UNIVERSITÉ DU LUXEMBOURG

PhD-FSTM-2025-112
The Faculty of Science, Technology and Medicine

DISSERTATION

Defence held on 25/09/2025 in Luxembourg

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN PHYSIQUE

by

Kyunghoon HAN
Born on 7 August 1990 in city of Seoul, (Republic of Korea)

ANALYSIS, CALCULATION AND UNDERSTANDING OF MOLECULAR VIBRATIONS:

DECOMPOSITION OF EMPIRICAL SPECTRA,
SIMULATION TARGETED TO A FREQUENCY BAND, AND
THEORETICAL FRAMEWORK OF OSCILLATORY MOTION AS FAT MANIFOLDS

Dissertation defence committee

Dr Joshua T. Berryman, dissertation supervisor
Research Scientist, Université du Luxembourg

Dr Aurélia Chenu, Chairman
Professor, Université du Luxembourg

Dr Francesca Ingrosso
Professor, Université de Lorraine

Dr Benjamin Stamm
Professor, Universität Stuttgart

Dr Christophe Ley
Professor, Université du Luxembourg

Affidavit / Statement of originality

I declare that this thesis:

- is the result of my own work. Any contribution from any other party, and any use of generative artificial intelligence technologies have been duly cited and acknowledged;
- is not substantially the same as any other that I have submitted, and;
- is not being concurrently submitted for a degree, diploma or other qualification at the University of Luxembourg or any other University or similar institution except as specified in the text.

With my approval I furthermore confirm the following:

- I have adhered to the rules set out in the University of Luxembourg's Code of Conduct and the Doctoral Education Agreement (DEA)[1], in particular with regard to Research Integrity.
- I have documented all methods, data, and processes truthfully and fully.
- I have mentioned all the significant contributors to the work.
- I am aware that the work may be screened electronically for originality.

I acknowledge that if any issues are raised regarding good research practices based on the review of the thesis, the examination may be postponed pending the outcome of any investigation of such issues. If a degree was conferred, any such subsequently discovered issues may result in the cancellation of the degree.

Approved on 2025-08-20

[1] If applicable (DEA is compulsory since August 2020)

Analysis, Calculation, and Understanding of Molecular Vibrations:

decomposition of empirical spectra, simulation targeted to a frequency band, and theoretical framework of oscillatory motion as fat manifolds

by

Kyunghoon Han

A thesis submitted to the University of Luxembourg
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

PHYSICS

Department of Physics and Materials Science
University of Luxembourg
Luxembourg

Abstract

The idea of molecular dynamics seems intuitively simple, that atoms should move through space on complex but well defined trajectories that reflect the chemical interactions between them. Unfortunately, atoms are quantum objects therefore to define even the concept of movement, for example as the change of a well-defined position versus a smoothly evolving time coordinate, is problematic.

In the present thesis I address molecular motion beginning from the fundamentals, presenting initially a software tool for analysis of empirically obtained spectra such that the optical (or acoustic) properties of a material can be connected to basic statements about its inherent vibrational resonances. Moving on from this empirically-grounded signal processing approach I develop a framework for selectively interrogating dynamics when confined to a specific window of frequency space, providing a toolkit to answer the question “what molecular motion connects to what observed peaks in a spectrum”.

The final mathematical approach offers a novel perspective on molecular motion derived from the tradition of Riemannian geometry, framing molecular trajectories as geodesics on a Riemannian manifold shaped by the Jacobi metric. In this setting, vibrational motion emerges as curvature-driven structure, captured without recourse to predefined coordinates or harmonic approximations. The Riemannian approach offers a novel, efficient, and highly robust integration algorithm which can be used to produce the geodesics of complex systems in a way that is naturally adapted to interrogate vibrational spectra, even in the presence of relativistic and quantum effects.

Résumé

L'idée de la dynamique moléculaire peut sembler, à première vue, intuitivement simple : les atomes devraient se déplacer dans l'espace selon des trajectoires complexes mais bien définies, reflétant les interactions chimiques qui s'exercent entre eux. Cependant, les atomes étant des objets quantiques, la définition même de la notion de mouvement devient problématique — par exemple, comme la variation d'une position parfaitement déterminée par rapport à une coordonnée temporelle évoluant de manière continue.

Dans cette thèse, j'examine le mouvement moléculaire à partir de ses principes fondamentaux. Je présente d'abord un outil logiciel d'analyse de spectres expérimentaux, reliant les propriétés optiques (ou acoustiques) d'un matériau à ses résonances vibrationnelles intrinsèques. Partant de cette approche empirique du signal, je propose ensuite un cadre permettant d'examiner sélectivement la dynamique dans une plage spécifique de fréquences, afin de répondre à la question : « Quel mouvement moléculaire correspond à quels pics observés dans un spectre ? »

Enfin, l'approche que je propose renouvelle l'étude du mouvement moléculaire en l'inscrivant dans le cadre de la géométrie riemannienne, où les trajectoires sont traitées comme des géodésiques définies par la métrique de Jacobi. Le mouvement vibrationnel y apparaît comme une structure dictée par la courbure, sans recours à des coordonnées prédéfinies ni à l'approximation harmonique. Cette formulation conduit à un algorithme d'intégration efficace et robuste, adapté au calcul des géodésiques de systèmes complexes ainsi qu'à l'analyse des spectres vibrationnels, y compris en présence d'effets relativistes et quantiques.

Zusammenfassung

Die Grundidee der Molekulardynamik wirkt zunächst schlicht: Atome bewegen sich auf komplexen, aber wohldefinierten Bahnen durch den Raum – geprägt von den chemischen Wechselwirkungen zwischen ihnen. Doch Atome sind Quantenobjekte; schon der Begriff der Bewegung, etwa als Änderung einer wohldefinierten Position entlang einer fortschreitenden Zeitkoordinate, erweist sich als problematisch.

Diese Arbeit setzt bei den Grundlagen an. Zunächst stelle ich ein Software zur Auswertung empirisch gewonnener Spektren vor, das optische (bzw. akustische) Materialeigenschaften mit grundlegenden Aussagen über die zugrundeliegenden Eigenresonanzen verknüpft. Aufbauend auf diesem empirisch fundierten, signalverarbeitenden Vorgehen entwickle ich einen Ansatz, mit den sich Dynamik gezielt innerhalb eines ausgewählten Frequenzfensters untersuchen lässt – ein methodischer Werkzeugkasten, der die Frage beantwortet, welche molekularen Bewegungen welchen beobachteten Peaks im Spektrum zuzuordnen sind.

Abschließend stelle ich einen auf der Riemannschen Geometrie basierenden Lösungsansatz vor: Molekulare Trajektorien erscheinen als Geodäten auf einer durch die Jacobi-Metrik bestimmten Riemannschen Mannigfaltigkeit. In dieser Darstellung treten Schwingungen als krümmungsgetriebene Strukturen hervor – erfasst ohne Bezug auf vorgegebene Koordinaten oder harmonische Näherungen. Der Riemannsche Ansatz bietet einen neuartigen, effizienten und äußerst robusten Integrationsalgorithmus, mit dem sich die Geodäten komplexer Systeme auf eine Weise erzeugen lassen, die sich auf natürliche Weise für die Untersuchung von Schwingungsspektren eignen, selbst wenn relativistische und quantenmechanische Effekte vorliegen.

초록

분자 동역학의 개념은 표면적으로는 직관적으로 단순해 보이지만, 원자는 본질적으로 양자적 대상이므로 시간에 따라 명확히 정의된 위치의 변화를 운동으로 기술하는 것은 근본적으로 문제가 된다.

본 논문은 이러한 분자 운동의 개념을 근본 원리에서 출발하여 새롭게 탐구한다. 먼저, 화학적 세부 지식에 의존하지 않고도 다양한 스펙트럼에 폭 넓게 적용 가능한 보편적 분석 체계를 제안한다. 이를 통해 재료의 광학적·음향적 성질을 고유한 진동 공명과 직접적으로 연결할 수 있게 된다. 이어서, 주파수 영역을 선택적으로 한정하여 분자 동역학을 분석하는 이론적 프레임워크를 구축함으로써, “특정 스펙트럼의 최고점의 근방은 어떤 분자 운동과 대응하는가?”라는 질문에 체계적으로 답할 수 있도록 한다.

마지막으로, 리만 기하학적인 전통에 기반한 새로운 수학적 접근을 제안한다. 분자의 궤적을 야코비 계량으로 정의된 리만 다양체 위의 측지선으로 해석하며, 이 틀에서 진동 운동은 곡률에 의해 드러나는 기하학적 구조로 나타난다. 이 방법은 좌표계나 조화 근사에 의존하지 않고도 분자 운동을 기술할 수 있으며, 효율적이고 안정적인 새로운 적분 알고리즘을 제공한다. 나아가 양자적·상대론적 효과가 작용하는 체계에서도 분자 궤적과 스펙트럼의 연결을 자연스럽게 탐구할 수 있게 한다.

Acknowledgments

यस्मा हि धम्मं पुरसिो वजिञ्ञा,
इन्दंव नं देवता पूजयेय्य;
सो पूजतिो तस्मि पिसन्नचत्तिो,
बहुस्सुतो पातुकरोति धम्मं।

गौतमबुद्धः
सूत्र-नपिात

If you would learn the teaching from someone, honour them as the gods honour Indra. Then they will have confidence in you, and being learned, they reveal the teaching.[1]

The Buddha
Sutta Nipāta (tr. Bhikkhu Sujato)

From a career perspective, entering doctoral studies at this stage—leaving behind a stable position and established credentials—may well have appeared a kind of folly. I was called upon to immerse myself in topics that, at first, stirred little curiosity. There were moments when regret settled in, and the path ahead seemed uncertain.

And yet, it was the people—friends, mentors, colleagues—who gave me reason to persevere. Through their support, I came to see that even in a field where my interests often stood apart, there remained, however modest, a place for my contributions. Some of the knowledge I carried could take root, and, in time, bear fruit.

The Kindly Light

My supervisor, Dr Joshua T. Berryman, showed rare generosity and faith in accepting me as his student. Though my background diverged from the typical path, he welcomed my perspective with openness and patience.

[1] Throughout this thesis, all translations are mine unless otherwise indicated.

As the project evolved, I found myself gradually turning toward mathematics—a long-standing compass for my intellectual life. The tools and approaches I brought to bear were sometimes unfamiliar, and occasionally at odds with those in use around me. Yet from this tension arose insight. Our exchanges—challenging and at times dissonant—deepened my understanding of the fragile interface between theory and empirical study.

I remain particularly grateful for Dr Berryman's patience. He entertained a number of ideas I proposed, some well beyond the scope of the group's work and others of doubtful utility. That he allowed me this freedom is no small thing; it helped shape not only this thesis, but also my understanding of what it means to think independently within a scientific field.

One of Dr Berryman's efforts to bridge theory and experiment was to introduce me to the laboratory of Dr Francesco Simone Ruggeri at Wageningen University. With great patience, Dr Ruggeri shared his perspectives and guided me through the methods that ground his experimental research. I am sincerely grateful for the chance to learn from him and to witness how experimental and theoretical work might inform each other.

I also wish to thank Dr Alexandre Tkatchenko, who gave me the opportunity to join the Theoretical Chemical Physics group at the University of Luxembourg. His openness in allowing me to pursue an independent line of inquiry—even when it lay outside the group's core interests—was deeply appreciated. In time, those mathematical explorations began to resonate with broader themes in the group's work. Our wide-ranging conversations—scientific and otherwise—were a source of great learning and lasting value.

To the members of my CET committee, I offer heartfelt thanks. Their advice and guidance were indispensable. I am especially grateful to Dr Francesca Ingrosso of Université de Lorraine for her warm support throughout this journey. It was a privilege to witness, if only briefly, the academic life of that institution through her. I am also grateful to Dr Ulf Leonhardt, whose thoughtful remarks on the Riemannian framework employed in this thesis helped sharpen

its direction.

Thanks are also due to Dr Christophe Ley, who introduced me to statistical copulas. It was a joy to explore how these structures might converse with the language of statistical mechanics.

Finally, I extend my sincere gratitude to Dr Benjamin Stamm and Dr Aurélia Chenu for generously agreeing to serve on my defence jury, even on short notice. Their presence at the conclusion of this work means more than I can easily express.

I acknowledge the financial support provided by the Fonds Nationale de la Recherche (FNR, C20/MS/14588607) and by the University of Luxembourg, without which this work would not have been possible.

Fellow Pilgrims

From the moment I joined the Theoreitcal Chemical Physics (TCP) group, I was received with warmth and generous camaraderie. Yet beyond such kindness, it was the conversations shared among colleagues that proved a continual source of intellectual joy. In their company, I encountered not only fellowship, but a wealth of perspectives that challenged and enriched my own.

I am grateful to Dr Matthieu Sarkis for the many intellectually rich discussions we shared. His perspective on mathematical structures often illuminated aspects of the problem space I might have otherwise overlooked. Our exchanges, particularly regarding algebraic approaches to molecular dynamics, contributed significantly to the conceptual breadth of this project.

To Dr Matteo Gori and Dr Loris Di Cairano, I owe thanks for the many conversations that drew connections between molecular dynamics, chaos theory, and differential geometry. Engaging with their prior work allowed me to refine the geometrical interpretation of molecular systems explored here. It was also through Dr Gori that I encountered the concept of fat manifolds, which later became a central object in this thesis.

In my collaboration with Dr Ariadni Boziki on the development of the application *Tihi* [1], I gained a clearer understanding of the spectroscopic perspec-

tive in chemical analysis. Her approach demonstrated how mathematical and chemical intuitions may complement each other in interpreting experimental data.

I also wish to acknowledge those with whom I shared the office, whose interdisciplinary expertise and day-to-day exchanges offered a productive and collegial environment. From Dr Jorge Alfonso Charry Martinez and Dr Szabolcs Góger, I gained deeper exposure to both theoretical and experimental chemistry. From Dr Gregory Cordeiro Fonseca, I found a means of integrating my background in artificial intelligence with methods in computational physics. Discussions with Tobias Henkes further expanded my understanding of chemistry, and our shared linguistic exchanges added an unexpected dimension to our doctoral studies. Anton Charkin-Gorbulin provided connections to research communities working in myrmecology, a field of longstanding personal interest. I am also grateful to Jaume Alexandre Solé Gómez, whose brief stay in our office revived fruitful discussions around developments in AI.

I wish to thank Sergio Suárez Dou and Nils Davoine, fellow doctoral researchers under the same supervision, whose advanced understanding of molecular simulations helped to clarify the role of molecular dynamics within broader biological and chemical contexts. Their practical insights and many discussions offered essential grounding for the more abstract components of this work.

Lastly, I extend my gratitude to the colleagues in the office across from ours. With Dhruv Sharma, every exchange brimmed with invention. To Andronikos Leventis, I owe thanks for discourse on the subtle kinship betwixt random matrices and molecular spectra. To Mirela Puleva, whose conversation ranged from ancient languages to the laws of nature, full of excitements and warm appreciations.

A Still Point in a Turning World

I extend my thanks to Dr Kyung Hoon Han of Suwon University (no relation), who permitted me to attend a private seminar series in mathematics during

my time in industry. That continued exposure to academic discourse played a quiet yet formative role in sustaining my engagement with mathematics. I also thank Dr Donghoon Park of Yonsei University for his valuable insights across mathematics, physics, and philosophy, and Dr Hun Hee Lee for his efforts during our collaborative work on a funding proposal.

I am further grateful to Dr Jiwoo Bae, CEO of SRuniverse, for supporting my return to academia despite my professional responsibilities at the time. I also acknowledge my former colleagues—Sanguk Han, Myeongjae Yi, Hanbin Kim, Seongbo Yim, Taehoon Goo, and Alena Shilina—for their openness to mathematically driven research directions in artificial intelligence. Their patience and goodwill during that period contributed, in no small part, to the path that followed.

I also wish to acknowledge my fellow alumni of the University of Waterloo, the Waterloo Physics Crew: Dr David Luong, Carlos Wang, Dr Ian Lam, and Jeeyoung Nam. My conversations with David ranged widely across subjects, and often yielded intuitive insights of direct relevance to the present research. Our continued exchanges, both individual and collective, provided a source of intellectual stimulation and support throughout the course of my doctoral work.

I must also acknowledge the steadfast support of my parents, Dr Wanok Han and Kyeonghee Lee, whose encouragement made possible my return from industry to the academic path. Their confidence in the value of scholarship, and their readiness to support its pursuit, formed a foundation upon which this work could be undertaken. I am, likewise, grateful to my parents-in-law, Mija Seo and Bongsoon Shin, for their understanding and unfailing goodwill during this transition.

Lastly, I wish to express my deepest gratitude to my wife, Younkyeong Shin. Her willingness to undertake the challenges of life in a foreign country, adapting to new customs and languages, required both courage and resolve. Her understanding, patience, and unwavering support have been indispensable to

the completion of this work.

August 20th, 2025
Luxembourg,
Grand Duchy of Luxembourg

Kyunghoon HAN 한경훈

Table of Contents

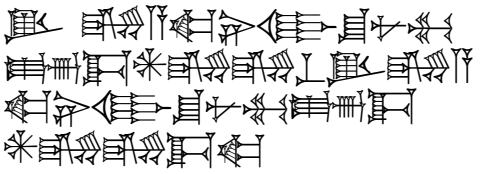
Ce livre est né d'un texte de Borges··· le charme énigmatique de cette taxinomie, l'exotisme de la pensée qui l'a rendue possible···

MICHEL FOUCAULT
Les Mots et les Choses

This book first arose from a text by Borges··· the enigmatic charm of that taxonomy, the exoticism of the thought that made it possible···

MICHEL FOUCAULT
The Order of Things

List of Tables



...Because the messenger, whose mouth was tired, was not able to repeat it, the lord of Kulaba patted some clay and wrote the message as if on a tablet...

ENMERKAR AND THE LORD OF ARATTA
A Sumerian epic (tr. Jeremy A. Black)

List of Figures

τὸ δὲ παρὸν ὅτε πάρεστιν, οἷον τοδὶ τὸ λευκὸν ὅτε ὁρᾷ, οὐδεὶς ἂν φαίη μνημονεύειν, οὐδὲ τὸ θεωρούμενον, ὅτε θεωρῶν τυγχάνει καὶ νοῶν, ἀλλὰ τὸ μὲν αἰσθάνεσθαί φησι, τὸ δ' ἐπίστασθαι μόνον·

ΠΕΡΙ ΨΥΧΗΣ
Ἀριστοτέλης

No one would claim to remember the present thing when it is present, such as this white thing here when one sees it, and not the contemplative object when one happens to be contemplating and considering it; rather one claims only to perceive the former and to know the latter.

ARISTOTLE
On the Soul (tr. Evan Strevell)

CHAPTER 1

Introduction

C'est le temps que tu as perdu pour ta rose qui fait ta rose si importante.

ANTOINE DE SAINT-EXUPÉRY
Le Petit Prince

It is the time you have lost for your rose that makes your rose so important.

ANTOINE DE SAINT-EXUPÉRY
The Little Prince

The study of molecular vibrations predates the modern understanding of quantum mechanics as known today. Long before the quantisation of energy levels was established, the scientific community had already begun probing the vibrational signatures of matter through their interaction with light.

During the 1860s, John Tyndall showed that gases such as carbon dioxide selectively absorb radiant heat—what is now identified as infrared radiation—thus providing the foundation for the contemporary understanding of the greenhouse effect [3]. Building on Kirchhoff's foundational work on thermal radiation [4], Tyndall went further, proposing a physical explanation for the observed absorption patterns. He suggested that molecules absorb radiation most efficiently when the wave frequencies match their internal modes of vibration, a view he attributed directly to Kirchhoff, as illustrated in the following passage from his work [3]:

> By Kirchhoff it has been conclusively shown that every atom absorbs in a special degree those waves which are synchronous with its own periods of vibration.

In this context, to the 19th century physicists, molecular vibrations could

be defined as:

> **Definition 1.1.** Regular patterns in the absorption of radiant energy by gases, observed at specific frequencies, is interpreted as evidence of internal periodic motions within matter—motions resonant with the incident waves and indicative of vibratory behaviour.

After the emergence of atomic theory and the birth of quantum mechanics in the early 20th century, the notion of internal periodic motion—previously treated as a speculative explanation—began to acquire precise mechanical meaning. With the formulation of Bohr's model and quantum quantisation rules, scientists could now describe internal vibrations as discrete energy levels of bounded particles, governed by dynamical laws.

Yet, in this formative stage, the tools of fully developed quantum mechanics were still lacking. As a result, early pioneers like Adolf Kratzer worked with a semi-classical framework: combining classical mechanics with quantisation rules (e.g. Bohr–Sommerfeld conditions [5, 6]) to describe the vibrational and rotational spectra of molecules [7].

> **Definition 1.2.** Molecular vibrations are the periodic relative motions of atomic centres in a molecule, treated as a classical oscillator. Allowed vibrational states are determined by the quantisation condition
>
> $$\oint p\,dq = nh, \tag{1.1}$$
>
> where $n \in \mathbb{N}$ and h is Planck's constant, following the Bohr–Sommerfeld rule. These states explain the discrete bands in infrared spectra.

This definition of molecular vibration captures a key conceptual advance over earlier, purely phenomenological interpretations: it treats internal motion as a concrete mechanical process between atomic centres, governed by dynamical laws and capable of being quantised. It explains the regularity and discrete-

ness observed in infrared spectra through the application of classical mechanics combined with early quantum rules. However, the definition remains limited in several respects. The vibrational motion is assumed to follow a prescribed, idealised trajectory—typically harmonic or weakly anharmonic—and the quantisation condition is imposed externally, rather than derived from a unified theory. Moreover, the approach applies cleanly only to simple diatomic systems and does not generalise easily to polyatomic molecules or account for mode coupling and energy redistribution. Thus, while the definition offers a tangible and computationally accessible model, it lacks generality and internal consistency in systems of higher complexity.

The major breakthrough in this field was initiated by Max Born and J. Robert Oppenheimer in their 1927 paper, *Zur Quantentheorie der Molekeln* (“On the Quantum Theory of Molecules”), in which they proposed that the full molecular Hamiltonian could be approximated in the following form (using their original notation):

$$H = H_0 + \kappa^4 H_1 \tag{1.2}$$

where κ is a small parameter proportional to the inverse square root of the nuclear and electronic mass ratio, H_0 represents the electronic Hamiltonian with fixed nuclei, and H_1 accounts for the nuclear kinetic energy and its coupling to the electronic motion [8].

By exploiting the smallness of κ in Equation (1.2), Born and Oppenheimer showed that the fast electronic degrees of freedom could be solved independently from the slow nuclear motion. In their 1927 theory, nuclear vibrations appear at second order, while lower-order terms vanish due to the existence of an equilibrium nuclear geometry where electronic energy is minimal. The essential consequence is that, to leading order, the fast-moving electrons adapt instantaneously to the slower motion of the atomic nuclei, which enables one to define a *potential energy surface* (PES)—the electronic energy as a function of frozen nuclear coordinates—and treat the nuclei as moving on this landscape

while obeying classical dynamics.

This separation of timescales and masses justifies the core assumption of *classical molecular dynamics* (MD): nuclei are approximated as classical particles evolving under forces derived from the PES. Once these forces are computed, one can directly apply Newton's laws, bypassing a full nuclear quantum treatment.

> **Definition 1.3** (Classical MD & molecular vibrations)**.** Molecular vibrations are the time-dependent, oscillatory motions of nuclei about equilibrium positions, simulated as classical trajectories driven by interatomic forces derived from the electronic potential energy surface. These vibrations are emergent phenomena of Newtonian dynamics on the PES, and can manifest well-defined frequencies via emergent quantisation.

It should be noted that Definition 1.3 remains an incomplete representation of the underlying physics, omitting effects such as nuclear quantum motion and zero-point energy.

Definition 1.3 summarises two essential advances: first, it supplies the equations of motion governing the atomic nuclei—treated classically as point masses interacting via interatomic forces—and second, it offers a framework for interpreting molecular vibrations as coherent, directional oscillations about equilibrium configurations. These vibrational motions occur at well-defined frequencies and along specific normal modes, which are dictated by the geometry and mass distribution of the molecule.

Had 19$^{\text{th}}$ century experimentalists with a strong theoretical inclination, such as Tyndall or Kirchhoff, been presented with such a framework, their natural question could have been: *How can this theory illuminate the structure behind vibrational spectra*? This question, though implicit in 19$^{\text{th}}$-century discussions of absorption and resonance, became a driving force for physicists and chemists in the past century. Their efforts—spanning from early infrared experiments

to the development of normal mode analysis—revealed deep connections between vibrational frequencies and molecular geometry, revolutionising the way chemists and molecular physicists interpret spectral data.

I hope to contribute to this enduring dialogue by giving new mathematical voice which may, in some small way, advance the understanding of molecular vibrations.

1.1 Main contributions of this thesis

This thesis revisits the problem of understanding molecular vibrations—not only through the conventional lenses of quantum eigenstates and linearised normal modes, but by reinterpreting early spectroscopists' intuition that vibrational spectra reflect internal molecular motions. Chapter 3 approaches this problem from a signal-processing perspective, treating spectral information as a signal to be decomposed. It introduces a step-by-step mathematical framework for identifying meaningful distributions centred around spectral peaks, without relying on prior physical or chemical assumptions. The techniques developed in this chapter are implemented in an open-source Python package, *Tihi* [1].

Chapters 4 and 5 extend this intuition into a dynamical framework, viewing molecular vibrations as quasi-periodic motions expressed either through their representation in Fourier space or directly in configurational space, the latter having an inherent Riemannian structure. These chapters employ tools from Riemannian geometry and harmonic analysis to show how vibrational modes arise from the geometric structure of molecular trajectories and their recurrent patterns over time. The theoretical foundations for these approaches are established in the early sections of each chapter, laying the groundwork for a geometric and frequency-resolved understanding of molecular motion.

The chapters are ordered to reflect the historical development of molecular vibration studies: beginning with empirical spectral analysis inspired by

19th-century observations, moving to frequency-domain isolation methods reminiscent of mid-20th-century normal mode theory, and culminating in a fully geometric framework rooted in modern differential geometry.

In addition to providing these perspectives, the thesis makes the following contributions:

- **Empirical spectral decomposition (Chapter 3):** Algorithms for denoising, baseline correction, peak detection, and decomposition of vibrational spectra without prior chemical or physical assumptions; implemented in the open-source Python package *Tihi* [1].
- **Fourier integrator framework (Chapter 4):** Development of Fourier-space integrators and mode-specific molecular dynamics algorithms, enabling frequency-band-targeted simulations while preserving the physical integrity of trajectories.
- **Geometric molecular dynamics (Chapter 5):** (i) Formulation of molecular trajectories as geodesics on manifolds defined by the Jacobi metric; (ii) analysis of quasi-periodicity via tubular trajectories and hyper-volume overlap; (iii) fat manifold construction to quantify vibrational mode coupling.

The Fourier-space integrators and mode-specific molecular dynamics algorithms developed in Chapter 4 provide the ability to isolate and simulate molecular trajectories within targeted frequency bands while preserving their physical integrity. These methods offer precise control over frequency resolution and avoid distorting anharmonic couplings, making them well suited for analysing the relationship between spectral features and underlying motions. Their main limitations stem from the requirement for finely sampled trajectories to resolve narrow frequency windows and the increased computational cost of maintaining accuracy when working in high-frequency bands.

The Riemannian framework presented in Chapter 5 enables a fully (and usefully) geometric, coordinate-free description of molecular motion. This method

supports the identification of quasi-periodic structure via tubular trajectory self-overlaps and the quantification of vibrational mode coupling through the construction of fat manifolds. These geometric algorithms excel in physical interpretability, robustness to coordinate choices, and their ability to reveal dynamical structure without predefined mode assumptions. However, they are sensitive to numerical errors in curvature estimation and can become computationally demanding for systems with many strongly coupled modes.

Building upon these perspectives, the central research questions guiding this thesis are as follows:

- Is it possible to extract a physically meaningful decomposition of molecular spectral data without prior knowledge of the underlying chemistry or physics? (Chapter 3)
- Can molecular trajectories be isolated within a specific frequency band without compromising their physical integrity? (Chapter 4)
- Can molecular trajectories be described usefully in geometrical terms? (Chapters 5.1 to 5.3)
- Can the internal vibrational motions of a molecule be understood as manifestations of quasi-periodicity? (Chapters 5.4 to 5.6)

These questions are revisited and answered in the concluding remarks.

CHAPTER 2

Theoretical Preliminaries

Nam si tu mihi ... eruditum hominem adduxeris, ... acrem et acutum in cogitando, ... si erit idem ... civitatis ... hospes, non multum ei ... proderunt ...: subacto mihi ingenio opus est, ut agro non semel arato, sed ... iterato, quo meliores fetus possit edere; subactio autem est usus, auditio, lectio, litterae.

MARCUS TULLIUS CICERO
De Oratore

If you were to bring me a person, however learned, however sharp and acute in thought... if this very person is a stranger to the land... then his brightness will not help his/her argument much. I need a mind that has been well-tilled—like a field, not ploughed only once, but worked again and again, so it may yield better fruits. And this cultivation consists of practice, listening, reading, and literary study.

MARCUS TULLIUS CICERO
On the Orator

The mathematical structures and physical principles introduced here—ranging from classical molecular dynamics to the description of vibrational motion—form the foundation for Chapters 4 and 5, upon which the more specialised treatments in these chapters are built.

In the spirit of the epigraph, this chapter serves not as a survey, but as a rigorous cultivation of the essential concepts required for what follows. While the results here are not revisited in detail later, they are assumed throughout the thesis. The aim is to equip the reader with a prepared and self-sufficient understanding of the theoretical landscape in which the main arguments are situated.

