. Today **58**, 37 (2005).

[57] G. Wentzel, "Eine Verallgemeinerung der Quantenbedingungen für die Zwecke der Wellenmechanik", Z. Phys. **38**, 518 (1926).

[58] H. A. Kramers, "Wellenmechanik und halbzahlige Quantisierung", Z. Phys. **39**, 828

(1926).

[59] L. Brillouin, “La mécanique ondulatoire de Schrödinger; une méthode générale de resolution par approximations successives”, C. R. Acad. Sci. **183**, 24 (1926).

[60] J. B. Keller, “Corrected Bohr-Sommerfeld quantum conditions for nonseparable systems”, Ann. Phys. **4**, 180 (1958).

[61] I. Percival and N. Pomphrey, “Vibrational quantization of polyatomic molecules”, Molec. Phys. **31**, 97 (1976).

[62] S. K. Knudson, J. B. Delos, and D. W. Noid, “Bound state semiclassical wave functions”, J. Chem. Phys. **84**, 6886 (1986).

[63] I. C. Percival, “Regular and irregular spectra”, J. Phys. B **6**, L229 (1973).

[64] K. Sohlberg and R. B. Shirts, “Semiclassical quantization of a nonintegrable system: Pushing the Fourier method into the chaotic regime”, J. Chem. Phys. **101**, 7763 (1994).

[65] M. C. Gutzwiller, “Periodic orbits and classical quantization conditions”, J. Math. Phys. **12**, 343 (1971).

[66] M. C. Gutzwiller, *Chaos in Classical and Quantum Mechanics*, Springer New York (1990).

[67] M. V. Berry and M. Tabor, “Closed orbits and the regular bound spectrum”, Proc. R. Soc. London. Ser. A **349**, 101 (1976).

[68] M. V. Berry and M. Tabor, “Calculating the bound spectrum by path summation in action-angle variables”, J. Phys. A **10**, 371 (1977).

[69] V. I. Arnold, *Mathematical Methods of Classical Mechanics*, Springer New York (1989).

[70] A. N. Kolmogorov, “On the conservation of conditionally periodic motions under small perturbation in Hamilton’s function”, Dokl. Akad. Nauk SSSR **98**, 527 (1954).

[71] V. I. Arnol’d, “Proof of a theorem of A. N. Kolmogorov on the invariance of quasi-periodic motions under small perturbations of the Hamiltonian”, Russ. Math. Surv. **18**, 9 (1963).

[72] V. I. Arnol’d, “Small denominators and problems of stability of motion in classical and celestial mechanics”, Russ. Math. Surv. **18**, 85 (1963).

[73] J. Moser, “On invariant curves of area-preserving mappings of an annulus”, Nachr. Akad. Wiss. Gött., II 1 (1962).

[74] S. K. Knudson and D. W. Noid, “The multidimensional WKB method”, J. Chem. Educ. **66**, 133 (1989).

[75] H. Poincaré, *Les méthodes nouvelles de la mécanique céleste*, volume 2, Gauthier-Villars et fils, imprimeurs-libraires, Paris (1893).

[76] D. W. Noid, S. K. Knudson, and B. G. Sumpter, “Exact semiclassical calculation of eigenvalues for multidimensional systems using SOS”, Comput. Phys. Commun. **51**, 11 (1988).

[77] H. R. Dullin and A. Wittek, “Complete Poincaré sections and tangent sets”, J. Phys. A **28**, 7157 (1995).

[78] D. W. Noid and R. A. Marcus, “Semiclassical calculation of eigenvalues for higher order resonances in nonseparable oscillator systems”, J. Chem. Phys. **85**, 3305 (1986).

[79] D. W. Noid, M. L. Koszykowski, and R. A. Marcus, “Semiclassical calculation of eigen-

values for a three-dimensional system", J. Chem. Phys. **73**, 391 (1980).