2.1 Classical molecular dynamics and symplectic integration

Viewing a molecule as an N-body system composed of interacting atoms, one can write the *internal* part of the Hamiltonian, $\mathcal{H}^{\text{int}} = T + U$ as

$$\mathcal{H}^{\text{int}} = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + \sum_{i<j}^{N} u^{(2)}(\mathbf{r}_i, \mathbf{r}_j) + \sum_{i<j<k}^{N} u^{(3)}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \cdots \tag{2.1}$$

where $\mathbf{p}_i$, $\mathbf{r}_i$, m_i are i-th atom's 3-momentum, 3-position and mass, respectively. The potential terms of Equation 2.1 can be grouped differently between the short range interactions and long range interactions as

$$U = \sum_{i<j}^{N} u(r_{ij} | r_{ij} < R_c) + U^{\text{long range}} \tag{2.2}$$

with some cut-off distance R_c, and the distance between the i-th and j-th atom $r_{ij} = \sqrt{(\mathbf{r}_i - \mathbf{r}_j) \cdot (\mathbf{r}_i - \mathbf{r}_j)}$. In this work, vector quantities are expressed in index notation, where atomic indices are represented by Latin letters and their corresponding coordinates by Greek letters. Rewriting Equation 2.1 in this notation yields the following equivalent expression for the Hamiltonian:

$$\mathcal{H}^{\text{int}} = \frac{1}{2m_i} p_{i\mu} p^{i\mu} + u^{(2)}_{ij\mu\nu}(r_{i\mu}, r_{j\nu}) H(j-i) + u^{(3)}_{ijk\mu\nu\xi}(r_{i\mu}, r_{j\nu}, r_{k\xi}) H(j-i) H(k-j) + \cdots . \tag{2.3}$$

Here, repeated Greek indices indicate summation over the corresponding components, following the Einstein summation convention. In this chapter, such conventions are *not* applied to Latin indices. The function $H(x)$ is the Heaviside step function, which takes the value 1 when $x > 0$ and 0 otherwise. The inner product between vectors is represented by a contraction of upper and lower indices, where summation occurs over the repeated indices in a manner consistent with standard index manipulation.

As an example, consider a neutral carbon dioxide molecule in vacuum without any external interaction. A simple description of the molecule can take

the following form of Hamiltonian.

$$\begin{aligned}\mathcal{H}^{\text{int}}_{\text{CO}_2} =& \frac{p_{O_1\mu}p^{O_1\mu}}{2m_O} + \frac{p_{O_2\mu}p^{O_2\mu}}{2m_O} + \frac{p_{C\mu}p^{C\mu}}{2m} \\ &+ \frac{k_{CO}}{2}(\|r_{O_1\mu} - r_{C\mu}\| - r_{OC})^2 + \frac{k_{CO}}{2}(\|r_{O_2\mu} - r_{C\mu}\| - r_{OC})^2 \\ &+ \frac{k^{\text{angle}}_{OCO}}{2}(\theta - \theta_{OCO})^2\end{aligned} \tag{2.4}$$

the indices O_1, O_2 and C were used to specify that the terms are relating to these atoms–oxygen, one carbon and the other carbon atoms, respectively. The *norm* of a quantity $a_{i\mu}$ used in Equation 2.4 is defined as

$$\|a_{i\mu}\| = \sqrt{a_{i\mu}a^{i\mu}}, \tag{2.5}$$

and the angle formed by the bond made the three atoms, θ, can be obtained by simple trigonometric equation below.

$$\theta = \cos^{-1}\left(\frac{(r_{O\mu} - r_{C_1\mu}) \cdot \left(r^{O\mu} - r^{C_2\mu}\right)}{\|r_{O\mu} - r_{C_1\mu}\|\|r^{O\mu} - r^{C_2\mu}\|}\right) \tag{2.6}$$

Note that the indices inside a function, e.g. square-root or inverse cosine, are summed prior to the function operation. The bond and valence angle force constants k_{CO} and k^{angle}_{OCO} to be used in this section are $2017.9\,\text{kcal/mol/}^2$ and $118.817\,\text{kcal/mol/rad}^2$ [9]. The equilibrium valence angle, θ_0 of carbon dioxide is π radians.

Hamilton's equations of motion for some quantity corresponding to a set of indices I is given as

$$\begin{aligned}\frac{\partial\mathcal{H}}{\partial r_I} &= -\dot{p}_I \\ \frac{\partial\mathcal{H}}{\partial p_I} &= \dot{r}_I\end{aligned} \tag{2.7}$$

where $\dot{a}$ for some physical quantity a stands for its time-derivative. Using this formula, one can then obtain the following set of equations of carbon dioxide

motion.

$$\begin{aligned}
\dot{r}_{i\mu} &= \frac{p_{i\mu}}{m_i} \\
\dot{p}_{C_i\mu} &= k_{OC}(\|r_{O\mu} - r_{C_i\mu}\| - r_{OC_i})\frac{r_{O\mu} - r_{C_i\mu}}{2\|r_{O\mu} - r_{C_i\mu}\|} + k^{\text{angle}}_{OCO}(\theta - \theta_0)\frac{\partial\theta}{\partial r_{C_i\mu}} \\
-\dot{p}_{O\mu} &= \sum_{i=C_1,C_2}(\|r_{O\mu} - r_{C_i\mu}\| - r_{OC_i})\frac{k_{OC}(r_{O\mu} - r_{C_i\mu})}{2\|r_{O\mu} - r_{C_i\mu}\|} + k^{\text{angle}}_{OCO}(\theta - \theta_0)\frac{\partial\theta}{\partial r_{O\mu}}
\end{aligned} \tag{2.8}$$

The small angle approximation allows one to rewrite θ as

$$\theta \approx \sqrt{2(1 - \cos\theta)} \tag{2.9}$$

and if one write $A_\mu = r_{O\mu} - r_{C_1\mu}$, $B_\mu = r_{O\mu} - r_{C_2\mu}$, $A = \sqrt{A_\mu A^\mu}$, and $B = \sqrt{B_\mu B^\mu}$, the derivatives of θ in Equation 2.8 become

$$\begin{aligned}
\frac{\partial\theta}{\partial r_{O\mu}} &= \frac{1}{\theta AB}\left(A_\nu B^\nu\left(\frac{A_\mu}{A} + \frac{B_\mu}{B}\right) - B_\mu - A_\mu\right) \\
\frac{\partial\theta}{\partial r_{C_1\mu}} &= \frac{1}{\theta AB}\left(B_\mu - A_\nu B^\nu\frac{A^\kappa}{A}\delta_{\kappa\mu}\right) \\
\frac{\partial\theta}{\partial r_{C_2\mu}} &= \frac{1}{\theta AB}\left(A_\mu - A_\nu B^\nu\frac{B^\kappa}{B}\delta_{\kappa\mu}\right)
\end{aligned} \tag{2.10}$$

Although the carbon dioxide molecule is relatively simple to simulate, solving the equations of motion (Equation 2.8) analytically presents significant challenges. A straightforward numerical approach to solving this system is to approximate the values at each time step using an iterative method. Specifically, if the time derivative of a quantity a_μ is known at ℓ-th step, i.e. $a_{\mu;\ell}$ is known, the approximation for a_μ at the $\ell + 1$-th time step, denoted $a_{\mu;\ell+1}$, is obtained by advancing the ℓ-th time-step value according to the following expression:

$$a_{\mu;\ell+1} \approx a_{\mu;\ell} + \Delta t\dot{a}_{\mu;\ell} \tag{2.11}$$

where Δt is the time step. This method, called the Euler's method of (numerical) integration, can be written algorithmically as in Algorithm 1.

Figures 2.1 and 2.2 show the evolution of bond lengths, bond angles, total energy, and temperature for a carbon dioxide molecule in vacuum, modelled

Algorithm 1 Simple iterative time-step approximation for the quantity a_μ: Euler's method of numerical integration

1: **Input:** Initial value $a_{\mu;0}$, time step Δt, and initial time derivative $\dot{a}_{\mu;0}$
2: **Output:** Approximate values $a_{\mu;\ell}$ at each time step
3: **for** $\ell = 0$ to N **do** ▷ Iterate over time steps
4: Compute $a_{\mu;\ell}$ using the previous time-step approximation

$$a_{\mu;\ell+1} = a_{\mu;\ell} + \Delta t\, \dot{a}_{\mu;\ell}$$

5: **end for**

by the Hamiltonian given in Equation (2.4), using the spring constants specified above. Two distinct time steps were used: the simulation in Figure 2.1 was performed with $\Delta t = 10^{-6}\,\mathrm{fs}$, while that in Figure 2.2 used $\Delta t = 1\,\mathrm{fs}$. The initial positions are defined to be in its geometric equilibrium state:

$$r^{O_1} = (-1.160\,, 0, 0), \quad r^{C} = (0, 0, 0), \quad r^{O_2} = (1.160\,, 0, 0). \tag{2.12}$$

The initial velocities were drawn from the Maxwell-Boltzmann distribution at $300\,\mathrm{K}$, i.e. for each atom i, the velocity vector components were sampled from a normal distribution with zero mean and variance

$${\sigma_i}^2 = \frac{k_B T}{m_i} \tag{2.13}$$

where k_B is the Boltzmann constant, and T is the target temperature, and m_i is the mass of atom i. The resulting initial phase space configurations thus reflect thermal fluctuations around the equilibrium structure at room temperature $T = 300\mathrm{K}$.

Inspection of Figure 2.1 reveals that the bond lengths oscillate around their equilibrium value of 1.160, the O-C-O bond angle fluctuates around 180° (or π radians), as expected for a linear configuration. The total energyremains well conserved over the duration of the simulation (up to $2\,\mathrm{fs}$ in the figure), consistent with the physics of the closed system. The instantaneous temperature,

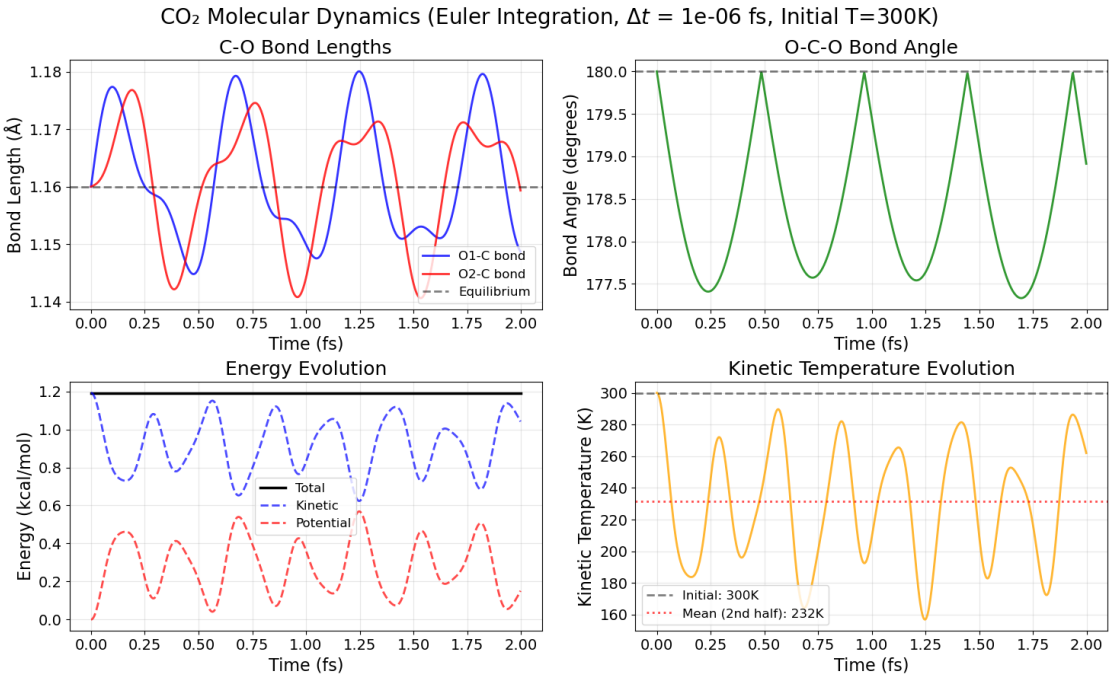


Figure 2.1. *Evolution of bond angles, bond lengths, energy, and temperature over time of carbon dioxide simulated via Algorithm 1 with a very small time-step of* $10^{-6}\,\mathrm{fs}$. *The simulation was run for* $2\,\mathrm{fs}$.

however, is observed to decrease from its initial value, and fluctuates around $232\,\mathrm{K}$. This behaviour arises because a portion of the initial kinetic energy is converted into potential energy as the system evolves. This is expected, as the initial configuration was chosen such that the potential energy at $t = 0$ is effectively zero. The initial velocities, sampled from the Boltzmann distribution, perturb this configuration and drive the system into regions of nonzero potential energy. As a result, some of the initial kinetic energy is inevitably transferred to potential energy, causing the instantaneous temperature to fluctuate around a value lower than the initial one.

The time-step of $\Delta t = 10^{-6}\,\mathrm{fs}$ is not realistic for systems larger or more complex than harmonic carbon dioxide model shown in Equation (2.4)–for $2\,\mathrm{fs}$ simulation, 2×10^6 number of computational steps are required. With the

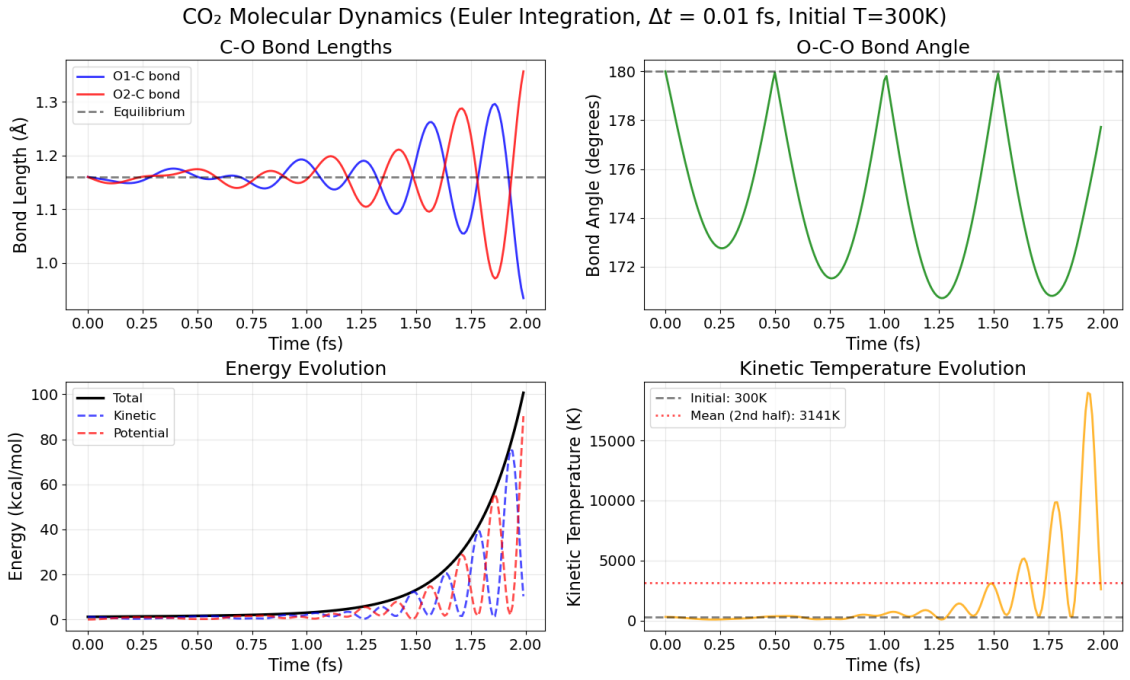


Figure 2.2. *Evolution of bond angles, bond lengths, energy, and temperature over time of carbon dioxide simulated via Algorithm 1 with a realistic time-step of* $1\,\mathrm{fs}$. *The simulation was run for* $2\,\mathrm{fs}$.

Euler integration algorithm, a more practical time step of $\Delta t = 0.01\,\mathrm{fs}$ leads to non-physical results, as shown in Figure 2.2. The bond lengths, total energy, and temperature diverge over time due to the accumulation of numerical error inherent in Euler's method. Since the same physical system yields stable and realistic dynamics when integrated using a smaller time step, the observed divergence can be attributed to the limitations of the integration method itself.

The limitations observed with non-symplectic methods, such as energy drift and numerical instability under practical time steps, underscore the need for integrators that maintain both stability and fidelity to the underlying physics over long simulation times. In particular, a desirable integrator should conserve key geometric structures of Hamiltonian dynamics, such as phase-space volume and the symplectic form, while accommodating reasonably large time

steps Δt. This motivates the use of *symplectic integration methods*, which are specifically designed to preserve the qualitative behaviour of Hamiltonian flows.

2.1.1 Symplectic maps & integrators

Prior to presenting the details of the integration scheme, the foundational concepts of symplectic maps (also known as symplectic transformations) and symplectic integration are outlined, as they form the theoretical basis for the subsequent analysis.

Definition 2.1 (Symplectic map)**.** Let $\Gamma = \mathbb{R}^d \times \mathbb{R}^d$ be the phase space of a given problem composed of the position and momentum coordinates $(r, p) \in \Gamma$, and inner-product over Γ, ω, defined as

$$\omega(a_\mu, b_\mu) = a_\kappa J^{\kappa\nu} b_\nu \tag{2.14}$$

for all $a, b \in \Gamma$, and

$$J = \begin{pmatrix} 0 & -I_d \\ I_d & 0 \end{pmatrix} \tag{2.15}$$

with a d-dimensional identity matrix I_d. A d-dimensional linear map $A : \Gamma \to \Gamma$ is *symplectic* if,

$$\omega(A_{\mu\nu} a^\nu, A_{\mu\nu} b^\nu) = \omega(a_\mu, b_\mu) \tag{2.16}$$

ω in Definition 2.1 is called a *symplectic form*.

If $d = 1$, i.e. in a 1-dimensional physical problem with a 2-dimensional phase-space, Definition 2.1 implies that symplectic mappings preserve the area of parallelograms. Similarly, for general d, the hypervolume of a $2d$-dimensional parallelepiped in phase space remains invariant under symplectic transformations [10].

For deterministic dynamics governed by Hamilton's equations of motion, Equation (2.7), is inherently symplectic. For a phase space element of an n-

dimensional system, the dynamical state can be represented as a single vector:

$$z = \begin{pmatrix} r \\ p \end{pmatrix} \in \mathbb{R}^{2n}, \tag{2.17}$$

and Hamilton's equations take the compact form:

$$\dot{z} = J\nabla_z H(z), \quad \text{with } J = \begin{pmatrix} 0 & -I_n \\ I_n & 0 \end{pmatrix}. \tag{2.18}$$

This representation highlights the structure-preserving nature of the dynamics. The solution to this system defines a flow map

$$\Phi_t : z(0) \mapsto z(t), \tag{2.19}$$

which advances the initial conditions through time along the Hamiltonian flow. In the case of a linear Hamiltonian system, or a system locally approximated by a linearised Hamiltonian (e.g., via Taylor expansion near a point), the flow map takes the form:

$$\Phi_t = e^{tJ\nabla^2 H} \tag{2.20}$$

where $\nabla^2 H$ is the Hessian matrix of the Hamiltonian.

To verify that this flow is symplectic in the sense of Definition 2.1, it suffices to show that

$$\Phi_t^T J \Phi_t = J. \tag{2.21}$$

This follows from the fact that the generator $A = J\nabla^2 H$ satisfies the identity $A^T J + JA = 0$, which implies that the matrix exponential e^{tA} preserves the symplectic form. Therefore the linearised flow Φ_t is symplectic.

In 1838, Joseph Liouville showed that the Hamiltonian vector field $J\nabla H$ generates canonical transformations that preserve both the symplectic form and phase space volume [11]. This result, known as *Liouville's theorem*, establishes that deterministic classical dynamics are symplectic in nature. Consequently, numerical integration schemes that preserve this structure—namely,

symplectic integrators—are particularly well suited for the accurate long-term simulation of Hamiltonian systems.

Symplectic integrators are constructed to mimic the geometric properties of exact Hamiltonian flows, particularly the preservation of the symplectic form ω, as in Equation (2.14), and, consequently, the conservation of phase-space volume [12,13]. Unlike general-purpose methods such as Euler or Runge–Kutta, which may introduce artificial dissipation or energy drift over long time scales, symplectic integrators maintain the qualitative features of the dynamics—such as bounded energy oscillations, conservation of invariants, and stability of long-term trajectories [14–16]. This makes them especially suitable for systems in applications such as molecular dynamics .

Mathematically, symplectic integrator can be defined as below.

Definition 2.2 (Symplectic integrator)**.** Let $\Gamma = \mathbb{R}^d \times \mathbb{R}^d$ be the phase space of a Hamiltonian system with canonical coordinates $(r, p) \in \Gamma$, and let the system evolve according to Hamilton's equations. A numerical integration scheme that defines a discrete-time map $\Phi_{\Delta t} : \Gamma \to \Gamma$, advancing the system from time t to $t + \Delta t$, is called a *symplectic integrator* if the map $\Phi_{\Delta t}$ is symplectic in the sense of Definition 2.1. That is,

$$\omega(D\Phi_{\Delta t}(a), D\Phi_{\Delta t}(b)) = \omega(a, b) \tag{2.22}$$

for all tangent vectors $a, b \in T\Gamma$, or equivalently,

$$(D\Phi_{\Delta t})^\top J D\Phi_{\Delta t} = J, \tag{2.23}$$

where J is the standard symplectic matrix defined in Equation (2.15), and $D\Phi_{\Delta t}$ is the Jacobian of the map $\Phi_{\Delta t}$.

2.1.2 Symplectic integration via Liouville operators

While traditional derivations of numerical integrators often rely on Taylor series expansions, a more elegant approach emerges from the perspective of Liouville operators. Before introducing this formalism, it is necessary to establish the concept of the Poisson bracket, which underlies the fundamental structure of classical mechanics.

For any two phase space functions f and g, their Poisson bracket is defined as:

$$\{f, g\} = \frac{\partial f}{\partial r_\mu}\frac{\partial g}{\partial p_\mu} - \frac{\partial f}{\partial p_\mu}\frac{\partial g}{\partial r_\mu} \tag{2.24}$$

where r_μ and p_μ are conjugate position and momentum variables, meaning they are the canonical coordinates of the physical system that satisfy Hamilton's equations of motion: Equation 2.7. The Poisson bracket provides a natural way to express the time evolution of any phase space function f through Hamilton's equations:

$$\frac{df}{dt} = \{f, \mathcal{H}\} \tag{2.25}$$

where $\mathcal{H}$ is the Hamiltonian of the system.

The solution to this differential equation defines a one-parameter family of mappings $\phi_t : \Gamma \to \Gamma$ called the *Hamiltonian flow*, which maps any initial phase space point $(r_\mu(0), p_\mu(0))$ to its evolved state $(r_\mu(t), p_\mu(t))$ at time t:

$$\phi_t(r_\mu(0), p_\mu(0)) = (r_\mu(t), p_\mu(t)) \tag{2.26}$$

By construction, this flow preserves the Hamiltonian and satisfies the group property $\phi_t \circ \phi_s = \phi_{t+s}$ for all $t, s \in \mathbb{R}$. The flow Φ_t introduced in the previous section is also a Hamiltonian flow [10].

However, solving Hamilton's equations analytically for complex systems is often not feasible, and numerical methods are required. This is where integrators come into play. Specifically, integrators allow us to approximate the continuous Hamiltonian flow ϕ_t over discrete time steps, enabling the simulation of long-term dynamics.

This formalism leads naturally to the definition of the Liouville operator, which provides a powerful and computationally efficient framework for constructing second-order symplectic integrators, such as the velocity Verlet method.

Definition 2.3 (Liouville Operator)**.** In a phase space composed of position variables $\mathbf{r} = (r_1, r_2, \cdots, r_{3N})$ and momentum variables $\mathbf{p} = (p_1, p_2, \cdots, p_{3N})$, the *Liouville operator* of the problem defined by the Hamiltonian $\mathcal{H}$ is

$$\hat{L}[\cdot] = \{\cdot, \mathcal{H}\} = \frac{\partial \mathcal{H}}{\partial p_\mu}\frac{\partial}{\partial r_\mu} - \frac{\partial \mathcal{H}}{\partial r_\mu}\frac{\partial}{\partial p_\mu} \tag{2.27}$$

Note that from the Hamilton's equations of motion, Equation (2.27) becomes:

$$\left(\hat{L}f\right)^\nu = \dot{r}_\mu \frac{\partial f^\nu}{\partial r^\mu} + \dot{p}_\mu \frac{\partial f^\nu}{\partial p^\mu}, \tag{2.28}$$

for some vector-valued function f. The square bracket is omitted here, following the usual convention.

If the function $f = P(t)$ is a tuple of momentum and position coordinates, at time t, the Hamilton's equations of motion gives

$$\dot{P}(t) = \hat{L}P(t). \tag{2.29}$$

This equation can be solved directly as

$$P(t) = e^{\hat{L}t}P(0). \tag{2.30}$$

To solve Equation (2.29) efficiently in numerical simulations, it is desirable to apply position and momentum updates separately. This can be achieved by decomposing the Liouville operator, $\hat{L}$, as

$$\hat{L} = \hat{L}_r + \hat{L}_p \tag{2.31}$$

where $\hat{L}_r = \dot{r}_\mu \frac{\partial}{\partial r^\mu}$ and $\hat{L}_p = \dot{p}\frac{\partial}{\partial p^\mu}$. It should be noted, however, that in general

$$e^{\hat{L}_r + \hat{L}_p} \neq e^{\hat{L}_r} e^{\hat{L}_p}; \tag{2.32}$$

the exponential of the sum is not equal to the product of the exponentials due to the non-commutativity of the operators.

While the operators $\hat{L}_r$ and $\hat{L}_p$ generally do not commute, an approximate factorisation of the time evolution operator is still possible. This is formalised by a result known as *Trotter's product formula* (or *Trotter's theorem*), introduced by Hale F. Trotter in a 1959 publication [17], which provides a way to approximate the exponential of a sum of operators by a product of exponentials.

> **Theorem 2.4** (Trotter's product formula)**.** *Let A and B be linear operators such that the exponentials e^{At} and e^{Bt} are well-defined for $t \geqslant 0$. Then for any $t \geqslant 0$,*
>
> $$e^{(A+B)t} = \lim_{n\to\infty}\left(e^{At/n}e^{Bt/n}\right)^n. \tag{2.33}$$

Equation (2.33) can also be written as

$$e^{(A+B)t} = \lim_{n\to\infty}\left(e^{At/2n}e^{Bt/n}e^{At/2n}\right)^n \approx \left(e^{At/2n}e^{Bt/n}e^{At/2n}\right)^n + O\left(t^3/n^2\right) \tag{2.34}$$

where the truncation assumes that n is large [18]. Using Trotter's theorem, Equation (2.30) can be rewritten as

$$P(t) \approx \left(e^{\hat{L}_p\Delta t/2}e^{\hat{L}_r\Delta t}e^{\hat{L}_p\Delta t/2}\right)^n + O(n\Delta t). \tag{2.35}$$

with $\Delta t = t/n$. Taylor expansion of the exponentials of $\hat{L}_r$ and $\hat{L}_p$ shows:

$$e^{\hat{L}_r\Delta t}P(t) = \left[1 + \hat{L}_r\Delta t + \frac{1}{2}\left(\hat{L}_r\Delta t\right)^2 + \cdots\right]P(t) \tag{2.36}$$

$$= \left[1 + \dot{r}(t)\Delta t + \frac{1}{2}(\dot{r}(t)\Delta t)^2 + \cdots\right]P(t) \tag{2.37}$$

$$\implies e^{\hat{L}_r\Delta t}P(t) = [r_1(t) + \dot{r}_1(t)\Delta t, \cdots, r_k(t) + \dot{r}_k(t)\Delta t, p_1(t), \cdots, p_k(t)], \tag{2.38}$$

and similarly,

$$e^{\hat{L}_p\Delta t/2}P(t) = \left[, r_1(t), \cdots, r_k(t), p_1(t) + \tfrac{1}{2}\dot{p}_1(t)\Delta t, \cdots, p_k(t) + \tfrac{1}{2}\dot{p}_k(t)\Delta t\right] \tag{2.39}$$

where k is the number of degrees of freedom. These observations show that the symmetric Trotter splitting in Equation (2.35) leads to an integration scheme

Algorithm 2 Velocity Verlet integration using symmetric Trotter splitting

1: **Input:** Initial values $r_{\mu;0}$, $p_{\mu;0}$; time step Δt; force function $F_\mu(p, r)$; velocity $v_\mu = \frac{\partial \mathcal{H}}{\partial p^\mu}$
2: **Output:** Approximate phase space trajectory $(p_{\mu;\ell}, r_{\mu;\ell})$ for $\ell = 0, \ldots, N$
3: **for** $\ell = 0$ to $N - 1$ **do** ▷ Iterate over time steps
4: Half-step momentum update: ▷ a *kick* step

$$p_{\mu;\ell+\frac{1}{2}} = p_{\mu;\ell} + \tfrac{1}{2}\Delta t\, F_\mu(p_{;\ell}, r_{;\ell})$$

5: Full-step position update: ▷ a *drift* step

$$r_{\mu;\ell+1} = r_{\mu;\ell} + \Delta t\, v_\mu(p_{;\ell+\frac{1}{2}})$$

6: Final half-step momentum update: ▷ a *kick* step

$$p_{\mu;\ell+1} = p_{\mu;\ell+\frac{1}{2}} + \tfrac{1}{2}\Delta t\, F_\mu(p_{\ell+\frac{1}{2}}, r_{\ell+1})$$

7: **end for**

composed of n steps. In each step, momenta and positions are updated as follows:

$$p_\mu\big(\tau + \tfrac{1}{2}\Delta t\big) = p_\mu(\tau) + \tfrac{1}{2}F_\mu(p(\tau), r(\tau))\Delta t, \tag{2.40}$$

$$r_\mu(\tau + \Delta t) = r_\mu(\tau) + v_\mu(\tau + \tfrac{1}{2}\Delta t)\Delta t, \tag{2.41}$$

$$p_\mu(\tau + \Delta t) = p_\mu\big(\tau + \tfrac{1}{2}\Delta t\big) + \tfrac{1}{2}F_\mu(p(\tau + \tfrac{1}{2}\Delta t), r(\tau + \Delta t))\Delta t, \tag{2.42}$$

where $v_\mu = \partial\mathcal{H}/\partial p^\mu$ denotes the velocity, $F_\mu = -\partial\mathcal{H}/\partial r^\mu$, and τ is the dummy variable for time.

In Algorithm 2, the subscript $;\ell$ was used to indicate the iteration step number.

Theorem 2.5. *Velocity Verlet method is a symplectic integrator*.

Proof. Velocity Verlet integrator of time-step Δt, $V^{\Delta t}$, can be decomposed as the following:

$$V^{\Delta t} = V_{\text{kick}} V_{\text{drift}} V_{\text{kick}}, \tag{2.43}$$

which gives $DV^{\Delta t} = DV_{\text{kick}} DV_{\text{drift}} DV_{\text{kick}}$. From Equation (2.40), it becomes immediate that

$$DV_{\text{kick}} = \begin{pmatrix} I & 0 \\ \frac{1}{2}\Delta t \frac{\partial F_\mu}{\partial r_\nu} & I \end{pmatrix}, \quad DV_{\text{drift}} = \begin{pmatrix} I & \Delta t \frac{\partial v_\mu}{\partial p_\nu} \\ 0 & I \end{pmatrix}. \tag{2.44}$$

Thus,

$$\left(DV^{\Delta t}\right)^T J DV^{\Delta t} = J \tag{2.45}$$

follows immediately. □

Theorem 2.6. *Let $\Phi_{\Delta t} : \Gamma \to \Gamma$ be a symplectic integrator on the phase space $\Gamma = \mathbb{R}^{2d}$, with Jacobian matrix $D\Phi_{\Delta t}$. Then, $\Phi_{\Delta t}$ preserves local phase-space volume; that is,*

$$\det D\Phi_{\Delta t} = 1. \tag{2.46}$$

Proof. Since $\Phi_{\Delta t}$ is a symplectic map, by Definition 2.1, it satisfies

$$\left(D\Phi_{\Delta t}\right)^T J D\Phi_{\Delta t} = J. \tag{2.47}$$

Taking determinant on both sides, one gets:

$$\det\Big(\left(D\Phi_{\Delta t}\right)^T J D\Phi_{\Delta t}\Big) = \det\left(D\Phi_{\Delta t}\right)^T \det J \det D\Phi_{\Delta t} = 1 = \det J, \tag{2.48}$$

which implies

$$\det D\Phi_{\Delta t} = \pm 1. \tag{2.49}$$

Since, $\Phi_{\Delta t}$ is a continuous deformation of the identity map, and the determinant of the identify is $+1$, the continuity implies that $\det D\Phi_{\Delta t} = 1$. Therefore, the map preserves local volume in phase space. □

Theorem 2.5 and 2.6 thus implies that the velocity Verlet integrator preserves local phase-space volume along the numerical trajectory as

$$\det DV_{\text{kick}} = \det DV_{\text{drift}} = 1. \tag{2.50}$$

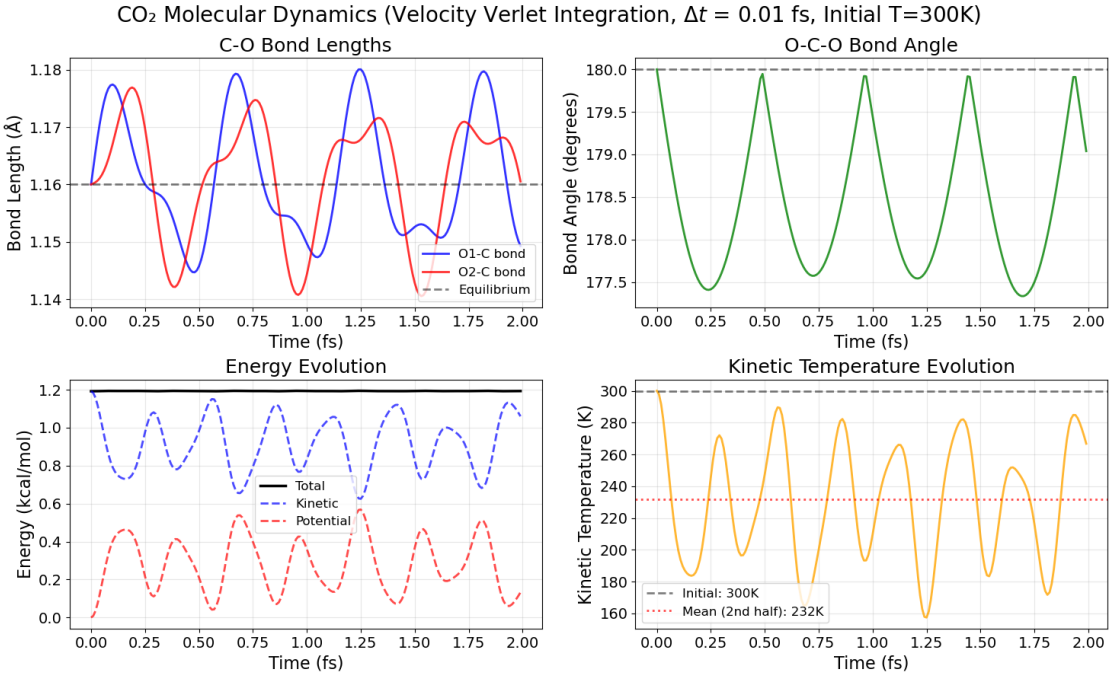


Figure 2.3. *Evolution of bond lengths, bond angles, energy and temperature over time of carbon dioxide simulated via Algorithm 2. The simulation had the time-step of* $0.01\,\mathrm{fs}$ *and was run for* $2\,\mathrm{fs}$ *in total.*

This volume preservation reflects the symplectic nature of the method and ensures that the evolution remains consistent with the incompressibility of Hamiltonian flows [14]. As a result, the integrator avoids any dissipation or contraction of phase-space structures, contributing to its excellent long-term stability and qualitative accuracy.

Figure 2.3 shows the evolution of bond lengths, bond angles, total energy, and temperature for a carbon dioxide molecule, simulated using the same system parameters and initial velocity sampling scheme as in Figures 2.1 and 2.2. In this case, the integration was performed using the velocity Verlet method with a time step of $0.01\,\mathrm{fs}$.

In contrast to the results shown in Figure 2.2 with the same time step, which exhibits numerical instability and rapidly diverging energy and temper-

ature, the velocity Verlet scheme maintains stable and physically consistent trajectories throughout the simulation window. The bond lengths and bond angle oscillate smoothly about their equilibrium values, reflecting the expected internal vibrational modes. The total energy remains conserved, with kinetic and potential energies exchanging harmonically as expected for a system governed by quadratic potentials. The temperature oscillates in a bounded and periodic fashion, with a stable mean value lower than $300\,\mathrm{K}$ resulting from the initial redistribution of kinetic energy.

One may argue that the Euler method is too simplistic for a fair comparison; however, even when compared to other non-symplectic integrators, symplectic methods demonstrate clear advantages for Hamiltonian systems. This superiority stems from their inherent ability to preserve phase-space volume and other geometric invariants of the underlying continuous dynamics. In doing so, symplectic integrators maintain the long-term stability of energy and prevent the cumulative error typically observed with non-symplectic schemes [19].

2.1.3 Thermostats

Figures 2.3 and 2.1 demonstrate that, in the absence of explicit temperature control, energy conservation inherently leads to temperature fluctuations during the dynamics. This arises because changes in potential energy are exactly compensated by changes in kinetic energy, in accordance with the conservation of total energy. Consequently, any drift or oscillation in potential energy directly affects the instantaneous temperature, which is determined from the kinetic energy of the system [20, 21].

In many practical simulations, it is necessary to maintain the temperature close to a specified target. This is particularly important in:

1. **Canonical ensemble (NVT) simulations**, where temperature must remain fixed by definition [22–24];

2. **Biomolecular simulations**, to mimic physiological conditions [25];
3. **Materials simulations under thermal stress**, where temperature gradients or thermostats are applied to control the heat flow [26, 27];
4. **Equilibration stages**, where temperature needs to be rapidly and stably brought to a desired value before production runs [28, 29].

In such cases, thermostatting mechanisms are employed to regulate the kinetic energy and stabilise the temperature around a target value. In an NVT ensemble, the resulting energy fluctuations can also facilitate exploration of the accessible phase space, helping the system traverse different regions of its configurational landscape.

Definition 2.7 (Thermostat in general dynamical systems)**.** Let Γ be a smooth finite-dimensional phase space equipped with canonical coordinates $(r, p) \in \Gamma$, and let $H : \Gamma \to \mathbb{R}$ be a smooth Hamiltonian. A *thermostatted dynamical system* is a system of differential equations on an extended phase space $\tilde{\Gamma} = \Gamma \times \mathbb{R}^m$ of the form:

$$\frac{d}{dt}\tilde{x}(t) = \tilde{X}(\tilde{x}(t)), \quad \tilde{x} \in \tilde{\Gamma} \tag{2.51}$$

where:

1. $\tilde{X} : \tilde{\Gamma} \to T\tilde{\Gamma}$ is a smooth (possibly stochastic) vector field, where $T\tilde{\Gamma}$ denotes the tangent bundle of $\tilde{\Gamma}$, i.e. the union of all tangent spaces at each point of $\tilde{\Gamma}$,
2. the projection $\pi : \tilde{\Gamma} \to \Gamma$ defines the physical trajectory $x(t) = \pi\tilde{x}(t)$, and
3. the dynamics are constructed that a chosen observable (e.g. temperature) is regulated dynamically by the evolution of the auxiliary variables in $\mathbb{R}^m$.

A thermostatted system is called *deterministic* if $\tilde{X}$ is a smooth vector field, and *stochastic* if $\tilde{X}$ includes stochastic noise terms.

Building on Definition 2.7, two commonly used thermostats: Nosé-Hoover and Langevin thermostats are defined and demonstrated in the next paragraphs.

2.1.3.1 Nosé-Hoover thermostat

The Nosé-Hoover thermostat is a deterministic method for simulating systems at constant temperature, designed to generate dynamics consistent with the canonical ensemble. The thermostat modifies the equations of motion by coupling the physical system to an auxiliary variable that acts as a dynamical friction coefficient. This coupling introduces a feedback mechanism that regulates the system's kinetic energy, thereby maintaining an average temperature close to a prescribed target.

Due to its deterministic nature, the Nosé-Hoover thermostat preserves the time-reversibility and generates smooth, continuous trajectories, making it particularly useful for studies where deterministic structure and trajectory continuity are essential [30].

Formally, the thermostat can be defined as in Definition 2.8.

Definition 2.8 (Nosé-Hoover thermostat)**.** Let $\Gamma \cong \mathbb{R}^{2n}$ be the phase space of a system with Hamiltonian $H(r,p) = \sum_{i=1}^{n} \frac{{p_i}^2}{2m_i} + V(r)$. The *Nosé-Hoover thermostat* at temperature T defines an extended phase space:

$$\tilde{\Gamma}_{NH} = \Gamma \times \mathbb{R}, \quad \tilde{x} = (r, p, \zeta) \tag{2.52}$$

where $\zeta \in \mathbb{R}$ is an auxiliary variable representing a dynamical friction coefficient. The *Nosé-Hoover vector field* $\tilde{X}_{NH} : \tilde{\Gamma}_{NH} \to T\tilde{\Gamma}_{NH}$, $\tilde{x} \mapsto \dot{\tilde{x}}$ is defined by the following:

$$\begin{cases} \dot{q}_i &= \frac{p_i}{m_i} \\ \dot{p}_i &= -\frac{\partial V}{\partial r_i} - \zeta p_i \\ \dot{\zeta} &= \frac{1}{Q}\left(\sum_{n}^{i=1} \frac{{p_i}^2}{m_i} - n k_B T\right) \end{cases} \tag{2.53}$$

where $Q > 0$ is a thermostat mass parameter, n is the number of degrees of freedom, and k_B is Boltzmann's constant.

The Nosé-Hoover thermostat is a deterministic dynamical system that extends classical Hamiltonian dynamics to include a feedback mechanism regulating the temperature. It operates on an extended phase space $\tilde{\Gamma}_{NH} = \{\tilde{X}\}$. This coupling introduces an effecitve non-Hamiltonian flow in the physical subspace Γ, but the total system remains volume-preserving (with proper measure) in the extended space [31, 32].

Proposition 2.9. *Let the Nosé–Hoover equations of motion be given by Equation* (2.53). *Then the flow generated by these equations does not preserve the standard volume in the extended phase space, but it preserves a modified volume measure given by*

$$e^{n\int_0^t \zeta(s)\,ds}\, dr_1 \cdots dr_n\, dp_1 \cdots dp_n\, d\zeta. \tag{2.54}$$

Proof. Define the state vector of the Nosé-Hoover system as

$$x = \begin{pmatrix} r_1 \\ \vdots \\ r_n \\ p_1 \\ \vdots \\ p_n \\ \zeta \end{pmatrix} \in \mathbb{R}^{2n+1}. \tag{2.55}$$

Let $F(x) = \dot{x}$ denote the right-hand side of the system shown in Equation (2.53), and let $J^x = \partial F/\partial x$ be the Jacobian of the flow map.

Consider the time-dependent Jacobian matrix

$$M(t) = \frac{\partial x(t)}{\partial x(0)}, \tag{2.56}$$

then the evolution of the determinant $\det M(t)$ satisfies the following identity:

$$\frac{d}{dt}\det M(t) = (\operatorname{Tr} J^x(t)) \cdot \det M(t). \tag{2.57}$$

From Equation (2.53), one obtains $\partial \dot{r}_i/\partial r_i = 0$, $\partial \dot{p}_i/\partial p_i = -\zeta$, and $\partial \dot{\zeta}/\partial \zeta = 0$. This gives $\mathrm{Tr}\, J^x(t) = -n\zeta$. With this information, one can rewrtie Equation (2.57) as

$$\frac{d}{dt} \log \det M(t) = \mathrm{Tr}\, J^x(t) = -n\zeta(t). \tag{2.58}$$

Integrating both sides, one gets

$$\log \det M(t) = -n \int_0^t \zeta(s) ds. \tag{2.59}$$

Now, consider the modified flow variable:

$$\tilde{M}(t) = e^{n \int_0^t \zeta(s) ds} M(t), \tag{2.60}$$

with the determinant

$$\det \tilde{M}(t) = e^{n \int_0^t \zeta(s) ds} \cdot \det M(t). \tag{2.61}$$

This then gives

$$\frac{d}{dt} \log \det \tilde{M}(t) = n\zeta(t) + \frac{d}{dt} \log \det M(t) = n\zeta(t) - n\zeta(t) = 0. \tag{2.62}$$

As the logarithmic derivative of $\det \tilde{M}(t)$ vanishes, $\det \tilde{M(t)}$ is constant.

Evaluating at $t = 0$ yields $\tilde{M}(0) = M(0) = I$, and therefore $\det \tilde{M}(t) = 1$ for all $t \geqslant 0$. This shows that although the standard Jacobian determinant $\det M(t)$ is not constant, the rescaled flow $\tilde{M}(t)$ evolves in such a way that the combination

$$e^{n \int_0^t \zeta(s)\, ds} \cdot \det M(t) \tag{2.63}$$

is conserved. This implies that the Nosé–Hoover flow preserves a *modified volume measure*

$$d\mu = e^{n \int_0^t \zeta(s)\, ds}\, dr_1 \cdots, dr_n, dp_1, \cdots, dp_n\, d\zeta, \tag{2.64}$$

which evolves under the flow such that its total measure is invariant in time.

Therefore, the flow of the Nosé–Hoover system is volume-preserving in the extended phase space with respect to this time-dependent measure, completing the proof.

□

Proposition 2.9 established that the Nosé–Hoover flow preserves a modified volume form in the extended phase space [33, 34]. In physical terms, this implies that the total *weight* of a collection of nearby initial conditions remains constant over time when evolved by the dynamics. Such volume preservation is a generalization of Liouville's theorem.

In molecular dynamics and statistical mechanics, the central object of interest is the equilibrium distribution—in particular, the canonical Gibbs measure, which describes the probability density of states at thermal equilibrium with fixed temperature T. A physically meaningful thermostat must not only preserve the volume but also leave this distribution invariant under the dynamics.

The appropriate framework to analyse the evolution of probability densities under a deterministic flow $\dot{x} = F(x)$ is the Liouville equation:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho F) = 0, \tag{2.65}$$

which describes the conservation of probability (with density ρ) in phase space. Equation (2.65) ensures that the probability density flows along the trajectories of the system. A time-independent (stationary) solution to Equation (2.65) satisfies:

$$\nabla \cdot (\rho F) = 0, \tag{2.66}$$

representing the probability density invariance under the flow. In the case of the Nosé-Hoover thermostat, it is natural to ask whether the canonical Gibbs measure is invariant in this sense [35, 36]. Proposition 2.10 confirms this property for the use of Nosé-Hoover dynamics in canonical ensemble simulations.

Proposition 2.10 (Invariance of the Canonical Gibbs Measure)**.** *Assume that the potential V is defined such that the partition function is finite. Define*

$$\rho_T(q, p, \zeta) = \frac{1}{Z_\beta} e^{-\left(H(q,p) + \frac{Q}{2}\zeta^2\right)\frac{1}{k_B T}}, \qquad H(q,p) = \sum_{i=1}^{n} \frac{p_i^2}{2m_i} + V(q), \tag{2.67}$$

with $Q > 0$. Then ρ_T is an invariant density for the Nosé–Hoover flow: if $\rho(t, \cdot)$

> *solves the Liouville equation* (*Equation* (2.65)) *with initial condition* $\rho(0,\cdot) = \rho_\beta$, *then* $\rho(t,\cdot) \equiv \rho_\beta$ *for all* t.

Proof. It suffices to show that $\nabla\cdot(\rho_\beta F) = 0$. Given the following

$$\nabla\cdot(\rho_T F) = (\nabla\rho_T)\cdot F + \rho_T\,(\nabla\cdot F) = \rho_T\big(\nabla\log\rho_T\cdot F + \nabla\cdot F\big), \tag{2.68}$$

from simple calculations, one can observe

$$\nabla\cdot F = -n\zeta, \quad \nabla\log\,\rho_T\cdot F = n\zeta. \tag{2.69}$$

This result gives $\nabla\cdot(\rho_T F) = 0$, and this implies that ρ_T is stationary for the Liouville equation.

Equivalently, for any smooth compactly supported function φ,

$$\frac{d}{dt}\int\varphi\,\rho_T = \int\frac{\partial\varphi}{\partial t}\,\rho_T + \int\varphi\,\frac{\partial\rho_T}{\partial t} = \int\varphi\,\big(-\nabla\cdot(\rho_T F)\big) = 0, \tag{2.70}$$

which shows the measure $\rho_T(q,p,\zeta)\,dq\,dp\,d\zeta$ is invariant under the flow. □

To implement the Nosé-Hoover thermostat in practice, Equation (2.53) must be integrated numerically using an appropriate integrator. Algorithm 3 describes a velocity Verlet-type scheme adapted to include the thermostat variable ζ, providing a time-reversible and second-order accurate method for simulating canonical ensemble dynamics. As done in Algorithm 2, subsctipt $;\ell$ is used to indicate the iteration step number.

Figures 2.4 and 2.5 show the bond lengths, bond angle, energy, and temperature evolution of a carbon dioxide system in vacuum with initial temperature of $300\,\mathrm{K}$ and time step size of $0.01\,\mathrm{fs}$. The bond lengths and bond angles from the two figures are showing the expected behaviours. The bond angle is fluctuating around $180°$, and the bond lengths are periodically getting stretched and contracted around the expected length of 1.160. The thermostat mass parameter Q used to produce the figures is computed as

$$Q = nk_BT \approx 5.3649\,\mathrm{kcal/mol} \tag{2.71}$$

Algorithm 3 Velocity Verlet algorithm with Nosé-Hoover thermostat

Input: Initial positions r_μ, momenta p_μ, dynamic friction variable ζ, timestep Δt, target temperature T, thermostat mass Q, total number of steps N. Number of degrees of freedom: n.
Output: Approximate phase space trajectory $(p_{\mu;\ell}, r_{\mu;\ell})$ for $\ell = 0, \ldots, N$
Compute initial forces $F_\mu = -\partial_\mu V(r)$
for $\ell = 0$ to $N-1$ **do** ▷ Iterate over time steps
    Half-step update of momenta:

$$p_{\mu;\ell+\frac{1}{2}} = p_{\mu;\ell} + \frac{\Delta t}{2}(F_\mu - \zeta_{;\ell} p_{\mu;\ell})$$

    Full-step update of positions: ▷ $v_{\mu;\ell}$ is the velocity obtained by $p_{\mu;\ell+\frac{1}{2}}$

$$r_{\mu;\ell+1} = r_{\mu;\ell} + \Delta t \cdot v_{\mu;\ell}$$

    Compute new forces $F_{\mu;\ell} = -\partial_\mu V(r_{;\ell+1})$
    Full-step update of thermostat variable: ▷ $m(\mu)$: corresponding mass

$$\zeta_{;\ell+1} = \zeta_{;\ell} + \frac{\Delta t}{Q}\left(m(\mu)\sqrt{v^{\mu;\ell} v_{\mu;\ell}} - n k_B T\right)$$

    Half-step update of momenta:

$$p_{\mu;\ell+1} = p_{\mu;\ell+\frac{1}{2}} + \frac{\Delta t}{2}\left(F_\mu - \zeta_{;\ell}\, p_{\mu;\ell+\frac{1}{2}}\right)$$

end for

where $n = 9$ is the degree of freedom of a carbon dioxide, $k_B = 1.987 \times 10^{-3}\,\mathrm{kcal}/(\mathrm{mol}\cdot\mathrm{K})$, and $T = 300\,\mathrm{K}$.

The key differences from the system without a thermostat are evident in the energy and temperature evolution plots. In Figure 2.4, the energy appears to diverge monotonically. However, Figure 2.5 reveals that the initially increasing trend eventually reverses, returning toward the initial energy level and repeating this behaviour periodically Throughout the simulation, the average temperature is maintained at $300,\mathrm{K}$, as intended. Notably, the peaks of these

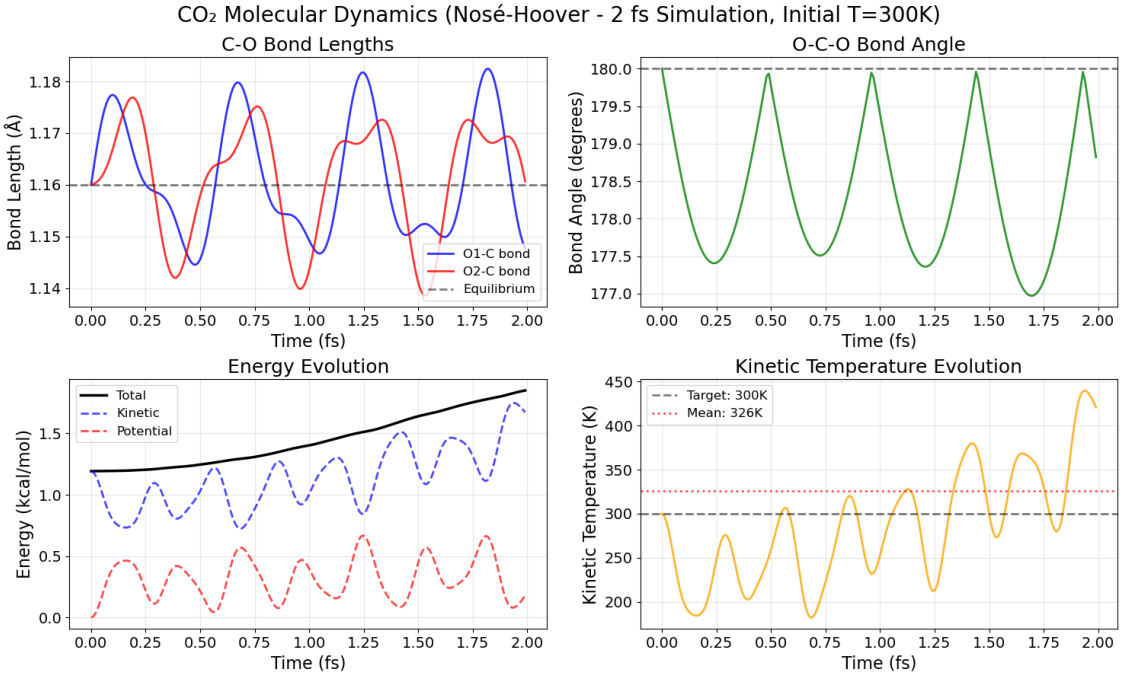


Figure 2.4. *Evolution of bond lengths, bond angles, energy and temperature over time of carbon dioxide simulated via Algorithm 3. The simulation had the time-step of* $0.01\,\mathrm{fs}$ *and was run for* $2\,\mathrm{fs}$ *in total. The initial temperature was set to* $300\,\mathrm{K}$*, and the target temperature was also set to* $300\,\mathrm{K}$*.*

periodic oscillations grow taller over time. This gradual increase is attributed to the accumulation of numerical errors at each time step.

The observed violation of energy conservation is expected, as the thermostat maintains the target temperature by adjusting the system's kinetic energy via the Nosé–Hoover equations, Equation (2.53), which introduce an additional degree of freedom to mimic the effect of a thermal reservoir.

2.1.3.2 Langevin thermostat

While the Nosé–Hoover thermostat provides a deterministic and time-reversible method for regulating temperature, it suffers from important limitations in practice. Most notably, the Nosé–Hoover dynamics are not guaranteed to be

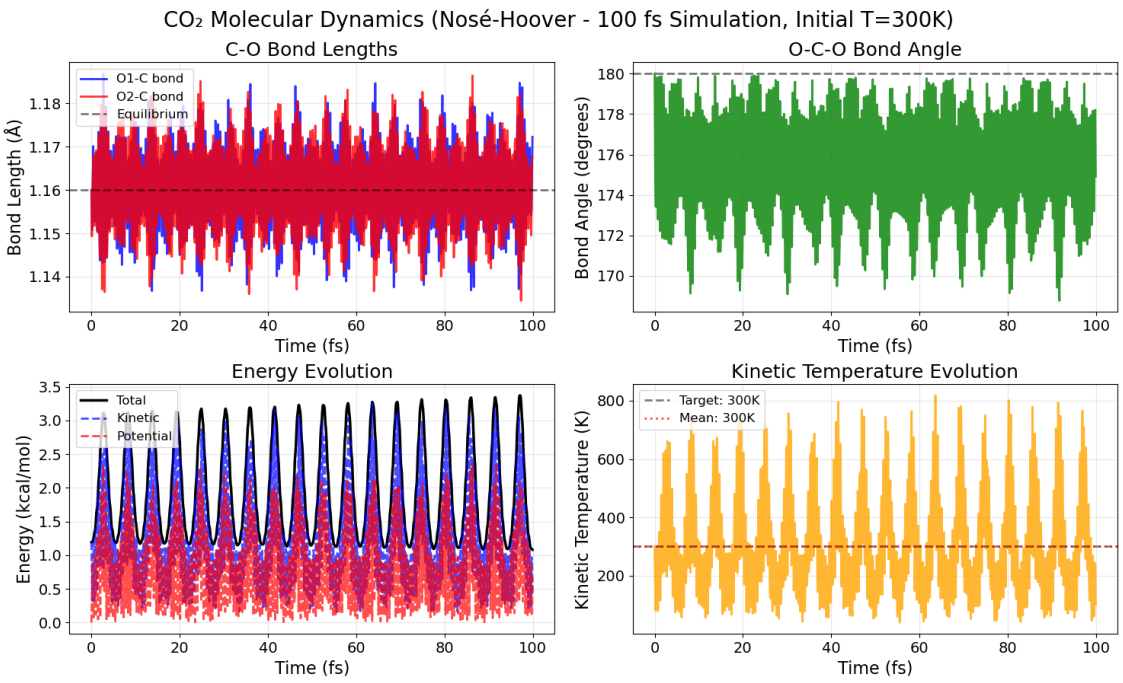


Figure 2.5. *Evolution of bond lengths, bond angles, energy and temperature over time of carbon dioxide simulated via Algorithm 3. The simulation had the time-step of* $0.01\,\mathrm{fs}$ *and was run for* $100\,\mathrm{fs}$ *in total. The initial temperature was set to* $300\,\mathrm{K}$*, and the target temperature was also set to* $300\,\mathrm{K}$.

ergodic, particularly in systems with few degrees of freedom or near-integrable behaviour [37]. In such cases, the system may fail to explore the full phase space, resulting in incorrect sampling of the canonical ensemble despite the formal invariance of the Gibbs measure. Moreover, the coupling mechanism through a single friction variable ζ may not be sufficient to induce sufficient mixing, especially in systems with weakly chaotic dynamics.

An alternative is the *Langevin thermostat*, originally introduced to describe Brownian motion [38] and later adapted to molecular systems by Prigogine and Balescu [39]. It combines deterministic forces with linear friction and stochastic fluctuations to mimic the interaction of a system with an implicit thermal reservoir. Unlike deterministic thermostats, Langevin dynamics ensure

ergodicity under broad conditions and guarantee convergence to the canonical ensemble [40].