[80] S. K. Knudson and D. W. Noid, "The semiclassical surface-of-section method in three dimensions", Chem. Phys. Lett. **145**, 16 (1988).

[81] C. W. Eaker, G. C. Schatz, N. D. Leon, and E. J. Heller, "Fourier transform methods for calculating action variables and semiclassical eigenvalues for coupled oscillator systems", J. Chem. Phys. **81**, 5913 (1984).

[82] J. Binney and D. Spergel, "Spectral stellar dynamics - II. the action integrals", Mon. Not. R. Astron. Soc. **206**, 159 (1984).

[83] C. C. Martens and G. S. Ezra, "EBK quantization of nonseparable systems: A Fourier transform method", J. Chem. Phys. **83**, 2990 (1985).

[84] B. G. Sumpter and D. W. Noid, "Method for semiclassical calculation of grid of eigenvalues", Chem. Phys. Lett. **126**, 181 (1986).

[85] D. W. Noid, M. L. Koszykowski, and R. A. Marcus, "Comparison of quantal, classical, and semiclassical behavior at an isolated avoided crossing", J. Chem. Phys. **78**, 4018 (1983).

[86] M. J. Davis and E. J. Heller, "Quantum dynamical tunneling in bound states", J. Chem. Phys. **75**, 246 (1981).

[87] W. H. Miller, "Periodic orbit description of tunneling in symmetric and asymmetric double-well potentials", J. Phys. Chem. **83**, 960 (1979).

[88] E. Thiele and D. J. Wilson, "Anharmonicity in unimolecular reactions", J. Chem. Phys. **35**, 1256 (1961).

[89] P. M. Morse, "Diatomic molecules according to the wave mechanics. II. Vibrational levels", Phys. Rev. **34**, 57 (1929).

[90] M. L. Sage, "Morse oscillator transition probabilities for molecular bond modes", Chem. Phys. **35**, 375 (1978).

[91] C. G. Schlier, "Collinear collisions in a double-Morse well", Chem. Phys. **105**, 361 (1986).

[92] W. E. Smyser and D. J. Wilson, "Quantum dynamics of triatomic molecules", J. Chem. Phys. **50**, 182 (1969).

[93] W. H. Press, *Numerical recipes 3rd edition: The art of scientific computing*, Cambridge University Press (2007).

[94] R. M. Hedges and W. P. Reinhardt, "Very long-lived doubly vibrationally excited states above dissociation threshold: non-RRKM behavior in model triatomic systems", Chem. Phys. Lett. **91**, 241 (1982).

[95] R. M. Hedges and W. P. Reinhardt, "Classical and quantum dynamics of long lived doubly excited vibrational states of triatomic molecules", J. Chem. Phys. **78**, 3964 (1983).

[96] R. F. Nalewajski and R. E. Wyatt, "Collisional perturbation of regular and irregular intramolecular dynamics: A classical dynamical study", Chem. Phys. **81**, 357 (1983).

[97] T. Matsushita, A. Narita, and T. Terasaka, "Chaotic behavior of a classical coupled Morse system around the escape energy region", Chem. Phys. Lett. **95**, 129 (1983).

[98] T. Matsushita and T. Terasaka, "Mass dependence of the Kolmogorov-Arnold-Moser stability and low-order resonances in the kinetically coupled two-degree-of-freedom Morse system", Chem. Phys. Lett. **100**, 138 (1983).

[99] T. Matsushita and T. Terasaka, "A connection between classical chaos and the quantized energy spectrum: Level-spacing distributions in a kinetically coupled quantum Morse system with two degrees of freedom", Chem. Phys. Lett. **105**, 511 (1984).

[100] T. Terasaka and T. Matsushita, "Statistical properties of the quantized energy spectrum of a Hamiltonian system with classically regular and chaotic trajectories: A numerical study of level-spacing distributions for two-dimensional coupled Morse-oscillator systems", Phys. Rev. A **32**, 538 (1985).