To rigorously characterise such dynamics, one models the state of the system as a time-dependent stochastic process

$$X_t = (r(t), p(t)) \in \Gamma \cong \mathbb{R}^{2n}, \tag{2.72}$$

satisfying a system of stochastic differential equations (SDEs) of the form

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \tag{2.73}$$

where $W_t \in \mathbb{R}^m$ is a vector of independent Wiener processes. For any twice continuously differentiable map $\varphi : \Gamma \to \mathbb{R}$, it is known that

$$d\varphi(X_t) = \nabla\varphi(X_t) \cdot b(X_t)dt + \tfrac{1}{2}\mathrm{Tr}\big[\sigma(X_t)\sigma(X_t)^T\nabla^2\varphi(X_t)\big]dt + \nabla\varphi(X_t) \cdot \sigma(X_t)dW_t. \tag{2.74}$$

This relation is known as *Itō's lemma*, named after Kiyoshi Itō. Equation (2.74) describes the infinitesimal evolution of observables along stochastic trajectories. This statement is crucial for determining whether the Langevin thermostat preserves the given probability distribution under the stochastic dynamics.

Formally, Langevin thermostat is defined as in Definition 2.11.

Definition 2.11. Let Γ be the phase space of a classical system with position-momentum coordinates $(r, p) \in \mathbb{R}^n \times \mathbb{R}^n$ with the Hamiltonian $H(r, p) = \sum_{i=1}^n \frac{{p_i}^2}{m_i} + V(r)$ for a smooth potential $V : \mathbb{R}^n \to \mathbb{R}$.

The *Langevin thermostat* at temperature $T > 0$ defines as stochastic process $(r(t), p(t)) \in \Gamma$ satisfying the system of stochastic differential equations

$$\begin{cases} dr_i &= \frac{p_i}{m_i}dt, \\ dp_i &= -\frac{\partial V}{\partial r_i}(r)dt - \gamma_i p_i dt + \sqrt{2\gamma_i m_i k_B T}dW_i(t) \end{cases}, \tag{2.75}$$

for $i = 1, \cdots, n$ where $\gamma_i > 0$ is the friction coefficient for particle i, $W_i(t)$ are independent standard Wiener processes, and k_B is Boltzmann's constant.

The random process obtained via Definition 2.11 is called *Langevin dynamics*.

Given that the momentum component of Equation (2.75) takes the form specified in Equation (2.73), one can apply Itō's lemma to Equation (2.75) and justify the invariance of the canonical measure under Langevin dynamics. This result is formally established in Proposition 2.12.

Proposition 2.12. *Let* $(r(t), p(t)) \in \mathbb{R}^{2n}$ *evolve according to Equation* (2.75) *in Definition 2.11. The Gibbs measure defined as*

$$\rho_T(r,p) = \frac{1}{Z_T} e^{-H(r,p)/k_B T}, \quad Z_T = \int_\Gamma e^{-H(r,p)/k_B T} dr dp, \tag{2.76}$$

is invariant under the Langevin dynamics.

Proof. Let $X_t = (r(t), p(t))$ be a solution of the Langevin SDEs with initial condition distributed as ρ_T, and let $\varphi \in C^\infty(\mathbb{R}^{2n})$ be a smooth observable. To show that ρ_T is invariant, it suffices to show that $\mathbb{E}[\varphi(X_t)]$ is time-independent.

Applying Itō's lemma to $\varphi(r(t), p(t))$, and noting that the noise terms have 0 expectation, the time derivative of the expected value is given by:

$$\frac{d}{dt}\mathbb{E}[\varphi(r(t), p(t))] = \mathbb{E}[\hat{\mathcal{L}}\varphi(r(t), p(t))] \tag{2.77}$$

where $\hat{\mathcal{L}}$ is the infinitesimal generator of the random process produced by the Langevin thermostat, acting on the observable φ by Equation (2.75), by:

$$\hat{\mathcal{L}}\varphi = \sum_{i=1}^{n} \left(\frac{p_i}{m_i} \frac{\partial \varphi}{\partial r_i} - \frac{\partial V}{\partial r_i} \frac{\partial \varphi}{\partial p_i} - \gamma_i p_i \frac{\partial \varphi}{\partial p_i} + \gamma_i m_i k_B T \frac{\partial^2 \varphi}{\partial t^2} \right). \tag{2.78}$$

By the initial condition $(r(0), p(0)) \sim \rho_T$,

$$\mathbb{E}[\varphi(r(t), p(t))] = \int_{\mathbb{R}^{2n}} \varphi(r,p) \mu_t(dr, dp) \tag{2.79}$$

where μ_t is the probability distribution of X_t. The quantity $\mu_t = \rho_T dr dp$ for all $t \geqslant 0$ by Equation (2.77) as:

$$\frac{d}{dt}\mathbb{E}[\varphi(r(t), p(t))] = \int \hat{\mathcal{L}}\varphi(r,p) \rho_T(r,p) dr dp. \tag{2.80}$$

Applying integration by parts to the right-hand side of Equation (2.80), it becomes evident that for all $\varphi \in C^\infty(\mathbb{R}^{2n})$, the following holds

$$\frac{d}{dt}\mathbb{E}[\varphi(r(t), p(t))] = 0, \tag{2.81}$$

as $\int \hat{\mathcal{L}}\varphi(r,p)\rho_T(r,p)drdp = 0$. Thus, for all $t \geqslant 0$ and X_t, ρ_T remains invariant. □

Remark. The validity of Proposition 2.12 crucially relies on the specific coefficient of the stochastic term in Equation (2.75). If this coefficient were not equal to $\sqrt{2\gamma_i m_i k_B T}$, the canonical Gibbs measure would, in general, no longer be invariant under the dynamics. This is because the balance between stochastic fluctuations and deterministic dissipation is necessary to ensure that the time derivative $\frac{d}{dt}\mathbb{E}[\varphi(r(t), p(t))]$ vanishes for all smooth observables φ.

The term $\sqrt{2\gamma_i m_i k_B T}$ determines the standard deviation of the Gaussian random variable sampled at each time step. Denoting this standard deviation by σ_i, the required identity is:

$$\sigma_i{}^2 = 2\gamma_i m_i k_B T. \tag{2.82}$$

This relation is known as the *fluctuation–dissipation relation* for the Langevin thermostat, and it ensures compatibility between thermal fluctuations and viscous damping in the Langevin thermostat.

While Proposition 2.12 establishes that the canonical Gibbs measure ρ_T is invariant under Langevin dynamics, this alone does not guarantee the time averages of observables will converge to ensemble averages. A system may preserve an equilibrium measure while failing to explore the entire phase space, a phenomenon that arises in deterministic thermostats such as Nosé–Hoover. In the Nosé–Hoover case, the dynamics may exhibit non-ergodic behaviour, especially in systems with few degrees of freedom or nearly integrable structures, leading to poor sampling of thermodynamic states.

In contrast, the Langevin thermostat includes stochastic forcing, which disrupts regular trajectories and drives the system to explore a much broader

portion of phase space. The presence of noise, balanced by dissipation through the fluctuation–dissipation relation, enables the system to overcome dynamical barriers and ensures robust mixing properties. As a result, the Langevin dynamics are ergodic under broad and physically realistic assumptions.

This ergodicity ensures that long-time averages of observables computed along individual trajectories converge almost surely to expectations with respect to the canonical measure. Thus, the Langevin thermostat not only preserves the desired equilibrium distribution but also provides a reliable mechanism for sampling it in practice. This is formalised in the following proposition.

Proposition 2.13. *Let $(r(t), p(t)) \in \mathbb{R}^{2n}$ evolve under the Langevin equations in Definition 2.11, with Hamiltonian $H(r,p) = \sum_{i=1}^{n} \frac{p_i^2}{2m_i} + V(r)$, where $V : \mathbb{R}^n \to \mathbb{R}$ is smooth and bounded from below, such that*

$$\lim_{\|r\| \to \infty} V(r) = \infty. \tag{2.83}$$

Assume that each friction coefficient $\gamma_i > 0$ and noise acts on all momentum coordinates. Then the Langevin process is ergodic with respect to the canonical measure

$$\rho_T(r,p) = \frac{1}{Z_T} \exp\left(-\frac{H(r,p)}{k_B T}\right), \tag{2.84}$$

in the sense that for any observable $\varphi \in L^1(\rho_T)$,

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T \varphi(r(t), p(t))\, dt = \int_{\mathbb{R}^{2n}} \varphi(r,p)\, \rho_T(r,p)\, dr\, dp \quad \textit{almost surely}. \tag{2.85}$$

The proof of Proposition 2.13 relies on several advanced mathematical tools that lie beyond the scope of this thesis. A concise version of the argument is presented below, with references to the relevant literature for further details.

Proof. Let $X_t = (r(t), p(t))$ denote the Langevin process. The stochastic forcing acts directly on the momentum variables p_i, while the deterministic drift couples p to r, ensuring sufficient mixing. Under mild regularity assumptions on V,

the associated generator satisfies Hörmander's condition, making the process *hypoelliptic*, strong Feller, and irreducible [41, 42]. These properties imply that the process admits at most one smooth invariant measure and explores the phase space fully. For the precise statements and proofs on this remark, see Section 3 of [41] or Chapter 6 of [42].

Take $V(r)$ such that $\lim_{\|r\|\to\infty} V(r) = \infty$. The total energy $H(r,p)$ grows at infinity and can serve as a Lyapunov function [43, 44]. Applying the Langevin generator $\hat{\mathcal{L}}$ to H yields

$$\hat{\mathcal{L}}H(r,p) = -\sum_{i=1}^{n} \gamma_i \frac{p_i^2}{m_i} + \sum_{i=1}^{n} \gamma_i m_i k_B T \frac{\partial^2 H}{\partial t^2}. \tag{2.86}$$

Thus, for sufficiently large $\|p\|$, the generator satisfies the drift condition:

$$\hat{\mathcal{L}}H \leqslant -\delta\|p\|^2 + C \tag{2.87}$$

for some constants $\delta, C > 0$, due to the boundedness of the Hamiltonian. This guarantees recurrence and tightness of trajectories in phase space [43, 45].

By the classical results of Sean P. Meyn and Richard L. Tweedie (cited as [45], see also [44]), the process is geometrically ergodic. Since the canonical measure ρ_T is invariant (Proposition 2.12) and unique, the ergodic theorem for Markov processes implies that for all $\varphi \in L^1(\rho_T)$, the following holds [46]:

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \varphi(r(t), p(t))\, dt = \int \varphi \, d\rho_T \quad \text{almost surely.} \tag{2.88}$$

Note that the uniqueness of the canonical measure ρ_T is immediate from the definition of Gibbs measure with known initial conditions. □

Proposition 2.12 establishes the canonical Gibbs measure as an invariant distribution of the Langevin dynamics, providing the statistical mechanical foundation for the thermostat. Proposition 2.13 further guarantees that the dynamics are ergodic, ensuring that, over sufficiently long simulation times, the system explores almost the entire phase space in accordance with the canonical ensemble.

Algorithm 4 Stochastic Velocity Verlet algorithm with Langevin thermostat

Input: Initial positions r_μ, momenta p_μ, timestep Δt, temperature T, friction coefficients γ_μ, particle masses $m(\mu)$, number of steps N
Output: Approximate phase space trajectory $(p_{\mu;\ell}, r_{\mu;\ell})$ for $\ell = 0, \ldots, N$
Compute initial forces $F_\mu = -\partial_\mu V(r)$
for $\ell = 0$ to $N-1$ **do** ▷ Iterate over time steps
    Half-step update of momenta with friction and noise:

$$p_{\mu;\ell+\frac{1}{2}} = p_{\mu;\ell} + \frac{\Delta t}{2} F_\mu - \frac{\Delta t}{2} \gamma_\mu p_{\mu;\ell} + \sqrt{\frac{1}{2}\gamma_\mu m(\mu) k_B T \Delta t} \cdot R_{\mu;\ell}$$

    where $R_{\mu;\ell} \sim \mathcal{N}(0,1)$ are independent standard Gaussian samples.
    Full-step update of positions using velocity:

$$r_{\mu;\ell+1} = r_{\mu;\ell} + \Delta t \cdot \frac{p_{\mu;\ell+\frac{1}{2}}}{m(\mu)}$$

    Compute new forces: $F_{\mu;\ell+1} = -\partial_\mu V(r_{\ell+1})$
    Half-step update of momenta with new force:

$$p_{\mu;\ell+1} = p_{\mu;\ell+\frac{1}{2}} + \frac{\Delta t}{2} F_{\mu;\ell+1} - \frac{\Delta t}{2} \gamma_\mu p_{\mu;\ell+\frac{1}{2}} + \sqrt{\frac{1}{2}\gamma_\mu m(\mu) k_B T \Delta t} \cdot \tilde{R}_{\mu;\ell} \tag{2.89}$$

    with fresh independent noise $\tilde{R}_{\mu;\ell} \sim \mathcal{N}(0,1)$.
end for

With both the invariant measure and ergodicity established, it remains essential to implement a numerical integration scheme that accurately captures the statistical properties of the Langevin dynamics. Algorithm 4 demonstrates how the Langevin thermostat can be combined with the velocity Verlet method to generate stochastic trajectories that sample from the canonical ensemble at a fixed temperature. Note that Einstein summation convention was not used in Algorithm 4.

It is important to note that, due to the presence of stochastic forcing, the Langevin thermostat requires careful consideration of both physical and numerical parameters. In particular, one must first ensure that the system

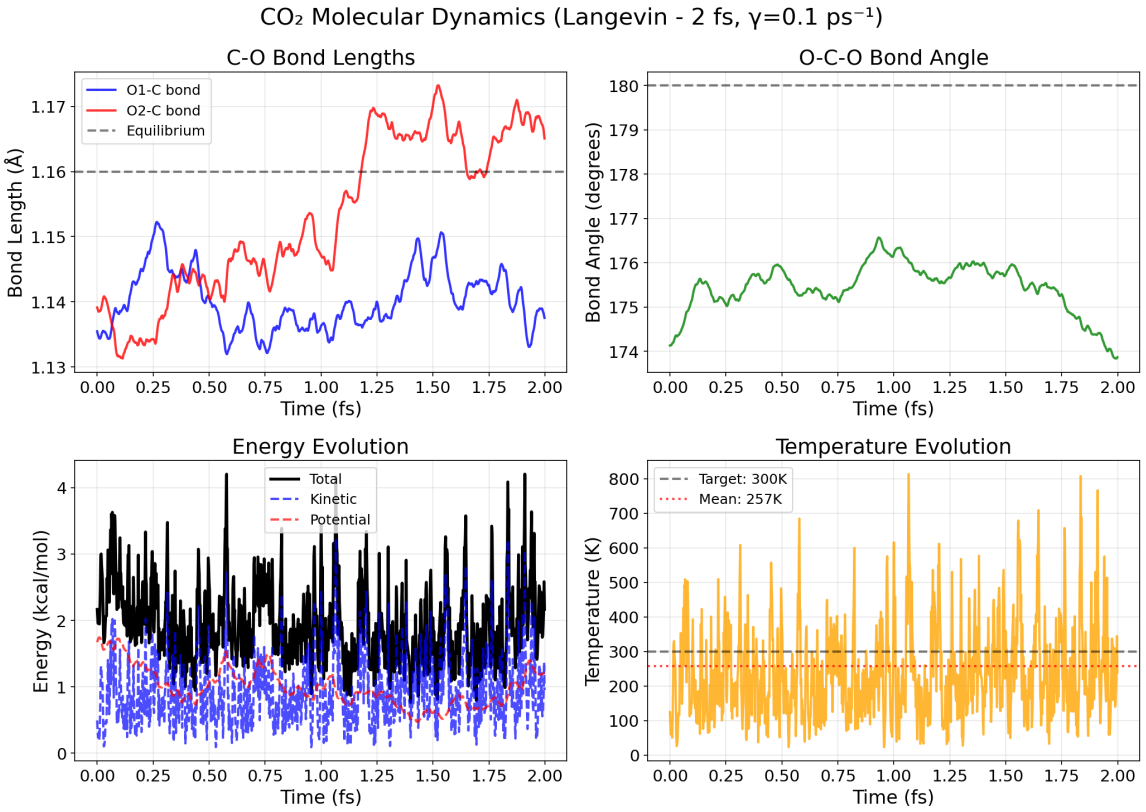


Figure 2.6. *Evolution of bond lengths, bond angles, energy and temperature over time of carbon dioxide simulated via Algorithm 4. The simulation had the time-step of* $0.001\,\mathrm{fs}$ *and was run for* $2\,\mathrm{fs}$ *in total. The initial and target temperatures were set to* $300\,\mathrm{K}$*, and the friction coefficient* γ *was set to* $0.1\,\mathrm{ps}^{-1}$*. The system was equilibrated for* $50\,\mathrm{ps}$ *prior to recording the data shown.*

is properly *equilibrated*, meaning that it has evolved from its initial conditions toward a statistical steady state representative of the target temperature. This typically requires a sufficient simulation time, particularly when the friction parameter γ is small. Additionally, the parameters γ and timestep Δt should satisfy $\gamma \Delta t \ll 1$.

The latter condition is essential for the numerical stability of the velocity Verlet integration scheme, especially when discretizing the stochastic differential equations. Since the stochastic term in Equation (2.89) involves a factor

of $\sqrt{\gamma \Delta t}$, even moderate values of $\gamma \Delta t$ can lead to disproportionately large noise contributions, potentially distorting the sampling if the timestep is not sufficiently small.

The time required for equilibration is expected to be *much longer* than the time step. The system can be considered equilibrated if macroscopic observables such as total energy, temperature, or pressure fluctuate around steady values over time, independent of the initial conditions. In practice, one typically discards an initial portion of the trajectory—referred to as the *burn-in period*—before collecting statistics.

Figures 2.6 and 2.7 show the system dynamics beyond an initial equilibration period of $50\,\mathrm{ps}$, simulated using Algorithm 4 with time steps of 2 and $100\,\mathrm{fs}$, respectively.

In Figure 2.6, the bond angles and bond lengths do not fluctuate around their target values, indicating that the system remains in the burn-in phase and has not yet reached equilibrium. Moreover, the mean temperature is approximately $257\,\mathrm{K}$, which deviates significantly from the target temperature of $300\,\mathrm{K}$.

By contrast, in the $100\,\mathrm{fs}$ simulation shown in Figure 2.7, the bond lengths and angles oscillate around their equilibrium values, and the system temperature fluctuates near $300\,\mathrm{K}$, with a mean of $297\,\mathrm{K}$. As Algorithm 4 is inherently stochastic, the mean temperature does not exactly match the target value.

Using a larger value of γ in the Langevin thermostat accelerates the approach of the system's kinetic temperature toward the target value. Figure 2.8 illustrates that, following equilibration, Langevin dynamics with $\gamma = 1.0, \mathrm{ps}^{-1}$ drives the system towards the target temperature more rapidly than with $\gamma = 0.1, \mathrm{ps}^{-1}$. However, this faster thermalisation does not imply that the system has reached equilibrium, as evidenced by deviations in bond lengths and bond angle distributions.

Comparing to the Nosé-Hoover thermostat, the Langevin thermostat's temperature management is not as strict due to its stochastic nature. However,

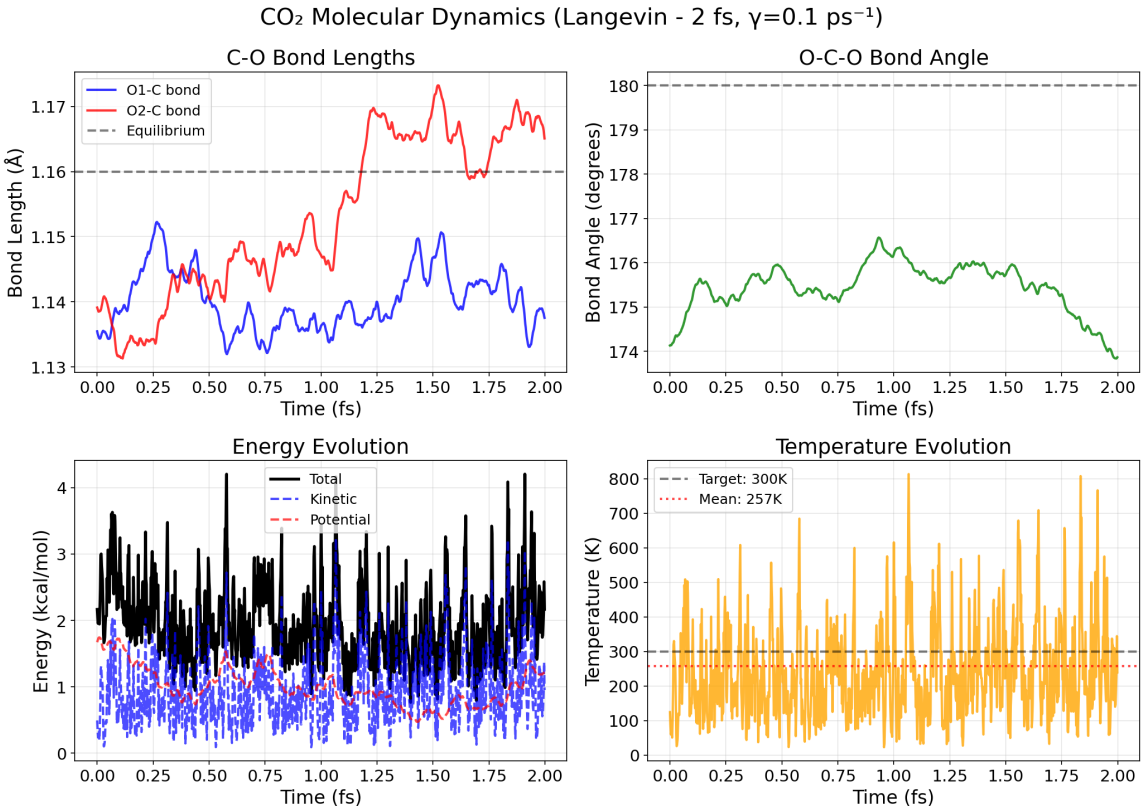


Figure 2.7. *Evolution of bond lengths, bond angles, energy and temperature over time of carbon dioxide simulated via Algorithm 4. The simulation had the time-step of* $0.001\,\mathrm{fs}$ *and was run for* $100\,\mathrm{fs}$ *in total. The initial and target temperatures were set to* $300\,\mathrm{K}$*, and the friction coefficient* γ *was set to* $0.1\,\mathrm{ps}^{-1}$*. The system was equilibrated for* $50\,\mathrm{ps}$ *prior to recording the data shown.*

as shown in Figure 2.7, the molecule explores a wider range of geometric configurations—as evident from the fluctuations in bond lengths and angles—as well as different regions of the potential energy surface. This suggests that the Langevin thermostat may be more suitable for generating physically realistic trajectories, especially when thorough sampling of the configuration space is desired.

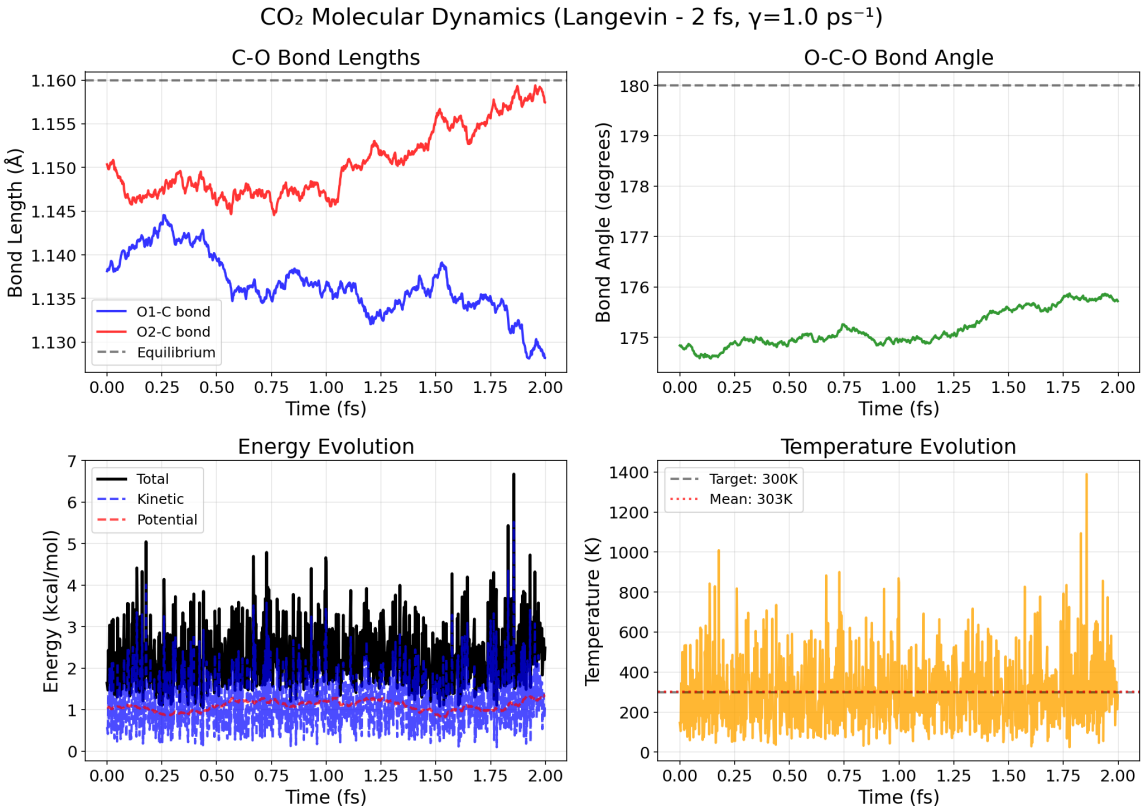


Figure 2.8. *Evolution of bond lengths, bond angles, energy and temperature over time of carbon dioxide simulated via Algorithm 4. The simulation had the time-step of* $0.001\,\mathrm{fs}$ *and was run for* $2\,\mathrm{fs}$ *in total. The initial and target temperatures were set to* $300\,\mathrm{K}$*, and the friction coefficient* γ *was set to* $1.0\,\mathrm{ps}^{-1}$*. The system was equilibrated for* $50\,\mathrm{ps}$ *prior to recording the data shown.*

2.2 Molecular vibrations

This thesis explores several mathematical frameworks to understand the vibrational behaviour of classical molecular systems. Molecular vibrations—periodic motions of atoms about their equilibrium positions—play a central role in determining a wide range of molecular properties, from thermal capacity to spectroscopic signatures. Despite being inherently quantum mechanical phenomena, much insight can be gained through classical approximations that reveal the structure of the vibrational modes and their interactions [47].

2.2.1 Classical description of molecular vibrations

Consider a classically described molecule with Hamiltonian given by

$$H(r,p) = \mathrm{KE}(p) + V(r), \tag{2.90}$$

where $\mathrm{KE}(p)$ denotes the kinetic energy of the system. Let the masses of the constituent particles be encoded in a mass matrix , $m^{\mu\nu}$, such that in matrix form:

$$m = [m^{\mu\nu}] = \begin{bmatrix} m_1 & 0 & 0 & & & & & \\ 0 & m_1 & 0 & & & & & \\ 0 & 0 & m_1 & & & & & \\ & & & \ddots & & & & \\ & & & & m_{n_{\mathrm{atoms}}} & 0 & 0 \\ & & & & 0 & m_{n_{\mathrm{atoms}}} & 0 \\ & & & & 0 & 0 & m_{n_{\mathrm{atoms}}} \end{bmatrix}. \tag{2.91}$$

Then the Hamiltonian can be expressed as

$$H(r,p) = \tfrac{1}{2}\dot{r}_\mu m^{\nu\mu}\dot{r}_\nu + V(r), \tag{2.92}$$

where the Einstein summation convention is assumed (i.e., repeated indices are implicitly summed over).

The global translational/rotational motions of a molecular system are not relevant to its internal vibrational dynamics. To isolate the vibrational degrees of freedom, one can transform the coordinate system such that 1) all positions are defined relative to the centre of mass, and 2) global rotational components are explicitly removed from the description [47].

After eliminating these external motions, the number of independent *vibrational degrees of freedom* for a nonlinear molecule composed of n_{atoms} atoms is given by $3n_{\mathrm{atoms}} - 6$. Here, the subtraction accounts for three degrees of freedom associated with translation and three with rotation. For linear

molecules, which lack one of the rotational degrees of freedom, the vibrational degrees of freedom are instead given by $3n_{\text{atoms}} - 5$.

Physically, vibrational modes correspond to the characteristic ways in which atoms in a molecule move relative to one another while preserving the overall molecular identity. These are not arbitrary jostlings but rather specific, repeatable internal motions determined by the structure and bonding of the molecule. One can picture a molecule as a collection of masses (atoms) connected by springs (chemical bonds); when the system is perturbed, it responds by oscillating in well-defined patterns. These include stretching motions, where atoms move back and forth along the line of the bond, and bending motions, where the angles between atoms open and close like hinges. Importantly, these modes involve only internal reconfigurations and exclude any global translation or rotation of the entire molecule.

As an example, for a carbon dioxide molecule, one can have bending, symmetric stretch, and antisymmetric stretch modes as illustrated in Figure 2.9. It is noted that only three vibrational modes are observed, since the bending mode may occur in two directions—within the molecular plane and orthogonal to it.

2.2.1.1 Small vibrations around an equilibrium configuration of a molecule

Suppose that a molecular configuration is initially given in its equilibrium state at a specified temperature. If the molecule is constrained to move only in the vicinity of this equilibrium configuration, the potential energy function can be approximated using a Taylor expansion about the equilibrium position:

$$V(r) = V_0 + \left.\frac{\partial V}{\partial r^\mu}\right|_0 r_\mu + \frac{1}{2}\left.\frac{\partial^2 V}{\partial r^\mu \partial r^\nu}\right|_0 r_\mu r_\nu + \cdots, \tag{2.93}$$

where the subscript 0 specifies that the corresponding values are computed at its equilibrium state. At equilibrium, the gradient of the potential vanishes,

$$\left.\frac{\partial V}{\partial r^\mu}\right|_0 = 0, \tag{2.94}$$

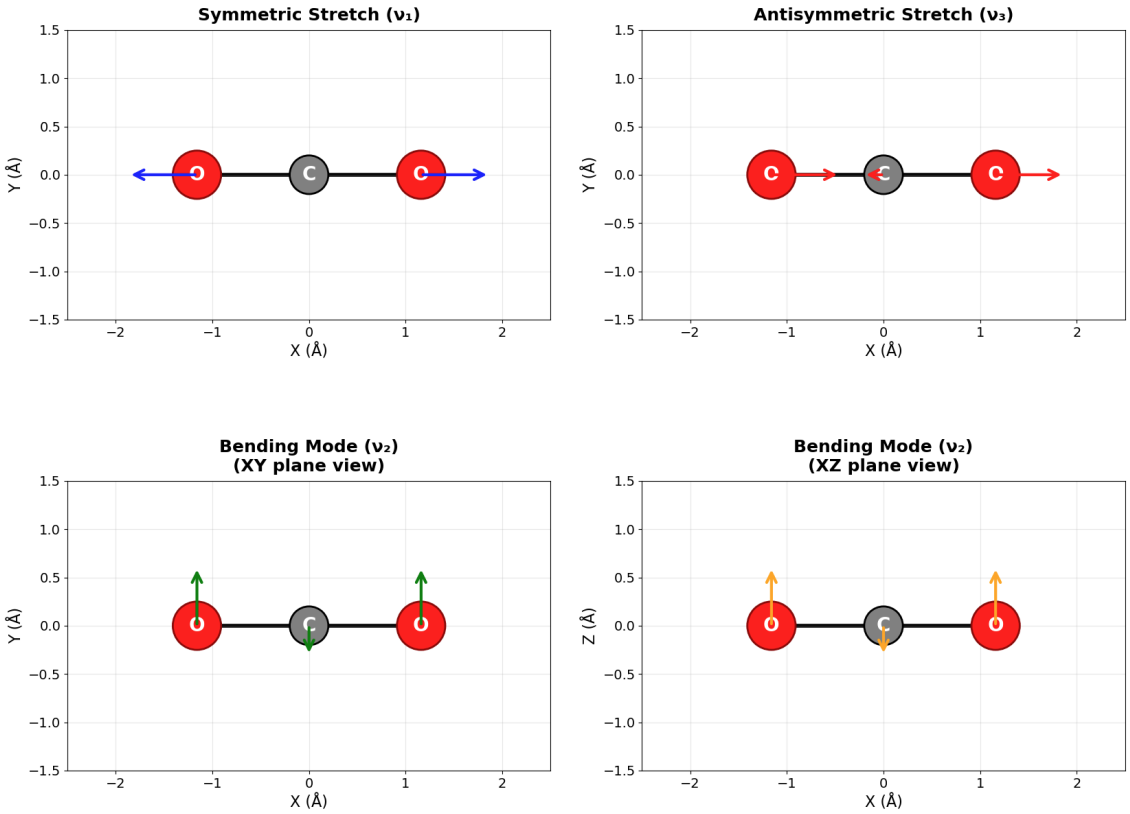


Figure 2.9. *Vibrational mode directions of carbon dioxide*

and one may, without loss of generality, choose the reference potential energy V_0 to be zero. The potential energy then simplifies to the harmonic approximation:

$$V(r) \approx \frac{1}{2} \left. \frac{\partial^2 V}{\partial r_\mu \partial r_\nu} \right|_0 r_\mu r_\nu. \tag{2.95}$$

The quadratic form captures the leading-order behaviour of the potential near equilibrium and underlies the theory of normal modes in molecular vibrations. The matrix $\frac{\partial^2 V}{\partial r^\mu \partial r^\nu}$ is called the Hessian of the potential. In this work, denoted as $\mathrm{Hess}_{\mu\nu}$.

As Newton's second law can be expressed in Lagrangian form as:

$$\frac{d}{dt}\left(\frac{\partial \mathrm{KE}}{\partial \dot{r}^\mu}\right) + \frac{\partial V}{\partial r^\mu} = 0, \tag{2.96}$$

it follows that, near the equilibrium configuration where the potential can be approximated harmonically, the equations of motion reduce to:

$$\ddot{r}_\mu + \mathrm{Hess}_{\mu\nu}\, r^\nu = 0, \tag{2.97}$$

This equation constitutes a system of $3n_{\mathrm{atoms}}$ coupled, second-order linear differential equations whose general solutions of the form:

$$r_\mu = A_\mu \cos\left(\sqrt{\lambda}t + \phi\right) \tag{2.98}$$

for constants A_μ, λ, and ϕ [47]. This demonstrates that the harmonic approximation of the potential naturally leads to oscillatory (vibrational) solutions, corresponding to the normal modes of the molecular system.

As an example, reconsider Equation (2.4), which may initially appear to be harmonic at first glance. However, it is not harmonic in the strict sense, as the angle θ is a nonlinear function of the atomic coordinates—it involves an inverse cosine of dot products of normalized vectors. Therefore, the potential energy term involving θ must also be expanded to second order in Cartesian displacements to obtain a true harmonic approximation.

The Hessian matrix can be interpreted as a sum of physical contributions from different types of internal interactions:

$$\mathrm{Hess}_{\mu\nu} = \sum_{\text{bonds}} \mathrm{Hess}_{\mu\nu}^{(\text{stretch})} + \sum_{\text{angles}} \mathrm{Hess}_{\mu\nu}^{(\text{bend})} + \cdots. \tag{2.99}$$

Each component $\mathrm{Hess}_{\mu\nu}^{(\cdot)}$ corresponds to the second-order derivative of the potential energy with respect to coordinates r^μ and r^ν, evaluated at the equilibrium configuration.

1. For a bond-stretching interaction between two atoms, the potential energy is typically given by
$$V_{\text{stretch}} = \frac{k_s}{2}\left(\sqrt{u^\alpha u_\alpha} - r_0\right)^2, \tag{2.100}$$
where $u^\alpha = {r_i}^\alpha - {r_j}^\alpha$ is the displacement vector between the bonded atoms indexed by i and j. The corresponding contribution to the Hessian tensor

is

$$\mathrm{Hess}_{\mu\nu}^{(\text{stretch})} = k_s \frac{\partial r_{ij}}{\partial r^\mu} \frac{\partial r_{ij}}{\partial r^\nu}\bigg|_0, \tag{2.101}$$

where $r_{ij} = \|u^\alpha\|$ denotes the scalar bond length.

2. For an angle-bending interaction involving three atoms, the potential energy is typically given by

$$V_{\text{bend}} = \frac{k_\theta}{2}(\theta - \theta_0)^2, \tag{2.102}$$

where θ is the angle between the vectors $u_1^\alpha = r_i^\alpha - r_j^\alpha$ and $u_2^\alpha = r_k^\alpha - r_j^\alpha$. Since θ is a nonlinear function of the positions, it must be expanded to second order around equilibrium. The harmonic approximation of its contribution to the Hessian is then given by

$$\mathrm{Hess}_{\mu\nu}^{(\text{bend})} = k_\theta \frac{\partial \theta}{\partial r^\mu} \frac{\partial \theta}{\partial r^\nu}\bigg|_0. \tag{2.103}$$

After summing over all such internal interactions, the total harmonic potential energy takes the quadratic form:

$$V(r) \approx \frac{1}{2} r^\mu \mathrm{Hess}_{\mu\nu} r^\nu, \tag{2.104}$$

where $r \in \mathbb{R}^{3n_{\text{atoms}}}$ denotes the displacement of each atomic coordinate from equilibrium. This reduces the equations of motion to the form given in Equation (2.97). Consequently, one can expect that the solution to the harmonically approximated carbon dioxide system, as described by the Hamiltonian in Equation (2.4), takes the form presented in Equation (2.98)—namely, a set of purely oscillatory solutions characteristic of vibrational motion.

The preceding analysis established that the solutions possess a vibrational form; however, two key questions remain unresolved:

1. How are the constants A^μ, λ and ϕ in Equation (2.98) determined?
2. In which directions are the atoms actually moving in the solution to Equation (2.97)?

These naturally leads to the next part of the discussion: normal mode analysis.

2.2.1.2 Normal mode analysis

By solving the secular equation

$$\det(\mathrm{Hess}_{\mu\nu} - \lambda\delta_{\mu\nu}) = 0 \tag{2.105}$$

one can find the eigenvalues of the Hessian of the potential in equilibrium. With these eigenvalues, Equation (2.98) transforms as below [48]:

$$\ddot{r}_\mu + \Lambda_{\mu\nu}\mathcal{D}^{\nu\alpha}r_\alpha = 0 \tag{2.106}$$

where $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_{3n_{\mathrm{atoms}}})$ is the diagonal matrix of the eigenvalues of Hess, and $\mathcal{D}$ is the matrix where the columns are eigenvectors of Hess, i.e.

$$\mathcal{D} = \begin{bmatrix} \hat{n}^{(1)} & \cdots & \hat{n}^{(3n_{\mathrm{atoms}})} \end{bmatrix} \tag{2.107}$$

where for all $i = 1, \cdots, 3n_{\mathrm{atoms}}$,

$$\lambda_i \hat{n}^{(i)} = \mathrm{Hess}\,\hat{n}^{(i)}. \tag{2.108}$$

Note that if λ_i is positive, it corresponds to a λ of a solution of the form given by Equation (2.98), vibrating in the direction $\hat{n}^{(i)}$ [47]. The amplitude and the phase of Equation (2.98) is given by the initial condition of the system. The eigenvector (or *eigenmode*) $\hat{n}^{(i)}$represents the direction of vibration and is referred to as the *vibrational mode* or the *normal mode*. The quantity $\sqrt{\lambda_i}$ corresponds to the frequency of this vibration, and is called the *vibrational frequency* associated with mode i.

If the λ_i is zero, the solution of Equation (2.106) becomes linear of the form:

$$r_\mu = W_\mu t + B_\mu \tag{2.109}$$

for some constant vectors W and B. This corresponds to the translational and rotational motions. For an isolated molecule six of the eigenvalues (five for linear molecules) are zero. Similarly, if $\lambda_i < 0$, the solution takes a form of

exponential growth or decay. These do not correspond to any of the vibrational modes, and are typically excluded from vibrational analysis.

The directions of Figure 2.9 shows the directions corresponding to each normal mode.

2.2.2 Vibrational density of states

While normal mode analysis provides detailed information about individual vibrational frequencies and their associated atomic motions, it is often useful—especially in large or disordered systems—to consider the collective distribution of these frequencies. This leads to the concept of the *vibrational density of states* (VDOS), which characterises how vibrational modes are distributed over frequency.

The VDOS, denoted as $g(\omega)$, is defined such that $g(\omega)\,d\omega$ gives the number of vibrational modes with frequencies in the interval $[\omega, \omega + d\omega]$. This quantity can be defined as in Definition 2.14 below.

Definition 2.14 (Vibrational density of states)**.** The *vibrational density of states* (VDOS) of a system of n_{atoms} atoms with n_{ext} external degrees of freedom removed is a function $g(\omega)$ defined by

$$g(\omega) = \frac{1}{3n_{\text{atoms}} - n_{\text{ext}}} \sum_{k=1}^{3n_{\text{atoms}}-n_{\text{ext}}} \delta(\omega - \omega_k), \tag{2.110}$$

where ω_k are the eigenfrequencies of the system's normal modes. The function $g(\omega)\,d\omega$ represents the fraction of vibrational modes with frequencies in the interval $[\omega, \omega + d\omega]$. The VDOS is, by convention, normalized so that

$$\int_0^\infty g(\omega)\,d\omega = 1. \tag{2.111}$$

Note that n_{ext} is typically 6 for non-linear and 5 for linear molecules, as discussed in the previous section. The peaks in $g(\omega)$ correspond to dominant

vibrational bands of the system.

In general, the distribution $g(\omega)$ is not known explicitly from the system's Hamiltonian or the initial conditions. At finite temperature, the vibrational frequencies ω_k are not directly accessible due to thermal fluctuations, which obscure the underlying harmonic modes. This is further complicated by the use of thermostats in molecular dynamics simulations, which modify the effective Hamiltonian through deterministic or stochastic feedback mechanisms, thereby driving the system away from purely Hamiltonian dynamics[49].

Consequently, it is not always feasible to extract vibrational frequencies by simply computing the square roots of the Hessian eigenvalues. The system is typically not in a local equilibrium configuration, and the exact form of the full Hamiltonian—including all thermal and environmental effects—is not available [50]. This necessitates the use of alternative methods that do not rely on knowledge of the Hessian or the underlying potential energy surface. One such approach is to extract vibrational information directly from the molecular dynamics trajectory, specifically via the *velocity autocorrelation function* [51, 52], as introduced in Definition 2.15.

Definition 2.15 (Velocity autocorrelation function)**.** The *velocity autocorrelation function* (VACF) of a system of n_{atoms} atoms is defined as

$$C_v(t) = \frac{1}{3n_{\text{atoms}}} \sum_{i=1}^{n_{\text{atoms}}} \big\langle {v_{(i)}}^{\mu}(0)\, v_{(i)\mu}(t) \big\rangle, \tag{2.112}$$

where ${v_{(i)}}^{\mu}(t)$ is the μ-th Cartesian component of the velocity of atom i at time t, and $\langle\cdot\rangle$ denotes either an ensemble average or, under suitable conditions, a time average over a sufficiently long molecular dynamics trajectory. The VACF measures the time correlation of atomic velocities and is used to obtain the vibrational density of states (VDOS) via the cosine Fourier transform:

$$g(\omega) = \frac{2}{\pi} \int_0^{\infty} C_v(t) \cos(\omega t)\, dt. \tag{2.113}$$

In practice, the VACF is typically computed as a time average over a single trajectory. For Equation (2.113) to yield a valid vibrational density of states, the system must be *ergodic*, meaning that time averages approximate ensemble averages over long timescales.

This assumption is especially important in the presence of thermal fluctuations or anharmonicity. Stochastic thermostats, such as Langevin or velocity rescaling, generally promote ergodicity but introduce random forces that can significantly perturb dynamical correlations, such as those in the VACF. Deterministic thermostats like Nosé–Hoover can also disturb the VACF, and may further fail to ensure ergodicity in small or weakly chaotic systems [28, 50, 53].

By Equation (2.98), the harmonic approximation allows one to decompose the atomic velocities into independent contributions from normal modes, each oscillating sinusoidally at its characteristic frequency ω_k. This is done by projecting the Cartesian velocities onto the eigenbasis $\hat{n}^{(k)\mu}$, yielding

$$v^k(t) := \hat{n}^{(k)\mu}\, \dot{r}_\mu(t), \tag{2.114}$$

where $v^k(t) \in \mathbb{R}$ is the time derivative of the k-th normal coordinate, i.e., the modal velocity along the k-th vibrational mode. Inverting this relation using the completeness of the eigenbasis, the full atomic velocity can be expressed as a sum over mode contributions:

$$\dot{r}_\mu(t) = \hat{n}^{(k)}{}_\mu\, v^k(t). \tag{2.115}$$

Substituting this expression into the definition of the VACF, one obtains:

$$C_v(t) = \frac{1}{3n_{\text{atoms}}} \langle \dot{r}_\mu(0)\, \dot{r}^\mu(t) \rangle \tag{2.116}$$

$$= \frac{1}{3n_{\text{atoms}}} \big\langle \hat{n}^{(k)}{}_\mu v^k(0)\, \hat{n}^{(\ell)\mu} v^\ell(t) \big\rangle \tag{2.117}$$

$$= \frac{1}{3n_{\text{atoms}}} \hat{n}^{(k)}{}_\mu\, \hat{n}^{(\ell)\mu} \big\langle v^k(0) v^\ell(t) \big\rangle. \tag{2.118}$$

Assuming the eigenvectors form an orthonormal basis with respect to the Euclidean inner product:

$$\hat{n}^{(k)}{}_\mu\, \hat{n}^{(\ell)\mu} = \delta^{k\ell}, \tag{2.119}$$

and that the normal modes are dynamically uncorrelated:

$$\big\langle v^k(0)v^\ell(t)\big\rangle = \delta^{k\ell}\big\langle v^k(0)v^k(t)\big\rangle, \tag{2.120}$$

one obtains the simplified form:

$$C_v(t) = \frac{1}{3n_{\text{atoms}}}\sum_k \big\langle v^k(0)\, v^k(t)\big\rangle. \tag{2.121}$$

In the harmonic approximation, each normal mode behaves like an independent oscillator, yielding:

$$\big\langle v^k(0)\, v^k(t)\big\rangle \propto \cos(\omega_k t), \tag{2.122}$$

so that

$$C_v(t) \propto \sum_k \cos(\omega_k t). \tag{2.123}$$

Taking the cosine Fourier transform then gives the vibrational density of states:

$$g(\omega) = \frac{2}{\pi}\int_0^\infty C_v(t)\cos(\omega t)\, dt \approx \sum_k \delta(\omega - \omega_k), \tag{2.124}$$

showing that the VACF encodes the spectral distribution of vibrational frequencies sampled dynamically.

Note that the derivation above assumes the system remains close to a local equilibrium configuration, such that the motion of atoms can be well-approximated by harmonic oscillations. In more general cases—including those with anharmonic interactions—the identification of the VACF with the vibrational spectrum is justified through a more rigorous statistical framework.

The *Wiener–Khinchin theorem* states that the power spectral density $S(\omega)$ of a wide-sense stationary stochastic process is the Fourier transform of its autocorrelation function [54, 55]:

$$S(\omega) = \int_{-\infty}^{\infty} C(t)\, e^{-i\omega t}\, dt. \tag{2.125}$$

This result does not require the system to be harmonic or linear—it only assumes stationarity and sufficient decay of correlations [56, 57].

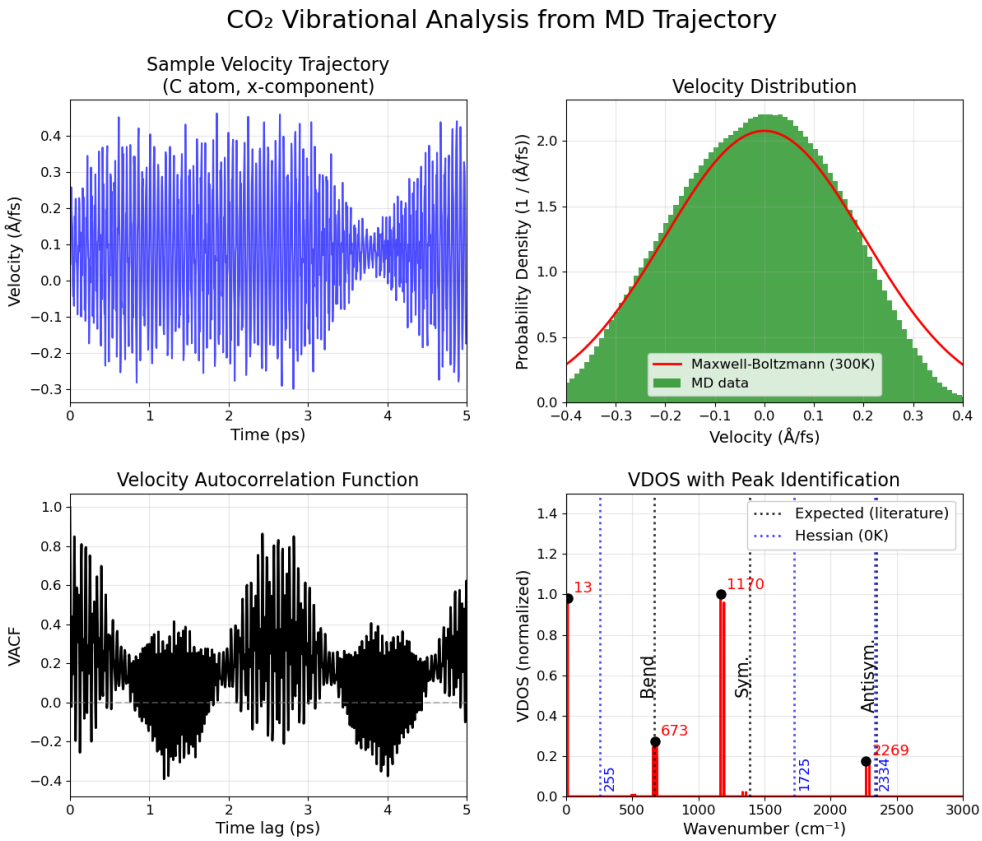


Figure 2.10. *Vibrational analysis of carbon dioxide from molecular dynamics simulation using the Hamiltonian defined in Equation* (2.4)*. The system was thermalised at* $300\,\mathrm{K}$ *for* $100\,\mathrm{ps}$ *using a Langevin thermostat, followed by a* $100\,\mathrm{ps}$ *microcanonical (NVE) trajectory for analysis. The top left panel shows a representative velocity trajectory of one of the carbon atoms* (x*-component*)*. The top right panel compares the velocity distribution from MD (green histogram) with the Maxwell–Boltzmann distribution at* $300\,\mathrm{K}$ *(red curve). The bottom left panel shows the velocity autocorrelation function (VACF), and the bottom right panel presents the vibrational density of states (VDOS) obtained via Fourier transform of the VACF. Vertical lines indicate reference vibrational frequencies: black lines correspond to experimental values from literature, and blue lines mark the normal mode frequencies computed from the Hessian at* $0\,\mathrm{K}$*.*

In the context of molecular dynamics, the VACF $C_v(t)$ behaves as the autocorrelation of a stationary observable (the atomic velocity), provided the trajectory is sampled from an equilibrated ensemble. Consequently, its Fourier transform yields a spectrum $g(\omega)$ that encodes the *distribution of frequency components actually sampled during the simulation*—that is, the dynamically accessible vibrational modes at finite temperature, including the effects of anharmonicity, mode coupling, and thermal noise [53, 58, 59].

The top right panel of Figure 2.10 shows the distribution of Cartesian velocity components sampled from the molecular dynamics trajectory, plotted as a histogram (green). Overlaid is the theoretical Maxwell–Boltzmann distribution at $300\,\mathrm{K}$ (red). The excellent agreement between the two confirms that the system was properly thermalised and is sampling from the canonical ensemble. This match validates the use of velocity-based observables such as the VACF and supports the interpretation of the trajectory as a representative sample of equilibrium dynamics.

The bottom left panel of Figure 2.10 displays the velocity autocorrelation function (VACF), computed from the same post-thermalisation trajectory. The VACF exhibits clear oscillatory structure. This indicates the presence of coherent vibrational motion in the system, dominated by the normal modes of CO_2. The well-resolved oscillations confirm that vibrational frequencies are dynamically accessible and justify the use of Fourier analysis to extract the vibrational density of states.

The bottom right panel of Figure 2.10 presents the VDOS of CO_2 computed from the VACF. The lowest-frequency peak, corresponding to the bending mode, shows good agreement with both the expected literature value. However, the symmetric and antisymmetric stretch modes exhibit more notable deviations. The main peak associated with the symmetric stretch appears significantly redshifted compared to the literature value, although two smaller adjacent peaks are present near the expected position. The antisymmetric stretch mode lies somewhat closer to the experimental reference, though still

exhibits a shift relative to the expected result from the literature. These discrepancies arise primarily due to the simplicity of the model Hamiltonian and the empirical spring constants used, which do not capture the full complexity of CO_2's anharmonic potential energy surface. Nevertheless, the extracted VDOS captures the qualitative structure of the vibrational spectrum and demonstrates that even a minimal model can recover the dominant vibrational features with some fidelity.

The discrepancy between the $0\,\mathrm{K}$ Hessian eigenvalue predictions, the expected vibrational spectral lines from the literature, and the VDOS obtained from the molecular dynamics trajectory arises from two main factors. First, the spring constants used in the model, taken from [9], are empirical parameters fit to experimental data at 300 K and are not optimized for zero-temperature predictions. Second, the Hamiltonian used in this analysis, given in Equation (2.4), represents a simplified approximation of the true interatomic interactions in CO_2 and lacks the complexity required for quantitative accuracy.

The VDOS plot in Figure 2.10 highlights the importance of using more sophisticated parameterisations when simulating molecular systems. In the main chapters of this work, the GAFF force field [60] was adopted to construct the Hamiltonian in order to overcome these limitations. Such parameter sets are referred to as *force fields*, and they provide empirically or semi-empirically derived coefficients that define the Hamiltonian and ensure that the resulting forces reproduce realistic molecular behaviour.

The unit cm^{-1} is commonly used to express vibrational frequencies in spectroscopy, and it relates to the angular frequency unit ps^{-1} through the conversion:

$$1\,\mathrm{ps}^{-1} = \frac{1}{c} \times 10^{12}\,\mathrm{cm}^{-1} \approx 33.356\,\mathrm{cm}^{-1}, \tag{2.126}$$

where c is the speed of light.

2.2.2.1 Infrared and Raman spectra

This work focuses on the computation of vibrational spectral information using classical molecular dynamics simulations. This is why in this work, the term *vibrational spectrum* is used synonymously to VDOS. In contrast, experimental access to vibrational information is typically achieved through indirect spectroscopic methods.

The most widely used techniques include infrared (IR) absorption spectroscopy, which detects vibrational transitions that involve a change in the dipole moment; and Raman spectroscopy, which measures inelastic light scattering due to fluctuations in polarizability [61–63].

Although the present work is based on classical MD trajectories and the resulting VDOS, IR and Raman experimental methods are discussed here for completeness.

In infrared spectroscopy, absorption arises when molecular vibrations induce a time-dependent dipole moment. In classical molecular dynamics, the *IR intensity* $I_{\mathrm{IR}}(\omega)$ is given by the Fourier cosine transform of the dipole moment autocorrelation function:

$$I_{\mathrm{IR}}(\omega) \propto \int_0^\infty \langle \mu^\alpha(0)\mu_\alpha(t)\rangle \cos(\omega t) dt \tag{2.127}$$

where μ^α is the α-component of the total dopole moment, and the angle brackets denote ensemble (or time, if ergodicity is guranteed) averages. This approach captures the dynamical dipolar fluctuations that correspond, in experiment, to the absorption of infrared radiation by vibrational modes that induce changes in the molecular dipole [64, 65].

Raman spectroscopy probes vibrational modes that induce time-dependent changes in the molecular polarizability tensor (denoted as $\pi^{\beta\nu}$ below). The molecular polarizability tensor is defined as the linear response of the induced dipole moment μ^α to an applied external electric field E^ν:

$$\pi^{\beta\nu} = \left.\frac{\partial \mu^\beta_{\mathrm{ind}}}{\partial E_\nu}\right|_{E=0}, \tag{2.128}$$

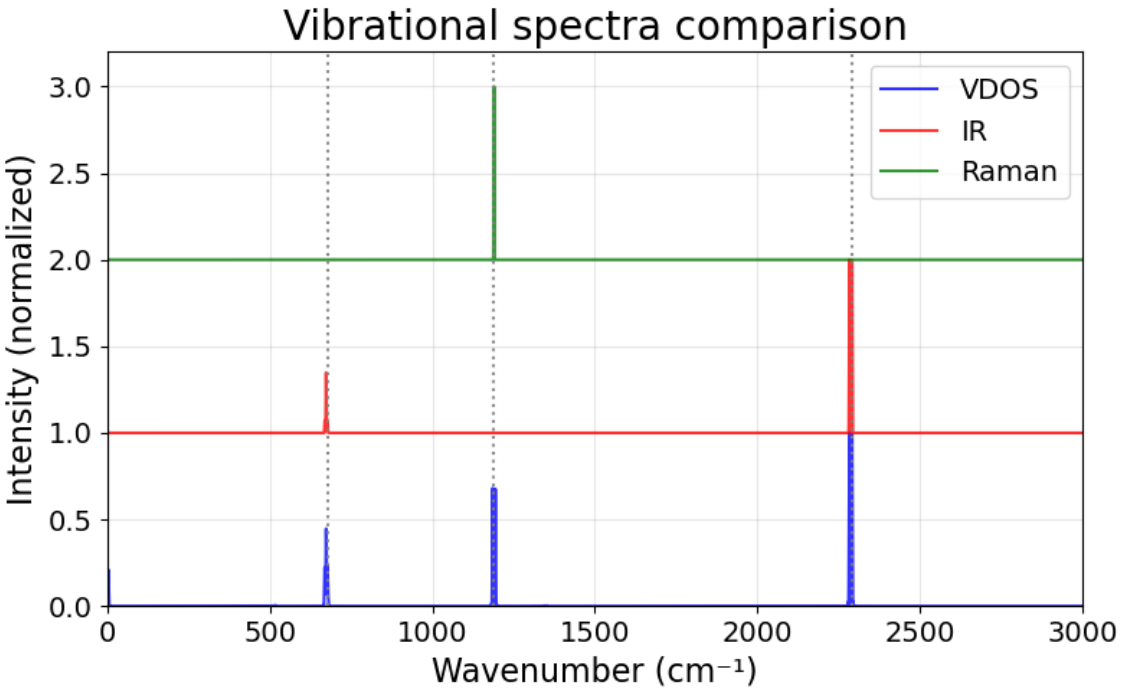


Figure 2.11. *Comparison of VDOS (blue, bottom), IR (red, middle), and Raman (green, top) spectra of CO_2, computed using the same simulation protocol as in Figure 2.10. Intensities are normalised to a maximum of 1, and the IR and Raman spectra are vertically offset by 1 and 2 units, respectively, for visibility.*

with $\mu^{\beta}_{\mathrm{ind}}$ is the component of the dipole moment in β-th Cartesian direction in response to a perturbation in the ν-th component of the external electric field E^{ν}.