[101] I. Hamilton and P. Brumer, "A minimally dynamic approach to unimolecular decay: CCH and coupled Morse dynamics", J. Chem. Phys. **82**, 1937 (1985).

[102] C. Löbner, S. Löck, A. Bäcker, and R. Ketzmerick, "Integrable approximation of regular islands: The iterative canonical transformation method", Phys. Rev. E **88**, 062901 (2013).

[103] C. Jaffé and P. Brumer, "Local and normal modes: A classical perspective", J. Chem. Phys. **73**, 5646 (1980).

[104] T. Feldmann, M. Cohen, and B. L. Burrows, "Coupled harmonic oscillator systems: An elementary algebraic decoupling approach", J. Math. Phys. **41**, 5897 (2000).

[105] G. D. Birkhoff, "An extension of Poincaré's last geometric theorem", Acta Math. **47**, 297 (1925).

[106] T. Uzer, D. W. Noid, and R. A. Marcus, "Uniform semiclassical theory of avoided crossings", J. Chem. Phys. **79**, 4412 (1983).

[107] R. T. Lawton and M. S. Child, "Local mode vibrations of water", Mol. Phys. **37**, 1799 (1979).

[108] D. Wintgen and H. Friedrich, "Correspondence of unstable periodic orbits and quasi-Landau modulations", Phys. Rev. A **36**, 131 (1987).

[109] H. Friedrich and H. Wintgen, "The hydrogen atom in a uniform magnetic field — an example of chaos", Phys. Rep. **183**, 37 (1989).

[110] A. R. Edmonds, "Studies of the quadratic Zeeman effect. I. Application of the Sturmian functions", J. Phys. B **6**, 1603 (1973).

[111] C. W. Clark and K. T. Taylor, "The quadratic Zeeman effect in hydrogen Rydberg series: application of Sturmian functions", J. Phys. B **15**, 1175 (1982).

[112] G. Wunner, M. Kost, and H. Ruder, ""Circular" states of Rydberg atoms in strong magnetic fields", Phys. Rev. A **33**, 1444 (1986).

[113] A. A. Kotzé, *Quantum Chaos and Analytic Structure of the Spectrum*, Ph.D. thesis, University of the Witwatersrand, Johannesburg, South Africa (1992).

[114] A. A. Kotzé and W. D. Heiss, "The quantum spectrum and the distribution of exceptional points: the hydrogen atom in a strong magnetic field", J. Phys. A **27**, 3059 (1994).

[115] G. Tanner, K. T. Hansen, and J. Main, "The semiclassical resonance spectrum of hydrogen in a constant magnetic field", Nonlinearity **9**, 1641 (1996).

[116] J. B. Delos, S. K. Knudson, and D. W. Noid, "Trajectories of an atomic electron in a magnetic field", Phys. Rev. A **30**, 1208 (1984).

[117] D. Delande and J. C. Gay, "Quantum chaos and statistical properties of energy levels: Numerical study of the hydrogen atom in a magnetic field", Phys. Rev. Lett. **57**, 2006 (1986).

[118] J. B. Delos, S. K. Knudson, and D. W. Noid, “Highly excited states of a hydrogen atom in a strong magnetic field”, Phys. Rev. A **28**, 7 (1983).

[119] D. W. Noid, M. L. Koszykowski, and R. A. Marcus, “Calculations related to quantum stochasticity, an example of overlapping avoided crossings”, Chem. Phys. Lett. **73**, 269 (1980).

[120] M. V. Berry, “Quantizing a classically ergodic system: Sinai’s billiard and the KKR method”, Ann. Phys. **131**, 163 (1981).

[121] N. Rosenzweig and C. E. Porter, ““Repulsion of energy levels” in complex atomic spectra”, Phys. Rev. **120**, 1698 (1960).

[122] R. L. Waterland, J. B. Delos, and M. L. Du, “High Rydberg states of an atom in parallel electric and magnetic fields”, Phys. Rev. A **35**, 5064 (1987).

[123] M. J. Davis, “Chaotic dynamics and approximate semiclassical quantization: Assigning highly excited vibrational eigenstates”, J. Chem. Phys. **107**, 106 (1997).

[124] C. C. Martens and G. S. Ezra, “Classical and semiclassical mechanics of strongly resonant systems: A Fourier transform approach”, J. Chem. Phys. **86**, 279 (1987).

[125] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory*, volume 3, Pergamon Press, Oxford (1991).

[126] R. Wallace, “Analytic quantum mechanics of the Morse oscillator”, Chem. Phys. Lett. **37**, 115 (1976).

[127] J. B. Tatum, *Classical Mechanics*, University of Victoria (2013).

[128] K. R. Meyer, “Scaling Hamiltonian systems”, SIAM J. Math. Anal. **15**, 877 (1984).

[129] T. M. Rassias and H. M. Srivastava, “The orthogonality property of the classical laguerre polynomials”, Appl. Math. Comput. **50**, 167 (1992).

Acknowledgements

I owe thanks to a number of people, without whom I would have been entirely lost in the vast space of science and research.

First and foremost I deeply thank my supervisor Prof. Dr. Jan-Michael Rost for this great honor to contribute to science in the name of the Max Planck Society. I am thankful for the guidance through this scientific adventure and for the introduction to machine learning, which has become a passion to me. I value the freedom to explore and the opportunity to work on subjects aligned with my interests. I have come to understand what it means to deeply comprehend a matter thoroughly, and how to work precisely and autonomously when overcoming challenges. His mentoring allowed me to truly grow in my abilities.

I deeply appreciate the willingness of Prof. Dr. Roman V. Krems to review my dissertation.

I’m thankful for all the help that was generously granted to me by my many colleagues and leaders at the MPI PKS. My gratitude goes to the scientific group leaders of the Finite Systems division for all the helpful insight on various matters: Prof. Dr. Ulf Saalmann, Dr. Alexander Eisfeld, Dr. Matthew Eiles, and Dr. Panagiotis Giannakeas. Also, I thank all my colleagues, especially Andrew Hunter and Dr. Jonathan Dubois, for all the helpful chats and quick advice.

I thank Prof. Dr. Frank Großmann for taking me in at the beginning and for paving the way to the MPI PKS.

Shout out to Dr. Mohammad Reza Eidi, with whom I had an exciting time juggling Gaussians, and for the additional challenges we faced that helped me to better understand my own research.

Special thanks goes to my friend Dr. Sebastian Gemsheim for all the time and the immense help over the years. It was an epiphany to witness him work and to soak up his brilliance.

I thank Dr. Felix Fritzsch for his expertise and helpful discussions, as well as for proofreading my dissertation.

For all the time off the field I thank my beloved wife for making this formidable task bearable, and for all the love and care she fills my life with every day. Ich danke meinen Eltern für ihre bedingungslose Unterstützung und Fürsorge zu jeder Zeit.

Lastly, my deepest gratitude goes to Stephanie and Ted Guerry, whose extraordinary generosity, care, and love have helped pave the way for my journey, leading me to where I am today.

’Well, this is the end, Sam Gamgee,’ said a voice by his side. And there was Frodo, pale and worn, and yet himself again; and in his eyes there was peace now, neither strain of will, nor madness, nor any fear. His burden was taken away. There was the dear master of the sweet days in the Shire.

Versicherung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Diese Dissertation wurde unter der Betreuung von Prof. Dr. Jan-Michael Rost am Max-Planck-Institut für Physik komplexer Systeme in Dresden angefertigt.

Ich erkläre hiermit, dass keine früheren erfolglosen Promotionsverfahren stattgefunden haben. Ich erkenne die Promotionsordnung des Bereiches Mathematik und Naturwissenschaften der Technischen Universität Dresden an.

Dresden, 30.01.2024
Stefan Benjamin Rabe