In the context of classical molecular dynamics, the *Raman intensity*, I_{Raman} is computed from the Fourier cosine transform of the autocorrelation function of the polarizability:

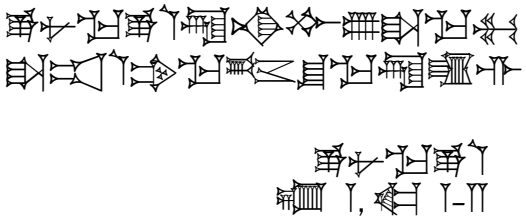
$$I_{\mathrm{Raman}}(\omega) \propto \int_0^{\infty} \langle \pi^{\mu\nu}(0)\pi_{\mu\nu}(t)\rangle \cos(\omega t) dt. \tag{2.129}$$

Raman-active vibrational modes are those that cause the polarizability to fluctuate over time, which in experiment leads to inelastic scattering of incident light at shifted frequencies.

Figure 2.11 compares the vibrational density of states (VDOS), infrared (IR), and Raman spectra of carbon dioxide, all derived from the system governed by the Hamiltonian in Equation (2.4), using the same molecular dynamics protocol described for Figure 2.10. The VDOS (blue) captures the full spectrum of vibrational activity, with peaks corresponding to all thermally sampled normal modes. The IR spectrum (red) shows peaks only for modes that involve time-dependent changes in the dipole moment—specifically, the bending and antisymmetric stretch modes. The Raman spectrum (green) displays only those modes that involve fluctuations in the polarizability tensor, with the symmetric stretch being prominently Raman-active. Note that the union of IR and Raman spectra recovers all the peaks visible in the VDOS; this reflects the complementary selection rules of the two spectroscopic techniques, which together can, in principle, resolve all vibrational modes in a centrosymmetric molecule like CO_2, although practical detection may be limited by weak intensities or overlapping bands [66].

CHAPTER 3

Empirical understanding of vibrational spectra: identifying peaks from spectral signals



At the peaks, the heaven is yet to be named, and rigid ground below is yet to be called by name

ENŪMA ELIŠ
Tablet 1, line 1 - 2

While the previous sections focused on the theoretical aspects of vibrational properties of molecules, it is important to acknowledge that most vibrational spectra are obtained from experiments. In many cases, researchers rely on pre-built software packages to identify spectral peaks and fit distributions without a thorough understanding of the underlying algorithms and methods employed. Moreover, many widely used software tools are commercial products, and their proprietary nature conceals the specific methodologies implemented. Consequently, researchers are frequently limited to learning how to use the software rather than comprehending the technical foundations of peak identification and spectral decomposition techniques.

The objective of this chapter is to provide a comprehensive introduction to implementing a custom, problem-specific peak decomposition scheme for the analysis of vibrational spectral data. This approach empowers researchers to develop tailored solutions that enhance their understanding of both the data

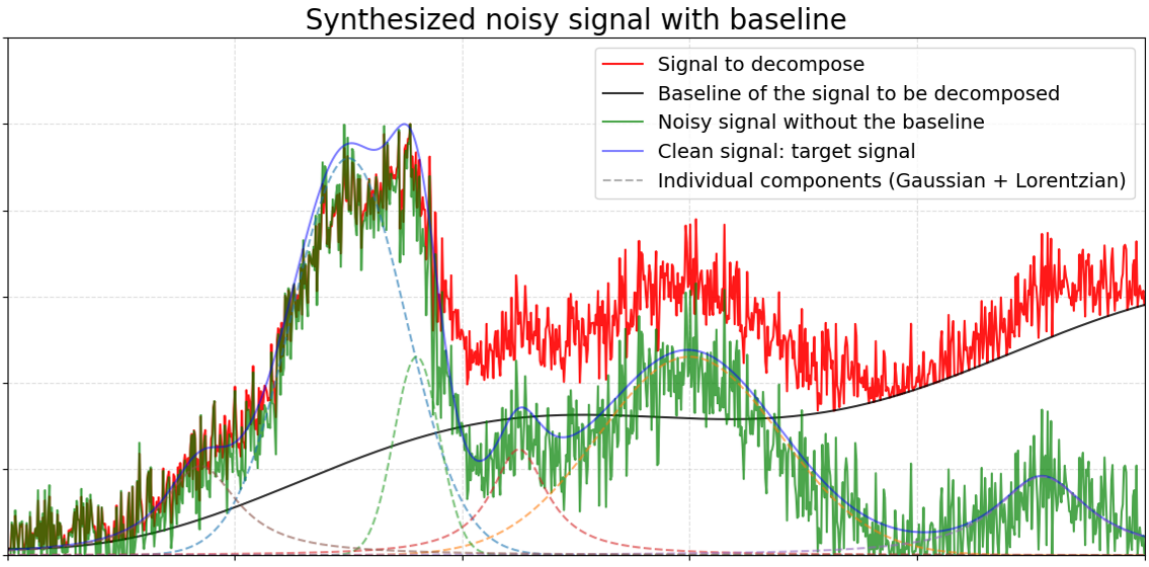


Figure 3.1. *A sample noisy signal generated by three Gaussian and three Lorentzian distributions with a sinusoidal baseline with linear trend. The target signal to be decomposed into distinct Lorentzian and Gaussian distributions are shown in red, where the target signal to retrieve is in solid blue. The expected baseline, Gaussian, and Lorentzian components of the signal are in dotted lines.*

and the methods used to interpret it.

In the theoretical sections of this chapter, a synthetic signal resembling a vibrational spectrum will be employed, while tests involving experimentally obtained spectra are presented in Section 3.5. The synthetic signal is designed to be positive, piecewise continuous, and twice differentiable within its continuous domains. These characteristics are essential since the signal cannot change its sign, and both continuity and second-order differentiability are required for the signal to be decomposable into known distributions. Figure 3.1 presents the synthetic signal utilized in this chapter, providing a visual representation of the problem addressed herein.

Figure 3.1 shows that identifying peaks and decomposing the signal into known distributions can be systematically divided into steps:

1. **Denoising:** If the signal contains significant noise, preprocessing is required. If the domain of interest is known in advance, it is advised to isolate the signal corresponding to the given domain for precision.

2. **Interpolation:** In cases where the dataset consists of a limited number of discrete points, interpolation may be necessary to approximate a continuous representation of the signal. The data can either be discarded (under-sampling) or expanded (over-sampling) by interpolation.

3. **Baseline removal:** A global trend, or baseline present in the signal may be spurious to the target signal. A baseline can obscure spurious signals that may hide small peaks or create false signals that could be mistaken for real peaks. [67]

4. **Peak identification:** Peaks, i.e. physically relevant local maxima of the signal, are then detected, using appropriate algorithms tailored to the characteristics of the signal.

5. **Signal decomposition:** Finally, the signal is decomposed into its constituent distributions, leveraging models that align with its physical or mathematical properties.

To ensure accuracy and reliability in each step, appropriate methods must be chosen based on the specific properties of the signal and the intended application. For example, noise filtering may employ wavelet denoising or Savitzky-Golay methods, depending on the frequency characteristics of the noise (compare: [68, 69]). Similarly, peak identification could leverage techniques such as window-based methods, derivative-based methods, or even machine learning algorithms for enhanced robustness. These considerations play a critical role in ensuring the validity of the overall decomposition process.

3.1 Denoising and interpolating signals

This section focuses on two critical preprocessing steps for spectroscopic signals: denoising and interpolation. The Savitzky-Golay filter and wavelet transform methods are introduced as the primary algorithms for denoising. The Savitzky-Golay filter is well suited for preserving the shape and width of spectral features while smoothing the signal, making it highly effective, in particular for molecular spectra that are expected to have underlying distributions of expected form. The wavelet transform provides a versatile approach for signals with varying frequency components, allowing efficient noise reduction across different scales.

For interpolation, cubic splines are employed here to reconstruct a smooth representation of the signal from a limited set of discrete data points. The ability of splines to ensure smoothness and second-order differentiability aligns with the requirements for subsequent signal decomposition and peak analysis.

3.1.1 Savitzky-Golay algorithm

The Savitzky-Golay filter, developed in 1964 at the Perkin-Elmer Corporation, was designed to smooth noisy chemical spectral data while preserving the shape and height of spectral peaks [70, 71]. This method combines least-squares polynomial fitting (see Appendix B for details) with moving average techniques, implementing smoothing operation through convolution with a fixed set of coefficients. The filter operates by fitting a polynomial of degree n to a local subinterval of the domain with $2m + 1$ data points. The window size and polynomial order are selected based on the characteristics of the spectral data [72].

This approach is especially effective in spectroscopic applications where maintaining peak characteristics is critical, offering superior performance compared to simple moving averages that tend to distort peak heights and widths.

This section introduces a modified version of the Savitzky-Golay method, developed by Michael Schmid, David Rath, and Ulrike Diebold [73], to better handle signal boundaries. The method is referred to as *Modified Sinc* smoothing because it uses a modified kernel based on the sinc function:

$$\operatorname{sinc}(x) = \frac{\sin(x)}{x} \tag{3.1}$$

The sinc function plays a crucial role in signal processing as it represents the impulse response of an ideal low-pass filter. If noise occurs at frequencies much higher than the signal, the sinc kernel effectively suppresses noise without distorting the signal.

Although the sinc function is theoretically ideal for low-pass filtering, its infinite extent and the resulting ringing artifacts make it impractical for direct use. To overcome this limitation, a modified kernel function is defined that can be effectively convolved with the input signal while maintaining the desirable properties of a low-pass filter. The modified sinc kernel is defined as:

$$k(x) = \begin{cases} 1, & \text{if } x = 0, \\ \frac{2}{\pi(d+\gamma)x} \sin\left(\frac{\pi(d+\gamma)x}{2}\right) \cdot w(x) + c(x), & \text{otherwise.} \end{cases} \tag{3.2}$$

This kernel incorporates three critical modifications to the basic sinc function:

1. **Scaling Factor:** The scaling factor $\pi(d+\gamma)/2$ in the argument controls the filter's cutoff frequency. Here, d is the (even) degree parameter controlling smoothing strength (typically between 2 and 10), and γ is a variant-dependent parameter.

2. **Window Function:** A window function $w(x)$ ensures the kernel has finite support and smooth decay to zero at its boundaries, making it practically implementable. The window function is defined as:

$$w(x) = e^{-\alpha x^2} + e^{-\alpha(x-2)^2} + e^{-\alpha(x+2)^2} - 2e^{-\alpha} - e^{-9\alpha} \tag{3.3}$$

3. **Correction Terms**: Correction terms $c(x)$, defined as a sum of weighted sine functions, are added to optimize the filter's frequency response:

$$c(x) = \sum_j \kappa_j x \sin((2j + \nu)\pi x) \tag{3.4}$$

where $\alpha \in \{2, 4\}$. This function ensures smooth decay and finite support. The correction coefficients κ_j depend on both the degree d and the kernel half-width m:

$$\kappa_j = a_j + \frac{b_j}{(c_j - m)^3} \tag{3.5}$$

The parameter ν is degree-dependent, taking the value 1 for degrees 6 and 10, and 2 for degree 8. The coefficients a_j, b_j, and c_j are empirically determined to optimize the filter's passband flatness. For lower degrees (2 and 4), no correction terms are needed ($c(x) = 0$).

This carefully constructed combination allows the kernel to maintain many desirable properties of the ideal sinc filter while being practical for real-world data. The window function ensures smooth decay at the boundaries, and the correction terms optimize the filter's frequency response, particularly for preserving signal features in the passband region.

To address boundary effects, the modified sinc approach uses weighted linear extrapolation instead of simple padding or reflection. The extrapolation employs a Hann-weighted linear regression (see Appendix B for details), with weights, ω_{fit}, that decay smoothly according to:

$$w_{\text{fit}}(p) = \cos^2\left(\frac{\pi p}{2 z_1 \beta}\right) \tag{3.6}$$

where z_1 is the first zero of the sinc function, calculated as $(m + 1)/(1 + 0.5d)$ if $\alpha = 2$ or $(m + 1)/(1.5 + 0.5d)$ if $\alpha = 4$, and β is a decay factor depending on the degree:

$$\beta = \begin{cases} 0.65 + 0.35e^{-0.55(d-4)}, & \text{if } \alpha = 2, \\ 0.70 + 0.14e^{-0.60(d-4)}, & \text{if } \alpha = 4. \end{cases} \tag{3.7}$$

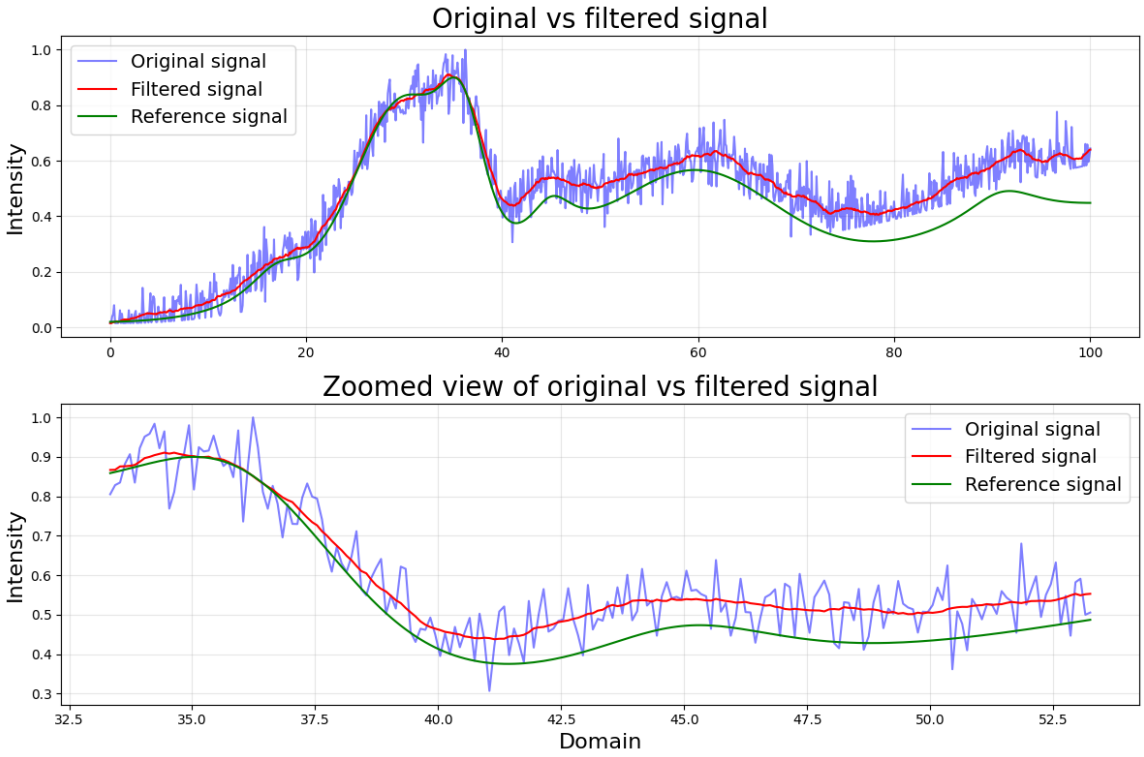


Figure 3.2. *A graphical demonstration of the denoising method introduced in this section. The parameters used to generate the denoised signal are* $d = 6$*, and* $m = 91$.

The final smoothed operation can be expressed as:

$$y_{\text{smooth}}[n] = \sum_{i=-m}^{m} k[i]y[n-i] \tag{3.8}$$

Here, the square bracket indicates the index within the window defined by the kernel half-width m.

The graphical example in Figure 3.2 illustrates how the noisy test signal from Figure 3.1 can be denoised using this modified method. There are four key features to note in Figure 3.2:

1. The peak locations and shapes are accurately preserved while noise is effectively suppressed. This demonstrates how the modified sinc kernel

maintains the signal's essential features through its carefully designed frequency response characteristics.

2. The boundary regions show smooth transitions without many artifacts, thanks to the Hann-weighted linear extrapolation. Unlike traditional methods, this approach provides more reliable treatment of signal endpoints, which is particularly valuable when analyzing spectroscopic data where edge features are significant.

3. The overall smoothing effect successfully distinguishes overlapping peaks while maintaining their relative intensities. The combination of the windowed sinc function and correction terms allows for effective noise reduction without compromising the resolution of closely spaced features, making it particularly suitable for analyzing complex spectroscopic profiles.

4. When the signal-to-noise ratio is low, the algorithm may incorrectly identify noise fluctuations as genuine peaks, leading to false-positive peak detections. This is where the wavelet transform method introduced in the next section outperforms.

3.1.2 Continuous wavelet transform method

The limitations of Fourier transform in analyzing non-stationary noisy signals, where frequency content varies over time, motivated scientists and mathematicians of the 20^{th} century to extend the Gabor transform. This led to the development of alternative methods that replace Fourier components with other sets of orthonormal basis functions. One of the most popular of such methods is the wavelet transformation, which overcomes the fixed frequency resolution of Fourier transforms by allowing variable resolution at different scales.

The wavelet transformation is based on the concept of a **mother wavelet**, a prototype function that generates a family of wavelets. By scaling and trans-

lating the mother wavelet, an orthonormal set of wavelets is created, enabling a more flexible time-frequency analysis. With this in mind, one can define the continuous wavelet transformation as in Definition 3.1.

Definition 3.1 (Continuous wavelet transform)**.** Given a continuous signal $s : \mathbb{R} \to \mathbb{R}$ and a continuous function $\psi : \mathbb{C} \to \mathbb{R}$, called the *mother wavelet*, the (*continuous*) *wavelet transformation* of the signal s under the mother wavelet ψ is defined as

$$W(a,b) = \int_{-\infty}^{\infty} s(x)\psi_a{}^*\left(\frac{x-b}{a}\right)dx. \tag{3.9}$$

$a, b \in \mathbb{R}$ are referred to as the scale and translation parameters, respectively, and ψ^* denotes the complex conjugate of the mother wavelet.

This definition can be generalized to higher dimensional spaces, but in this work, the 1-dimensional defintion is sufficient.

One of the simplest wavelets is the Haar wavelet, which is defined as:

$$\psi_{\text{Haar}}(t) = \begin{cases} 1 & 0 \leqslant t < \frac{1}{2} \\ -1 & \frac{1}{2} \leqslant t < 1 \\ 0 & \text{otherwise} \end{cases}. \tag{3.10}$$

Substituting this to the Definition 3.1, one can immediately obtain the transformation coefficients at different scales and translations. Figure 3.3 demonstrates the continuous wavelet transform using the Haar wavelet on a non-stationary signal. The Haar mother wavelet (top) serves as a basis function, characterized by its piecewise constant nature with a positive step followed by a negative step. The input signal (middle) exhibits varying frequency content and amplitude modulation across its duration, with notable transitions around t=40.

The wavelet transform coefficients $|W(a,b)|^2$ (bottom) reveal the signal's time-frequency structure through a scalogram. The power spectrum intensity (shown in color) represents the strength of correlation between the signal and

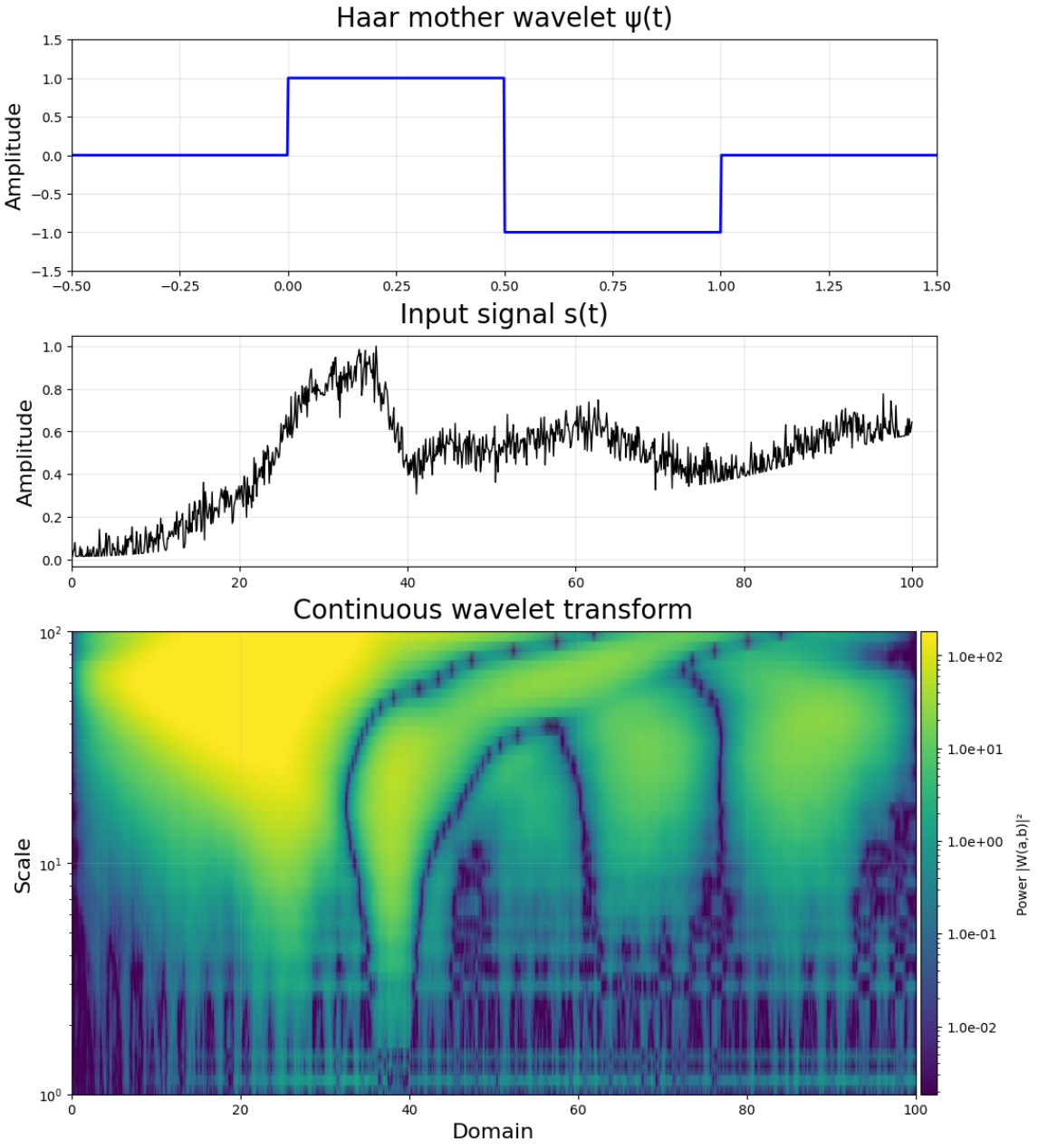


Figure 3.3. *Haar mother wavelet* (*TOP*), *our synthetic signal* (*MIDDLE*), *and its continuous Haar-wavelet transformation power spectrum* (*BOTTOM*).

the scaled wavelet at each time point. Lower scales (1-10) capture rapid signal variations and discontinuities, appearing as vertical structures in the transform, which correspond to the high-frequency noise fluctuations visible throughout the signal. Higher scales (10-100) detect slower variations and broader patterns, shown as horizontal bands of activity, revealing the signal's underlying trend changes. Strong coefficients (yellow regions) are concentrated in the interval $t \in [20, 40]$, corresponding to the principal transition period of the signal, characterised by a rapid increase in amplitude from approximately 0.2 to 1.0, followed by a sharp decrease to around 0.4. The transform also highlights subsequent minor transitions at $t \approx 60$ and $t \approx 80$, where the signal shows more subtle amplitude modulations around 0.6 and 0.4 respectively. The logarithmic scale spacing provides a natural way to analyze the signal across different time scales simultaneously, revealing both fine details and overall structure.

Despite its simplicity, Haar wavelets are not widely used in continuous signal denoising. Its discontinuity makes it only ideal for denoising discrete signals or signals with sharp transitions. Most popular and convenient wavelets are Daubechies wavelets, symlets, coiflets, Mexican hat wavelet, and biorthogonal wavelets.

Figure 3.4 presents the continuous wavelet transform using Daubechies-4 (DB4) wavelets, which offer improved frequency localization compared to Haar wavelets. The DB4 mother wavelet (top) exhibits smoother transitions and vanishing moments up to order 4, in contrast to the discontinuous step function of the Haar wavelet. This smoother basis function results in a more refined time-frequency decomposition, as evidenced in the wavelet transform coefficients (bottom). The transform reveals similar temporal features to the Haar analysis, with strong coefficients during the signal's major transition (t=20-40), but provides smoother scale transitions and better frequency discrimination, particularly in the higher scales (10-100) where the horizontal bands show clearer separation.

The wavelet denoising process relies on thresholding the wavelet coef-

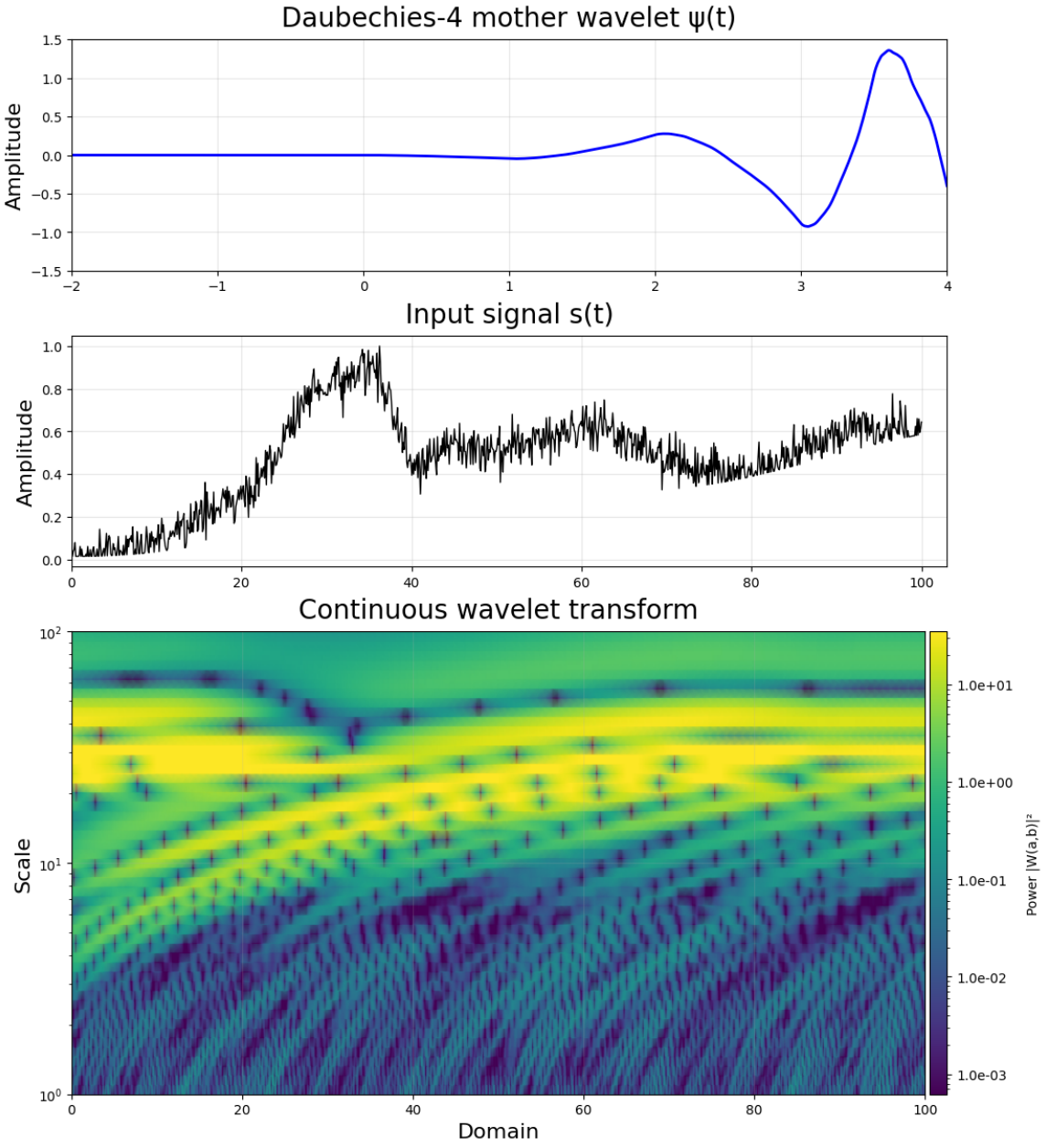


Figure 3.4. *Daubechies-4 mother* (*DB4*) *wavelet* (*TOP*), *our synthetic signal* (*MIDDLE*), *and its continuous DB4-wavelet transformation power spectrum* (*BOTTOM*).

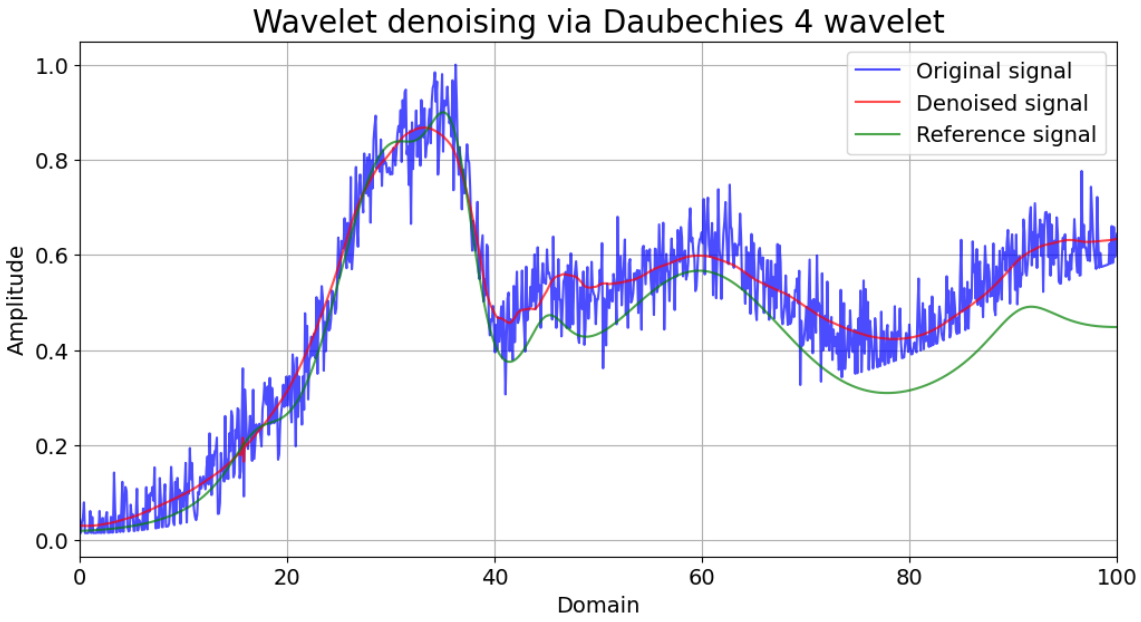


Figure 3.5. *Wavelet denoising using Daubechies-4 wavelets*.

ficients to separate signal from noise. The threshold is calculated using the universal threshold rule, which employs the Median Absolute Deviation (MAD) to estimate the noise level. For wavelet coefficients X, the MAD is defined as median$(|X - \text{median}(X)|)$, which is normalized to match Gaussian noise standard deviation: $\sigma = \text{MAD}/0.6745$. The universal threshold is then computed as $\lambda = \sigma\sqrt{2\log(N)}$, where N is the signal length. This threshold minimizes the maximum estimation error under certain statistical assumptions.

The thresholding operation can be performed using either hard or soft thresholding rules. Hard thresholding, defined as $\eta_H(x) = x\mathbb{I}_{|x|>\lambda}$, simply nullifies coefficients below the threshold. Soft thresholding, given by $\eta_S(x) = \text{sign}(x)(|x| - \lambda)\mathbb{I}_{|x|>\lambda}$, shrinks coefficients by the threshold amount, providing better continuity in the reconstructed signal. The signal reconstruction then proceeds by applying the inverse wavelet transform: $f(t) = \sum_k c_{A_n}[k]\tilde{\phi}_{n,k}(t) + \sum_{j=1}^{n}\sum_k c_{D_j}[k]\tilde{\psi}_{j,k}(t)$, where $\tilde{\phi}$ and $\tilde{\psi}$ are the dual scaling and wavelet functions, respectively, and c_{A_n}, c_{D_j} are the thresholded approximation and detail coeffi-

cients.

The resulting denoised denoised signal using this method can be obtained as shown in Figure 3.6

The signals obtained via Savitzky-Golay and wavelet transform methods have their own benefits and disadvantages. Each method exhibits distinct artifacts under specific conditions. The Daubechies-4 wavelet transform introduces a cusp at x = 18 where the peak magnitude falls below the noise level, while the Savitzky-Golay method creates spurious features near the boundary at x = 98 due to high-amplitude noise and insufficient data points to constrain the polynomial fit.

To incorporate the benefits and compensate for the limitations of both methods, one can apply a weighted sum of the denoised signals as shown below:

$$s^{\text{final denoised}}(x) = \alpha s^{SG}(x) + (1-\alpha)s^{CWT}(x) \tag{3.11}$$

where α is a number between 0 and 1, S^{SG}, s^{CWT} are denoised signals obtained via Savitzky-Golay and continuous wavelet transformation methods, respectively.

Figure 3.6 shows $s^{\text{final denoised}}(x)$ with different values of α. The curve given by $\alpha = 0.5$ yields the best outcome since it identifies the two overlapping peaks in $[28, 30]$ well, while having a less sharper drop at the boundaries compared to the cases with higher αs, and creates fewer peak-looking artefacts and the cusps (like the one shown at 18 in Figure 3.6) from the noise compared to the cases with lower αs.

3.1.3 Interpolating the denoised signal

Oftentimes, the data available might not be dense enough for correct baseline correction or peak-identification. This problem can be corrected via interpolation.

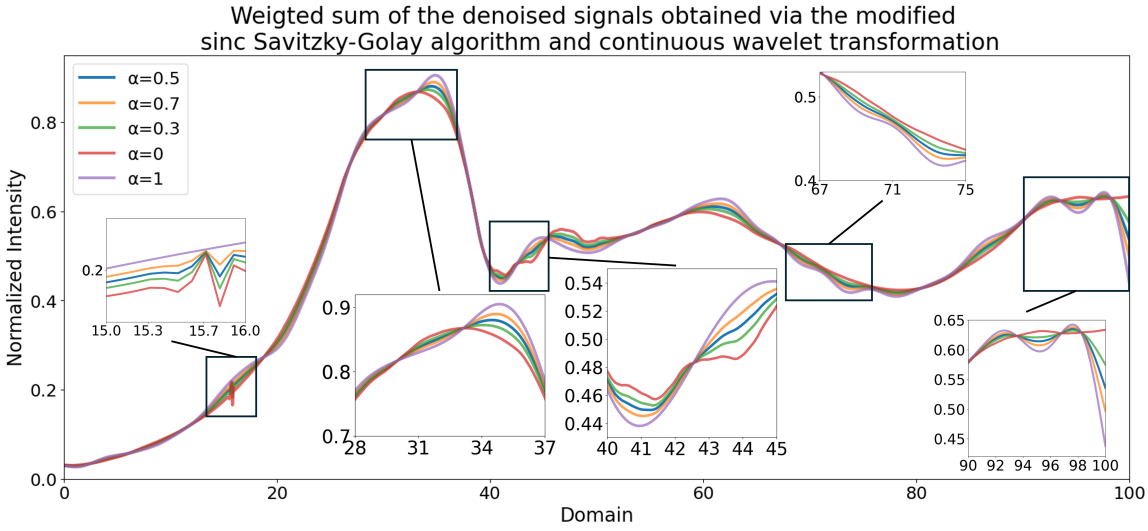


Figure 3.6. *Weighted sum of denoised signals obtained via Savitzky-Golay and continuous wavelet transform methods with different values of α as given in Equation 3.11.*

Definition 3.2 (Interpolation)**.** Let $\mathcal{V}$ be a linear space with norm, and for $B \subset \mathcal{V}$, let $\mathcal{V}(B)$ be the smallest subspace that contains B. With $A_0, A_1 \subset \mathcal{V}$, one can find a space called an *interpolation space*, $\mathcal{A}$, such that

$$\mathcal{V}(A_0) \cap \mathcal{V}(A_1) \subset \mathcal{A} \subset \mathcal{V}(A_0) + \mathcal{V}(A_1), \tag{3.12}$$

if there exists a function $\varphi^{\text{int}} : \mathcal{V}(A_0) \to \mathcal{V}(A_1)$ such that $\varphi^{\text{int}} : \mathcal{A} \to \mathcal{A}$ holds. This function is called the *interpolation* between A_0 and A_1. [74, 75]

For our application, A_0 can be thought of as the starting points to interpolate from and A_1 can be regarded as the end-points of the interpolation. As an illustrative example, the simplest interpolation method, namely piecewise linear interpolation, is considered, wherein interpolation is carried out between the ordered sets:

$$A_0 = \{(0,1),(1,2),(3,3),(7,1)\} \quad \text{the starting points} \tag{3.13}$$

and

$$A_1 = \{(1,2),(3,3),(7,1),(11,2)\} \quad \text{the end-points}, \tag{3.14}$$

via the interpolation map between the ℓ-th elements of A_0 and A_1:

$$\varphi^{\text{int}}_{\text{linear};\ell}(x) = \begin{cases} \frac{\Delta^1}{\Delta^2}(x - a^0_\ell[0]) + a^0_\ell[1] & \text{if } x \in (a^0_\ell[0], a^1_\ell[0]) \\ 0 & \text{otherwise} \end{cases} \tag{3.15}$$

with $\Delta^1 = a^1_\ell[1] - a^0_\ell[1]$ and $\Delta^2 = a^1_\ell[0] - a^0_\ell[0]$, where the superscript $k \in \{0,1\}$ indicates that $a^k \in A_k$ and $[m]$, $m \in \{0,1\}$, determines the position of the element in $a^k_\ell = \left(a^k_\ell[0], a^k_\ell[1]\right)$. Defining $\varphi^{\text{int}}_{\text{linear}}$ as

$$\varphi^{\text{int}}_{\text{linear}} = \sum_\ell \varphi^{\text{int}}_{\text{linear};\ell} \tag{3.16}$$

allows one to obtain the interpolation between A_0 and A_1. A visual representation of the linear interpolation process is provided in Figure 3.7 a), which illustrates how additional points are generated between A_0 and A_1.

Figure 3.7 b) demonstrates that the linear interpolation effectively approximates the signal. However, as shown in Figure 3.7 c), the numerical second derivatives exhibit substantial oscillations with excessively large magnitudes relative to the original signal.

The limitations of linear interpolation in accurately capturing second-order features underscore the need for an interpolation method that guarantees a continuously second-differentiable result. Several candidates, such as Bézier curves, B-splines, and cubic splines, are commonly used for this purpose. Among these, the cubic spline method provides a suitable balance of computational efficiency and smoothness, making it an ideal choice for addressing the deficiencies of linear interpolation. In the following sections, I will focus on implementing the cubic spline method to achieve the desired continuity and precision. For the details on least-squares methods, please consult Appendix B.

A function interpolating a given set of data is called a *spline* of order k if it consists of a set of smooth, piecewise-connected k-th order polynomials. That

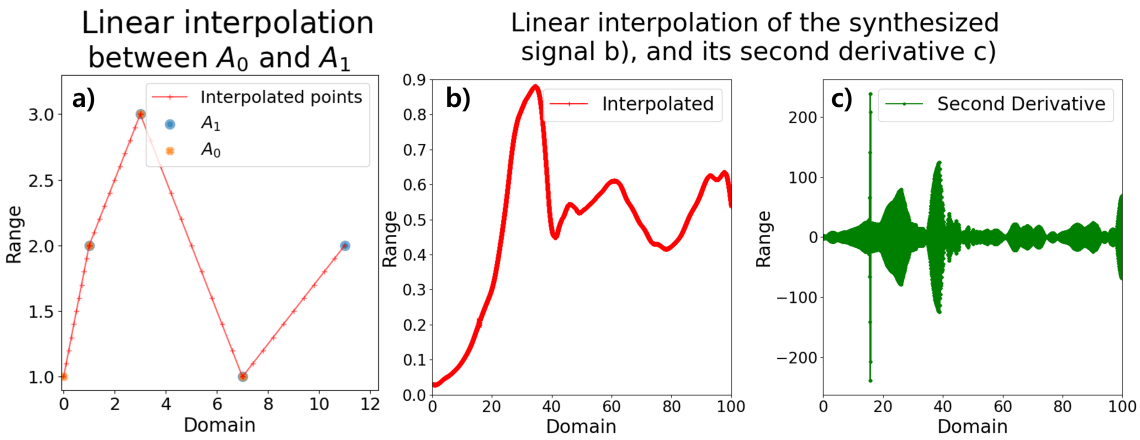


Figure 3.7. *a) Linear interpolation between A_0 and A_1 derived from Equations 3.13 and 3.14. b) Linear interpolation of the denoised synthetic signal. c) Numerical second derivative of the linearly interpolated result from b). Note the significant magnitude differences and rapid oscillations in the second derivative. The pronounced second derivative at domain = 18 is attributed to the cusp in the filtered signal.*

is, the domain is divided into intervals, and on each interval, the approximated function is a polynomial of degree k. In mathematical language:

> **Definition 3.3** (Spline)**.** Let $[x_i, x_f]$ and $[y_i, y_f]$ be the domain and the range of the given datapoints, N is the total number of datapoints. The *knots* of the domain are set of points $\{t_i\}_{i=0}^{i=N}$ such that
>
> $$x_i = t_0 < t_1 < \cdots < t_{N-1} < t_N = x_f. \tag{3.17}$$
>
> The *spline*, $s^{(k)} : [x_i, x_f] \to [y_i, y_f]$ is a function that
>
> 1) is a k-th order polynomial, p_j, in every $[t_j, t_{j+1}]$ for $j = 0, \cdots, N-1$,
> 2) the zeroth through $k-1$-th derivatives of p_js across adjacent intervals $[t_j, t_{j+1}]$, and
> 3) for all $j = 0, \cdots, N-1$, and $0 \leqslant \ell < k$, $\frac{d^\ell}{dx^\ell} p_j(t_{j+1}) = \frac{d^\ell}{dx^\ell} p_{j+1}(t_{j+1})$.[76]

One can obtain the A_0 and A_1 of Definition 3.2 via the definition of knots in Definition 3.3. Hence, a cubic spline interpolation map $\varphi^{\text{int}}_{\text{cubic}}$ can be constructed between A_0 and A_1. For the ℓ-th element, this map takes the form:

$$\varphi^{\text{int}}_{\text{cubic};\ell}(x) = \begin{cases} \alpha_\ell(x - a^0_\ell[0])^3 + \beta_\ell(x - a^0_\ell[0])^2 + \gamma_\ell(x - a^0_\ell[0]) + \delta_\ell & \text{if } x \in (a^0_\ell[0], a^1_\ell[0]) \\ 0 & \text{otherwise} \end{cases} \tag{3.18}$$

The coefficients α_ℓ, β_ℓ, γ_ℓ, and δ_ℓ are determined by enforcing the continuity and smoothness conditions at the knots from Definition 3.3:

$$\varphi^{\text{int}}_{\text{cubic};\ell}(a^0_\ell[0]) = a^0_\ell[1] \tag{3.19}$$

$$\varphi^{\text{int}}_{\text{cubic};\ell}(a^1_\ell[0]) = a^1_\ell[1] \tag{3.20}$$

$$\frac{d}{dx}\varphi^{\text{int}}_{\text{cubic};\ell}(a^1_\ell[0]) = \frac{d}{dx}\varphi^{\text{int}}_{\text{cubic};\ell+1}(a^1_\ell[0]) \tag{3.21}$$

$$\frac{d^2}{dx^2}\varphi^{\text{int}}_{\text{cubic};\ell}(a^1_\ell[0]) = \frac{d^2}{dx^2}\varphi^{\text{int}}_{\text{cubic};\ell+1}(a^1_\ell[0]) \tag{3.22}$$

These boundary conditions assume that the second derivatives of the spline are zero at the endpoints, reflecting a natural spline behavior where there is no abrupt change in curvature at the boundaries.

$$\frac{d^2}{dx^2}\varphi^{\text{int}}_{\text{cubic};1}(a^0_1[0]) = \frac{d^2}{dx^2}\varphi^{\text{int}}_{\text{cubic};n}(a^1_n[0]) = 0 \tag{3.23}$$

The complete interpolation map is then defined as:

$$\varphi^{\text{int}}_{\text{cubic}} = \sum_\ell \varphi^{\text{int}}_{\text{cubic};\ell}. \tag{3.24}$$

This construction guarantees the continuity of the second derivative across all interpolation points.

Cubic splines generate interpolations with minimal curvature, making them effective at approximating the original data-generating function.

Theorem 3.4 (Minimization of total extrinsic curvature in cubic spline interpolation)**.** *The cubic spline interpolation defined in* φ^{int}_{cubic} *guarantees the continuity of the second derivative across all interpolation points. Additionally, it minimizes the total curvature among all twice-differentiable cubic interpolants. Formally, the total extrinsic curvature is given by*:

$$I\left[\varphi^{int}_{cubic}\right] = \int (\varphi^{int}_{cubic})''(x)^2 dx, \tag{3.25}$$

which is minimized by the cubic spline interpolation compared to other possible interpolation methods.

Proof. This statement can be proved using variational principle via Euler-Lagrange equation. The integrand of Equation 3.25 simply is the second derivative of the function $\varphi^{\text{int}}_{\text{cubic}}$, meaning that the Euler-Lagrange equation becomes the following in (x_i, x_f)

$$\frac{d^2}{dx^2}\left(2{\varphi^{\text{int}}_{\text{cubic}}}''(x)\right) = 0 \implies {\varphi^{\text{int}}_{\text{cubic}}}''''(x) = 0. \tag{3.26}$$

At the endpoints, Equation 3.23 ensures that the Euler-Lagrange equation is zero in $[x_i, x_f]$. This completes the proof. □

Mechanical intuition. Equivalently, one can imagine an ideal one-dimensional, uniform elastic beam of constant flexural rigidity pinned so that it passes through each data point. When released, the beam "relaxes" in a shape that minimises its total bending energy,

$$E_{\text{bend}} \propto \int (y''(x))^2 \mathrm{d}x, \tag{3.27}$$

subject to the constraints $y(x_i) = y_i$. The Euler–Lagrange equation for that variational problem is exactly

$$y^{(4)}(x) = 0 \tag{3.28}$$

at each interval, with continuity of y, y' and y'' at the knots, precisely the conditions that define the natural cubic spline.

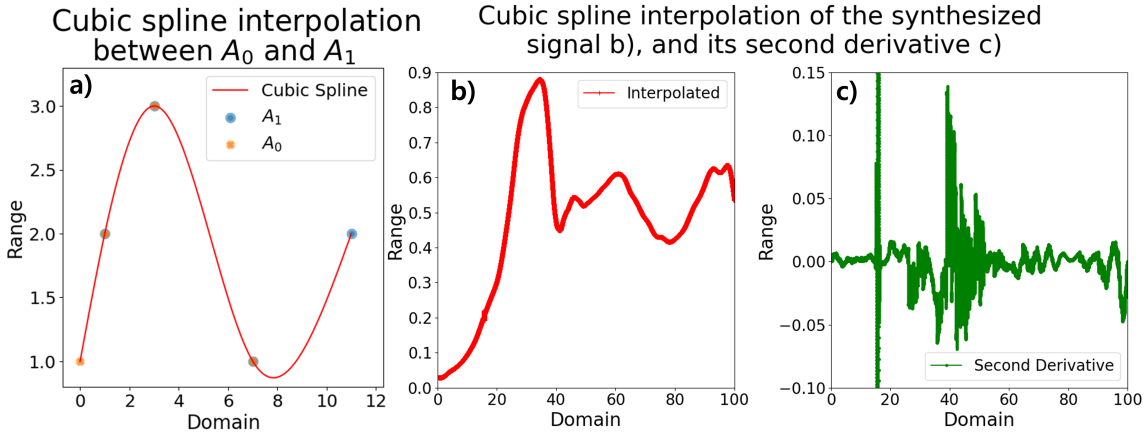


Figure 3.8. *a) Cubic spline interpolation between A_0 and A_1 derived from Equations 3.13 and 3.14—1000 datapoints are extended to have 10,000 datapoints. b) Cubic spline interpolation of the denoised synthetic signal. c) The second derivative of the cubic spline interpolated result from (b). Note that the local extrema of the second derivatives are more visible than in Figure 3.7 c) and the magnitudes are more realistic. The pronounced second derivative at domain = 18 is attributed to the cusp in the filtered signal.*

Figure 3.8 a) demonstrates that cubic spline interpolation effectively approximates the signal while maintaining the second-order smoothness. The second-order continuity guaranteed by the construction of $\varphi^{\text{int}}_{\text{cubic}}$ ensures smooth transitions between interpolation points, in contrast to the piecewise linear case where only zeroth-order continuity is maintained.

As shown in Figure 3.8 c), this higher-order smoothness manifests in the second derivatives, which exhibit significantly reduced oscillations compared to Figure 3.7 c). The elimination of artificial discontinuities is a direct consequence of enforcing the continuity conditions on $\frac{d^2}{dx^2}\varphi^{\text{int}}_{\text{cubic};\ell}$. Notably, while the second derivative still captures the cusp at domain = 18, the cusp manifests as a sharp peak with a maximum amplitude of 10.61. Since all other features in the signal's second derivative have magnitudes below 0.15, this pronounced difference in magnitude provides a clear criterion for distinguishing cusps from other signal features.

This combination of smoothness and accurate cusp detection, enabled by the minimal curvature property of cubic splines, makes them particularly suitable for applications requiring reliable second-order information.

3.2 Baseline correction of signals

Baseline correction aims to correct for the bias in the captured signal on coarse scales [1]. Empirically, one can view the baseline as the low-frequency or out-of-band content, including the 0-frequency (DC) offset introduced by instrument bias or drift, and hence treated as systematic error that must be removed to reveal the true features of the signal.

Formally, a baseline of a given signal can be defined as in Definition 3.5 below.

> **Definition 3.5** (Baseline)**.** Let $a, b \in \mathbb{R}$ with $a < b$, and let a signal be represented by a function $s : [a, b] \to \mathbb{R}$. The ***baseline*** of the signal s is a piecewise smooth function $B : [a, b] \to \mathbb{R}$ satisfying the following conditions:
>
> $$B(x) \leqslant s(x) \text{ for all } x \in [a, b], \tag{3.29}$$
>
> $$\operatorname{sgn}(B(x)) = \operatorname{sgn}(s(x)) \text{ for all } x \in [a, b], \tag{3.30}$$
>
> where $\operatorname{sgn}(\cdot)$ denotes the sign function. The function B approximates the lower envelope of s over $[a, b]$, capturing the trend or background level of the signal while excluding rapid variations or oscillatory components.

The simplest of baselines are linear baselines defined as a line connecting the first datapoint to the last datapoint by the linear relationship below

$$B^{\text{linear}}(x) = \frac{s(b) - s(a)}{b - a}(x - a) + s(a). \tag{3.31}$$

This method is intuitive, fast and straight-forward. As shown in Figure 3.9, the linear baseline correction works well without destroying the overall shape

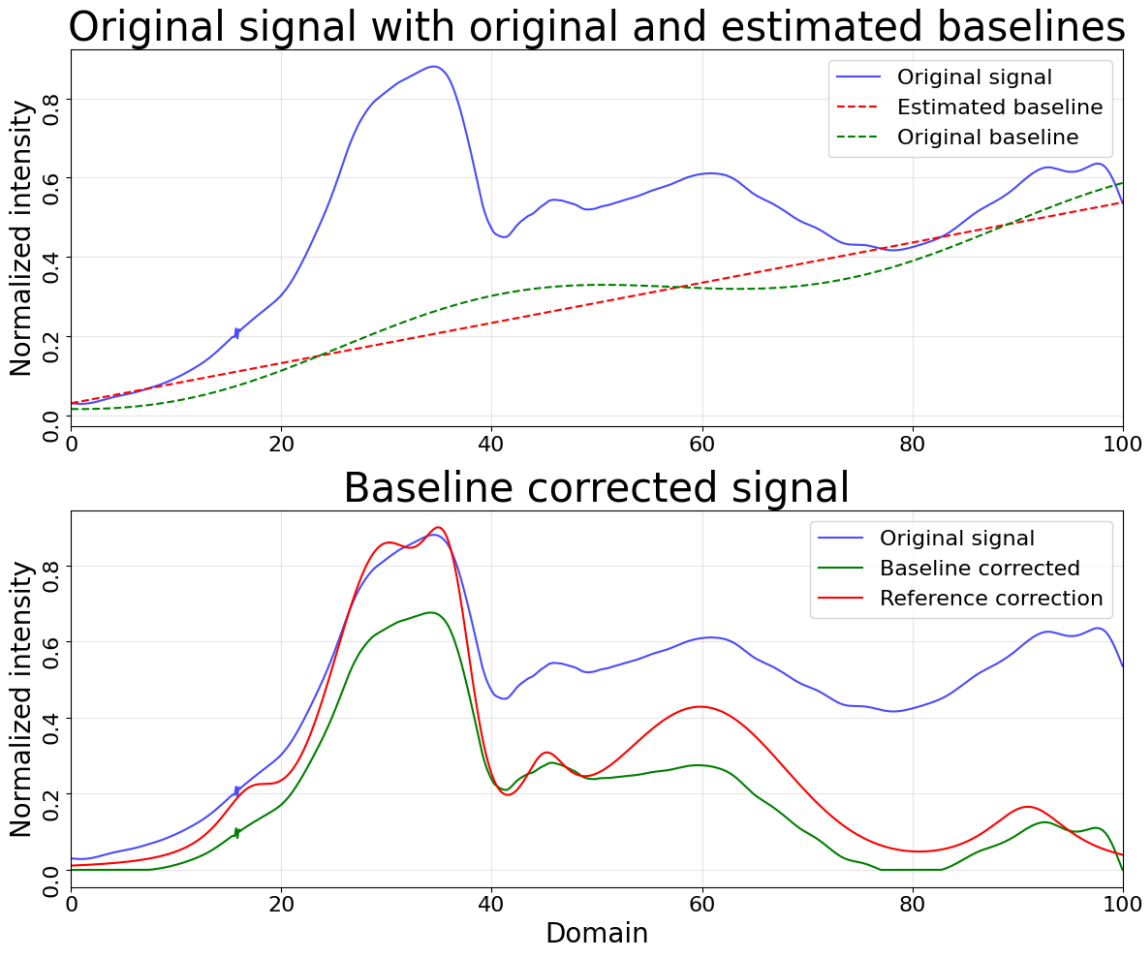


Figure 3.9. *a) The original signal with the predicted linear baseline and the original baseline used to construct the synthetic signal. b) The final corrected result in comparison to the original signal.*

of the initial signal. The points positioned below the predicted baseline are truncated to be zero as the signal is expected to not change the sign, and the definition of the baseline does not allow this. However, the supposedly separate signals in the middle of the sample synthetic signal in Figure 3.9 are not clearly separated. This can wrongly hint in the final signal decomposition stage that there is another distribution or peak hidden in the middle.

Penalised least squares (PLS) algorithms, introduced in detail in Appendix

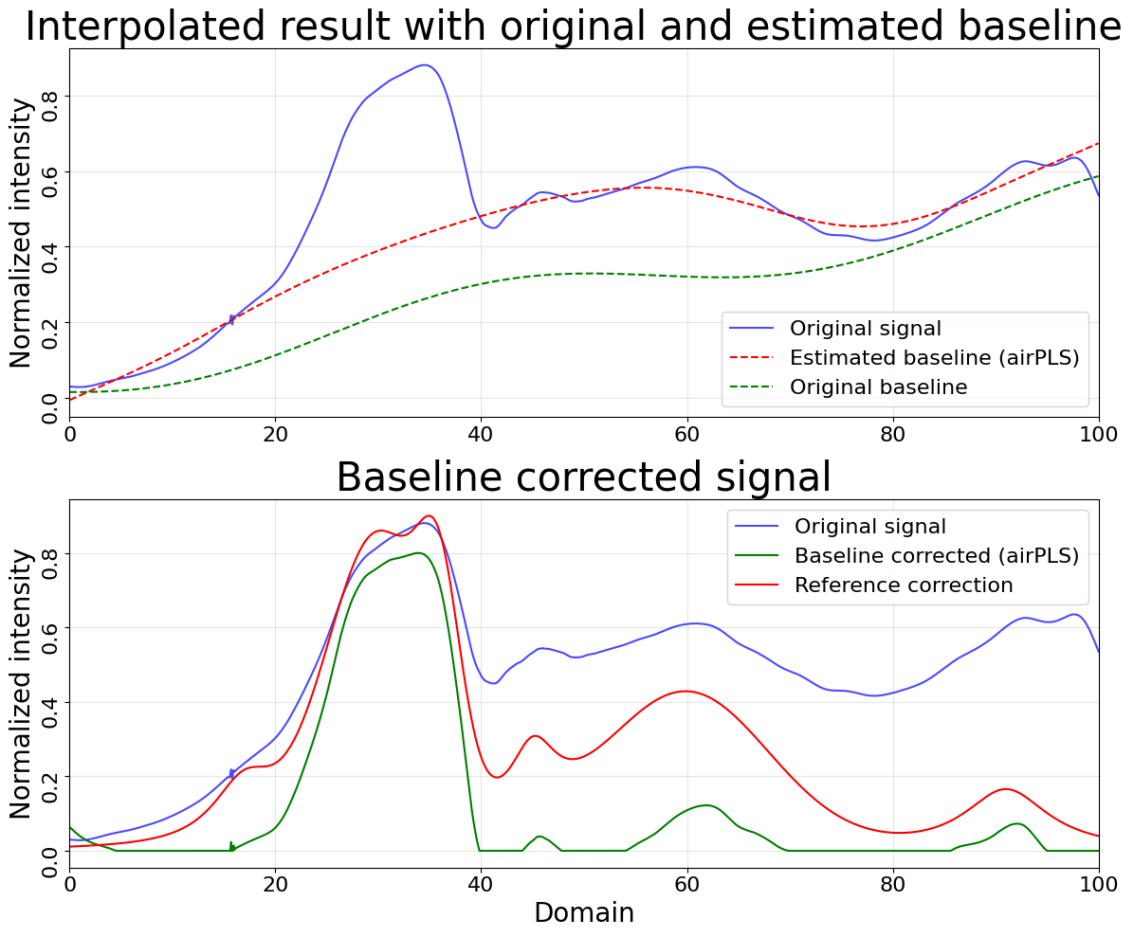


Figure 3.10. *Synthetic signal baseline correction via airPLS method with $\lambda = 10^{15}$. a) The original signal with the predicted airPLS-based baseline and the original baseline used to construct the synthetic signal. b) The final corrected result in comparison to the original signal.*

B.2, are employed to solve this problem in the rest of this section. One popular method is called adaptive iteratively reweighted PLS (airPLS)–introduced by Zhi-Min Zhang, Shan Chen, and Yi-Zeng Liang [77]–which is interested in finding the baseline by minimizing the following quantity:

$$Q = \sum_{i=1}^{n} w_i (s_i - z_i)^2 + \lambda \sum_{i=2}^{n} (z_i - z_{i-1})^2 \tag{3.32}$$

where s_is are the data-points of the i-th component of the signal,

$$\mathbf{s} = (s(x_1), \cdots, s(x_n)), \tag{3.33}$$

to be corrected, and z_is are the corresponding baseline components to be found. The weights are given by the relation below.

$$w_i^t = \begin{cases} e^{t\left(s_i - z_i^{t-1}\right)/|\mathbf{d}^t|} & s_i < z_i^{t-1} \\ 0 & s_i \geqslant z_i^{t-1} \end{cases} \tag{3.34}$$

The superscript t in Equation 3.34 refers to the iteration step in the alogrithm which terminates when

$$|\mathbf{d}_t| < 0.001\|\mathbf{s}\| \tag{3.35}$$

holds. The weights are initialized to be 1 when $t = 1$ and z_is are defined to match this initialization at step $t = 1$. Figure 3.10 demonstrates how the airPLS method (red dashed line) better captures the curved baseline behavior compared to the linear approximation shown in Figure 3.9. The superior baseline estimation results in more accurate peak separation in the corrected signal.

One problem of airPLS is that it is highly likely to position a substantial portion of the signal below the baseline. [1] In Figure 3.10, this effect was useful since it resulted in the removal of the peak generated by the noise contribution at the end of the signal. This fortuitous result should not be taken as a general advantage of the method, as such noise contributions can appear anywhere in the signal, including adjacent to critical peaks.

To overcome such disadvantages, Sung-June Baek, Aaron Park, Young-Jin Ahn and Jaebum Choo, developed a different weighting scheme with

$$w_i = \begin{cases} 1/e^{2(|\mathbf{d}| - (2\sigma_d - \mu_d))/\sigma_d} & s_i \geqslant z_i \\ 1 & s_i \leqslant z_i \end{cases} \tag{3.36}$$

where μ_d and σ_d are mean and standard deviations of the entries of $\mathbf{d}$. [78] The z_is are initialized to be slightly larger than s_is if such scheme can be computed, or set to be 1 in the beginning of the iterative process. The termination

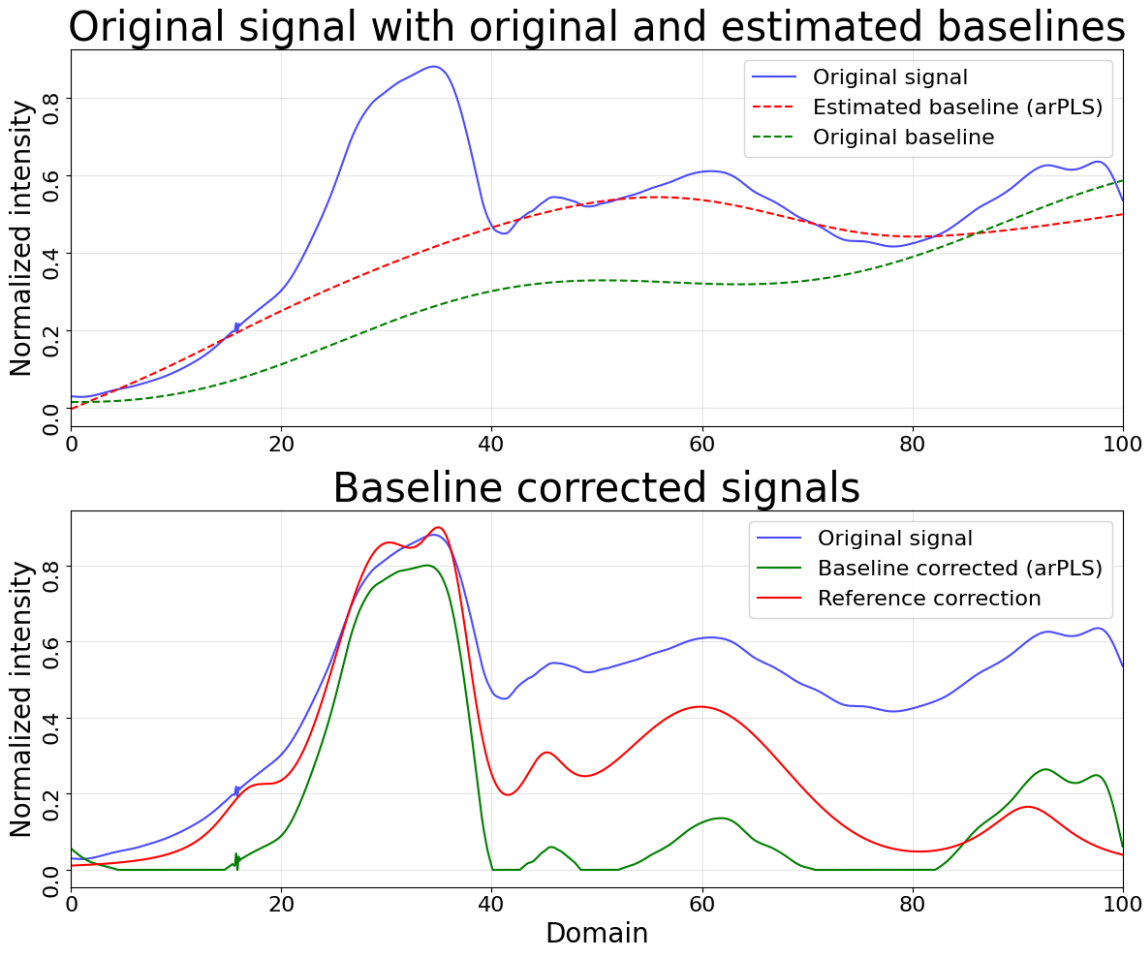


Figure 3.11. *Synthetic signal baseline correction via arPLS method with $\lambda = 10^{15}$ and termination ratio of 10^{-6}. a) The original signal with the predicted arPLS-based baseline and the original baseline used to construct the synthetic signal. b) The final corrected result in comparison to the original signal.*

condition of the algorithm is given as

$$\frac{\sum_i \left(w_i^t - w_i^{t+1}\right)^2}{\sum_i \left(w_i^t\right)^2} < \text{ratio} \tag{3.37}$$

where the ratio can be set by the user to be some very small number.

Figure 3.11 shows two benefits of the arPLS methods:

1. No peak candidates lose the curvature information of the original signal,

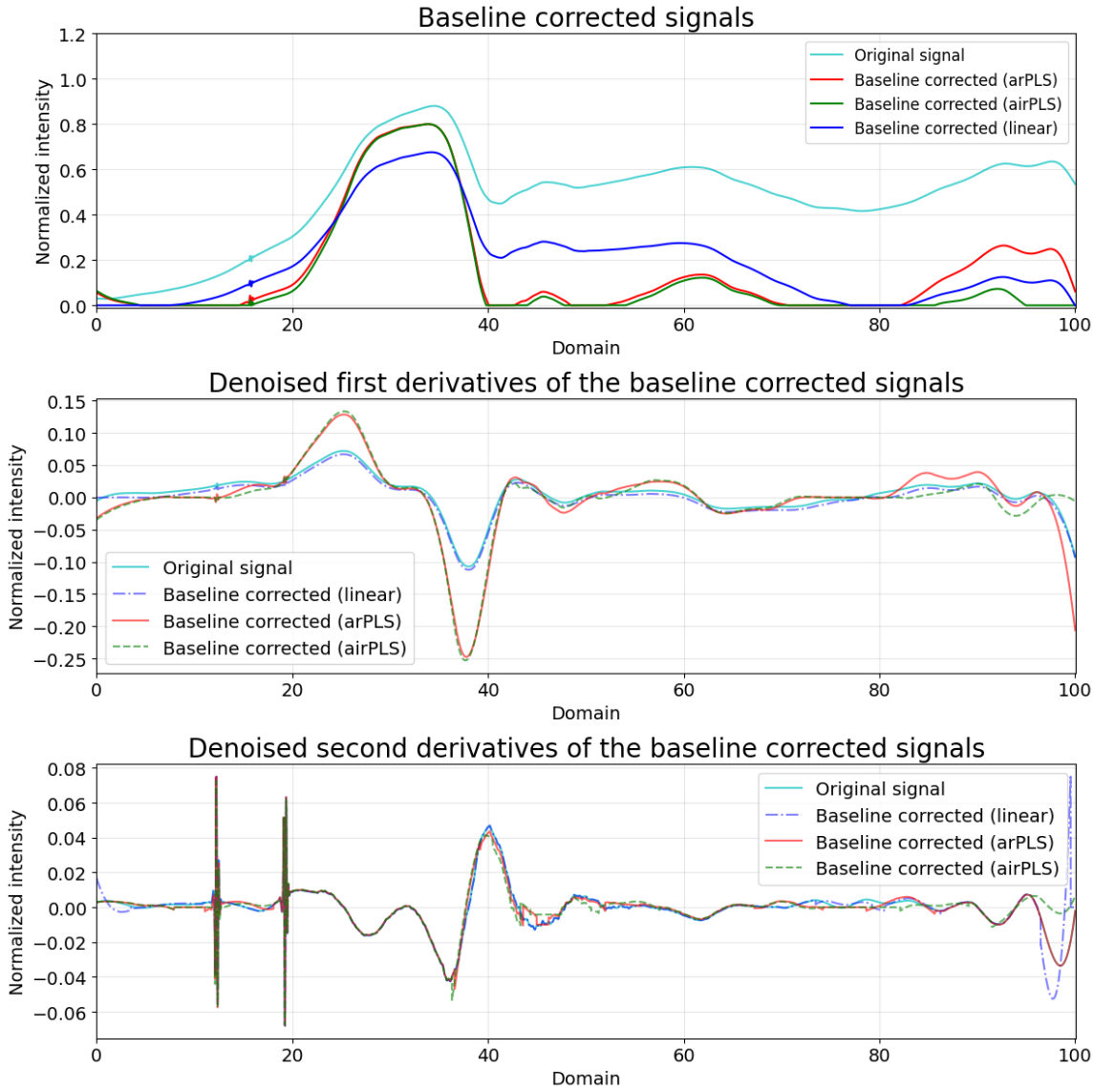


Figure 3.12. *Baseline corrected results and their first and second derivatives. The derivatives are filtered via Savitzky-Golay method with window size of 7001, where there are 10,000 datapoints in the baseline corrected signals and their derivatives.*

and

2. All peaks from the original signal are kept.

Figure 3.12 graphically compares the baseline corrected results obtained via all three methods introduced in this section. The first and second derivatives are shown to illustrate what model among the three methods introduced in this section could be useful in finding the peaks of the signal. With proper denoising, again via the Savitzky-Golay or wavelet transfomation methods, finding the first-derivative's zero-crossings and the negative peaks of the second derivatives becomes manageable due to their smooth profiles. Note that, although doable, the second derivatives of the linearly corrected signal are the noisest among the results of the other two results, making them less reliable.

3.3 Identifying peaks

The identification of peaks in a signal $s(x)$ defined over a domain $[a, b]$ can be formalised through the analysis of local maxima and their surrounding neighbourhoods. A peak at point $x^p \in [a, b]$ can be identified naïvely as a point where $s'(x^p) = 0$ and $s'' < 0$, where these derivatives exist. However, for discrete and noisy signals, a more robust definition is needed. This requires one to mathematically define what a local peak of a discrete signal is.

Definition 3.6 (Local median peak)**.** Let $\{t_i\}_{i=0}^{N}$ be the knots of the domain of a signal s. The *local median peak* of a subinterval $[t_{i-1}, t_i]$ is a pair $\big(\tilde{t}, s\big(\tilde{t}\big)\big)$ such that

$$\tilde{t} = \text{median}\{t : t \in (t_{i-1}, t_i) \cap t_P\} \tag{3.38}$$

such that

$$s\big(\tilde{t}\big) = \max\{s(t) : t \in t \in (t_{i-1}, t_i) \cap t_P\} \tag{3.39}$$

where t_P is a set of dependent variables of the signal given by the dataset.

One can then define the knots as equal-sized, say by the size δ, set of intervals, referred to as *sliding windows* of size δ. As the window slides across the domain by increments of δ, it generates a sequence of candidate peaks that satisfy both the median position and maximum amplitude criteria defined above. To consider all points equally, one can start defining the sliding windows by defining the knots starting from all or a few selected points of $[t_1, t_2) \cap t_P$ with spacings δ. In practice, one can define the sliding step k, as a fraction of the window size δ, such that:

$$k = \delta/m \tag{3.40}$$

where $m \in \mathbb{N}$ is chosen to ensure sufficient overlap between consecutive windows. This creates a sequence of windows $[t_i, t_i + \delta]$ where $t_{i+1} = t_i + k$ for each step i. The overlap between consecutive windows is then $100(m-1)/m$% of the window size.

Theorem 3.7 (Validity of the sliding window method)**.** *If the signal is smooth, the set of local median peaks contains the set of the local maxima of the entire signal*.

Proof. Suppose a point $(t^*, s(t^*))$ is a local maximum of the signal s that can not be defined as a local median peak. Then for all $\delta > 0$, the interval $(t^* - \delta, t^* + \delta)$ contains a pair $\left(t^{\text{faux}}, s\left(t^{\text{faux}}\right)\right)$ such that

$$s\left(t^{\text{faux}}\right) > s(t^*) \tag{3.41}$$

which contradicts the definition of local maximum. □

Note that the smoothness was required to justify the validity of the sliding window method. Since digital signals are discrete, attempting to identify local peaks with all values of $\delta > 0$ would not be feasible. As a result, an acceptance threshold is applied within a given window $[t_i, t_i + \delta]$, as defined in Definition 3.8.

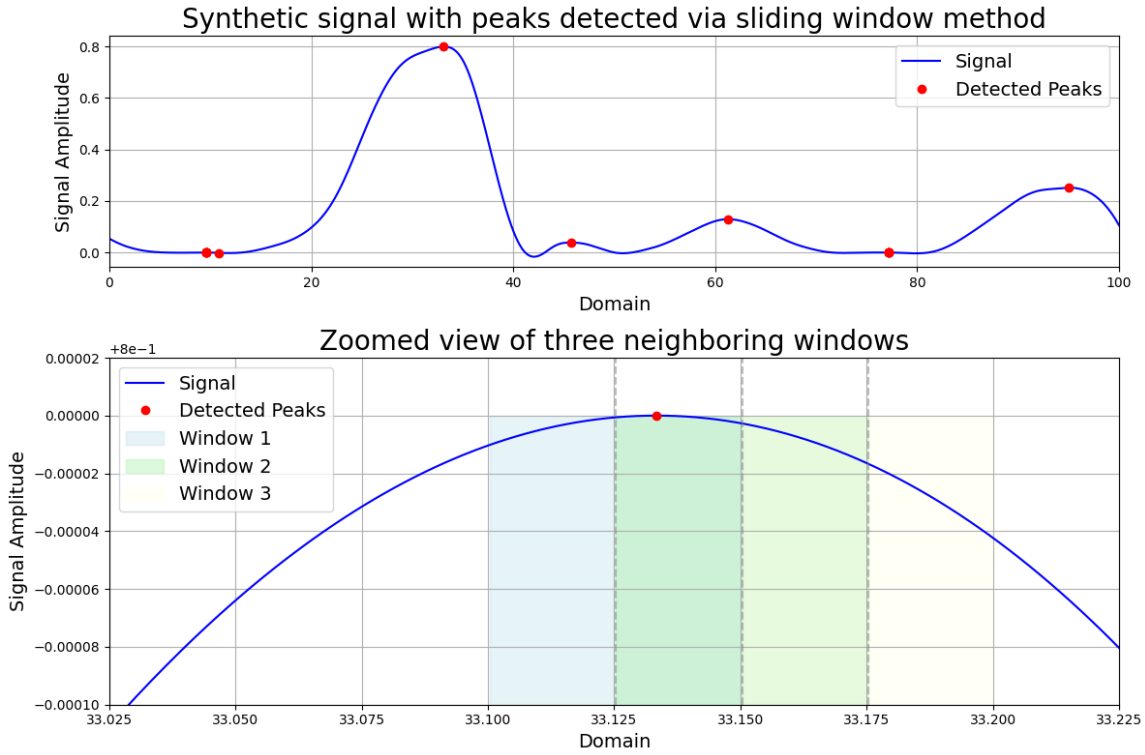


Figure 3.13. *Peak candidates found via the window-propagation method with window-size of 51, overlap factor of 27 (TOP) and visualisation of propagating windows (BOTTOM).*

Definition 3.8 (Acceptance threshold for window propagation method.)**.** If $t^* = \text{median}([t_i, t_i + \delta])$, and $s(t^*) = \max(s([t_i, t_i + \delta]))$, an *acceptance threshold*, $\xi > 0$ is a quantity that identifies t^* as a valid local median peak if

$$\frac{1}{\delta}(s(t^*) - \min(s([t_i, t_i + \delta]))) > \xi \tag{3.42}$$

holds.

Figure 3.13 illustrates how peaks are detected using the window-propagation method. While this approach successfully identifies all the true peaks, it may also report some false peaks. These false positives arise from small, numerically noisy variations near the base of the signal. To address this, a small

amplitude threshold can be applied to filter out these noise-induced peaks.

Figure 3.13 illustrates how the peaks can be found via the window-propagation method. Note that this method identified all the true peaks, but some false positives were reported. These are due to the small numerically noisy profiles near the base of the signal, which can be eliminated quickly by asserting a small threshold in the amplitude. For instance, in the example shown in Figure 3.13, setting a threshold of 2×10^{-3}, leaves the following peak candidates: $[(33.13, 0.8), (45.73, 0.039), (61.28, 0.129), (95, 0.252)]$. Despite the high noise level in the sample signal, these peaks are in close proximity to the centers of the distributions used to sample the original signal $x = 36, 45, 60, 91$. However, the method fails to recover the peaks at $x = 17$ and 30, which correspond to the *shoulders* of the nearby peaks. These shoulders can be detected using first and second derivative-based methods.

Definition 3.9 (Shoulder of a peak)**.** Let $s(x)$ be a signal. A pair $\left(x^{\text{shoulder}}, s\left(x^{\text{shoulder}}\right)\right)$ is called a shoulder of a peak $\left(x^{\text{peak}}, s\left(x^{\text{peak}}\right)\right)$ if it satisfies

$$\left|\frac{d^2 s}{dt^2}\left(x^{\text{shoulder}}\right)\right| < \xi_2 \quad \text{and} \quad \left|\frac{ds}{dt}\left(x^{\text{shoulder}}\right)\right| > \xi_1 \tag{3.43}$$

where $\xi_1 \geqslant 0$ and $\xi_2 \geqslant 2$ are tolerance parameters for the first and second derivatives respectively.

Note that the tolerance parameters can be obtained via statistical properties of the derivatives: ξ_1 can be chosen as a fraction of the standard deviation of the first derivative, while ξ_2 can be set relative to the maximum absolute value of the second derivative in regions where shoulders are expected to occur.

Figure 3.14 demonstrates the application of our shoulder detection method. The analysis identified two shoulders at $x = 24$ and $x = 31$. The shoulder at $x = 31$ closely matches one of the known peak centers in the synthetic signal at $x = 30$. However, the other shoulder at $x = 24$ deviates significantly from

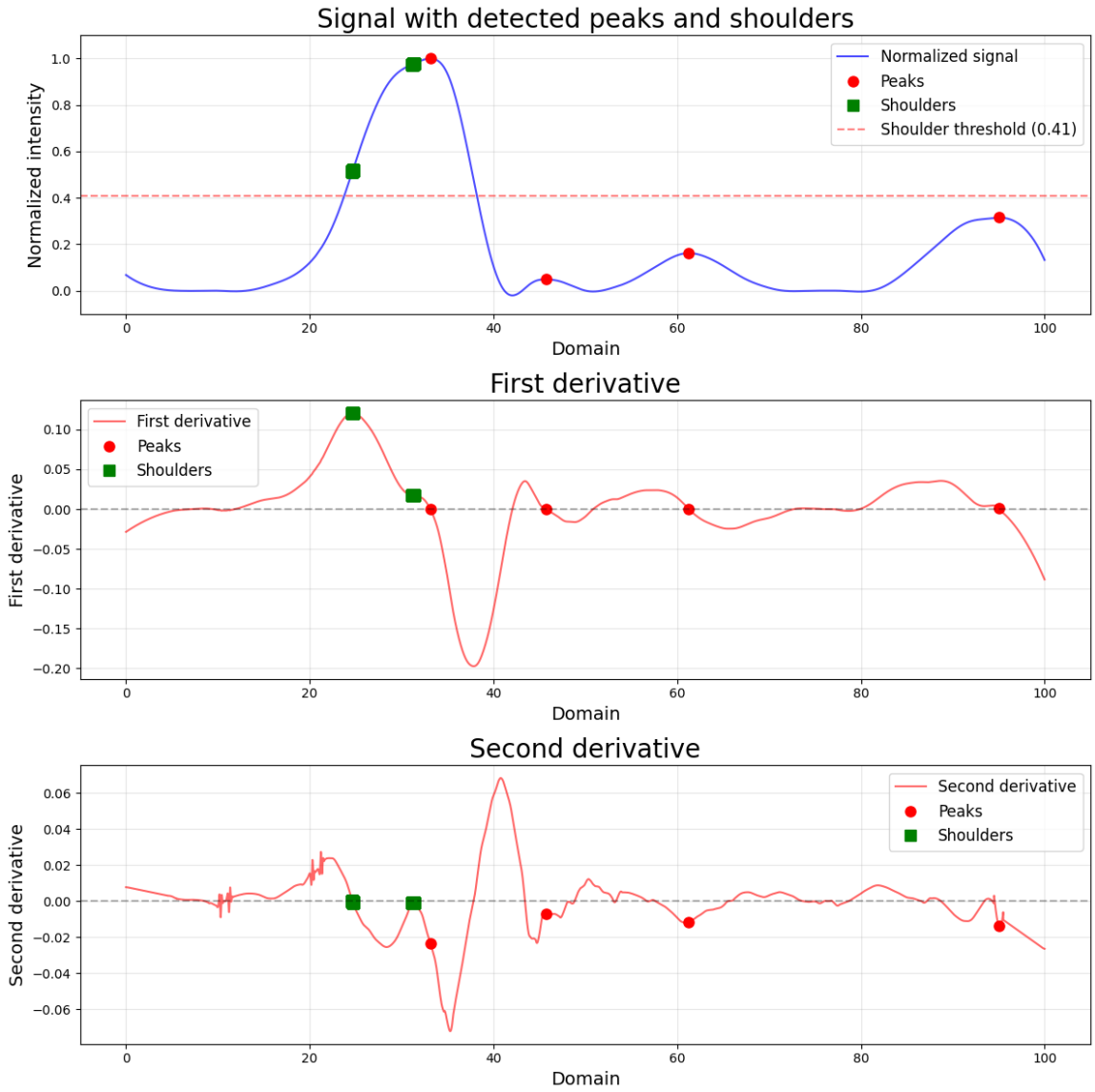


Figure 3.14. *Analysis of peak and shoulder detection in a signal. Top: Normalized signal (blue) with detected peaks (red circles) and shoulders (green squares). The dashed red line indicates the threshold below which shoulders are excluded. Middle: First derivative of the signal showing zero-crossings at peak locations. Bottom: Second derivative with near-zero values indicating shoulder locations. The positions of peaks and shoulders are marked consistently across all three plots to demonstrate their relationship with the derivatives.*

its expected position at $x = 17$. This discrepancy arose because the small shoulder feature at $x = 17$ was largely obscured by noise in the signal, and the subsequent denoising process further diminished its detectability.

3.4 Decomposition of signals

The analysis of spectroscopic signals often requires decomposing an input signal into multiple physically meaningful components. To carry out this decomposition rigorously, the signals are treated within the framework of *tempered distributions*, which provides a mathematical foundation for the analysis of functions that may not exhibit rapid decay at infinity—a characteristic frequently encountered in spectroscopic data.

A tempered distribution is a continuous linear functional on the Schwartz space $\mathcal{S}(\mathbb{R}^n)$. More precisely, let $\mathcal{S}(\mathbb{R}^n)$ be the space of infinitely differentiable functions $\phi : \mathbb{R}^n \to \mathbb{C}$ (called test functions) such that for all multi-indices $\alpha, \beta \in \mathbb{N}^n$:

$$\sup_{x \in \mathbb{R}^n} |x^\alpha D^\beta \phi(x)| < \infty \tag{3.44}$$

where $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ and $D^\beta = \frac{\partial^{|\beta|}}{\partial x_1^{\beta_1} \cdots \partial x_n^{\beta_n}}$.

Then, a tempered distribution $T \in \mathcal{S}'(\mathbb{R}^n)$ is defined as a linear functional $T : \mathcal{S}(\mathbb{R}^n) \to \mathbb{C}$ that is continuous in the following sense: for any sequence $\{\phi_k\}_{k=1}^\infty \subset \mathcal{S}(\mathbb{R}^n)$ converging to 0 in the Schwartz topology, it holds that

$$\lim_{k \to \infty} T(\phi_k) = 0 \tag{3.45}$$

Any locally integrable function f with at most polynomial growth at infinity defines a tempered distribution T_f via:

$$T_f(\phi) = \int_{\mathbb{R}^n} f(x)\phi(x)\, dx, \quad \phi \in \mathcal{S}(\mathbb{R}^n) \tag{3.46}$$

In spectroscopic analysis, measured signals naturally fit into this framework as tempered distributions since they satisfy the growth conditions at infinity and

are locally integrable. To decompose such a signal $f \in \mathcal{S}'(\mathbb{R}^n)$, the objective is to obtain a representation as a finite linear combination of reference functions $\{g_i\}_{i=1}^m \subset \mathcal{S}(\mathbb{R}^n)$:

$$f \approx \sum_{i=1}^{m} c_i g_i \tag{3.47}$$

where the coefficients $\{c_i\}_{i=1}^m$ are determined through least-squares optimization in the L^2 norm. The reference functions introduced in Appendix A are chosen from the Schwartz space $\mathcal{S}(\mathbb{R}^n)$ to ensure they decay rapidly at infinity and possess sufficient smoothness for spectroscopic analysis.

To apply this framework in practice, the locations of significant spectral features in the signal are first identified. Consider a signal s with M identified features (peaks and shoulders) at locations $\{(t_i, s(t_i))\}_{i=1}^M$, found using the methods described in the previous section. The decomposition of this signal into a sum of distributions can be formulated as an optimization problem. More precisely, the task is to find the set of distribution parameters $\{\mathbf{p}_i\}_{i=1}^M$ that minimise the residual given as the following:

$$Q^{\text{dist}} = \sum_{j=1}^{N} \left(s(t_i) - \sum_{i=1}^{M} \mathcal{D}(t_i; \mathbf{p}_j) \right)^2 \tag{3.48}$$

where $\mathcal{D}(t_i; \mathbf{p}_j)$ represents the value of a probability density function $\mathcal{D}$ evaluated at position t_i with parameters $\mathbf{p}_j$.

For instance, if the signal is composed of a sum of Gaussian density functions, Equation 3.48 becomes

$$Q^{\text{Gauss}} = \sum_{j=1}^{N} \left(s(t_j) - \sum_{i=1}^{M} a_i e^{-(t_j-\mu_i)^2/2\sigma_i^2} \right)^2 \tag{3.49}$$

where a_i, μ_i, and σ_i are the amplitude, mean, and standard deviation of the i-th Gaussian component, respectively. The parameters a_i and μ_i are initialised using the peak amplitudes and locations identified in the previous section.

For Lorentzian densities, the residual to optimise becomes

$$Q^{\text{Lorentz}} = \sum_{j=1}^{N} \left(s(t_j) - \sum_{i=1}^{M} a_i \frac{\gamma_i}{(t_j - t_i^0)^2 + {\gamma_i}^2} \right)^2 \tag{3.50}$$

where a_i, t_i^0, and γ_i are the amplitude, median, and median absolute deviation of the i-th Lorentzian component, respectively. The parameters a_i and t_i^0 are initialised using the peak amplitudes and locations identified by the peak-identification step.

For Voigt distributions, which are convolutions of Gaussian and Lorentzian distributions, the optimization becomes more complex due to the additional parameters needed to characterize both the Gaussian and Lorentzian components. Similar to the previous cases, the peak amplitudes and locations from the peak-identification step are used to initialize a_i and the central positions μ_i of each component. However, two width parameters must be initialized: the Gaussian width ω_G and the Lorentzian width ω_L. A common approach is to initially assume equal contributions from both broadening mechanisms by setting $y = \omega_L/\omega_G = 1$ as a starting point for the optimization procedure. The residual to optimise thus becomes

$$Q^{\text{Voigt}} = \sum_{j=1}^{N} \left(s(t_j) - \sum_{i=1}^{M} a_i K\left(\frac{t_i - \mu_i}{\omega_G}, y \right) \right)^2 . \tag{3.51}$$

Figure 3.15 illustrates the decomposition of the signal using three different profile types: Gaussian, Lorentzian, and Voigt distributions, each optimized through the least squares method. The comparative analysis reveals that the Gaussian profile provides the most accurate decomposition of the synthetic signal, capturing both the peak positions and their relative intensities with high fidelity.

All three fitting approaches demonstrate consistent identification of major peaks, with the Gaussian fit exhibiting particularly well-defined peak separations. However, it is noteworthy that none of the methods detected the minor peak at 17 units. This limitation can be attributed to the elevated noise level in this region, which effectively masks the subtle spectral feature. The signal-to-noise ratio in this region falls below the detection threshold necessary for reliable peak identification, regardless of the chosen profile type.

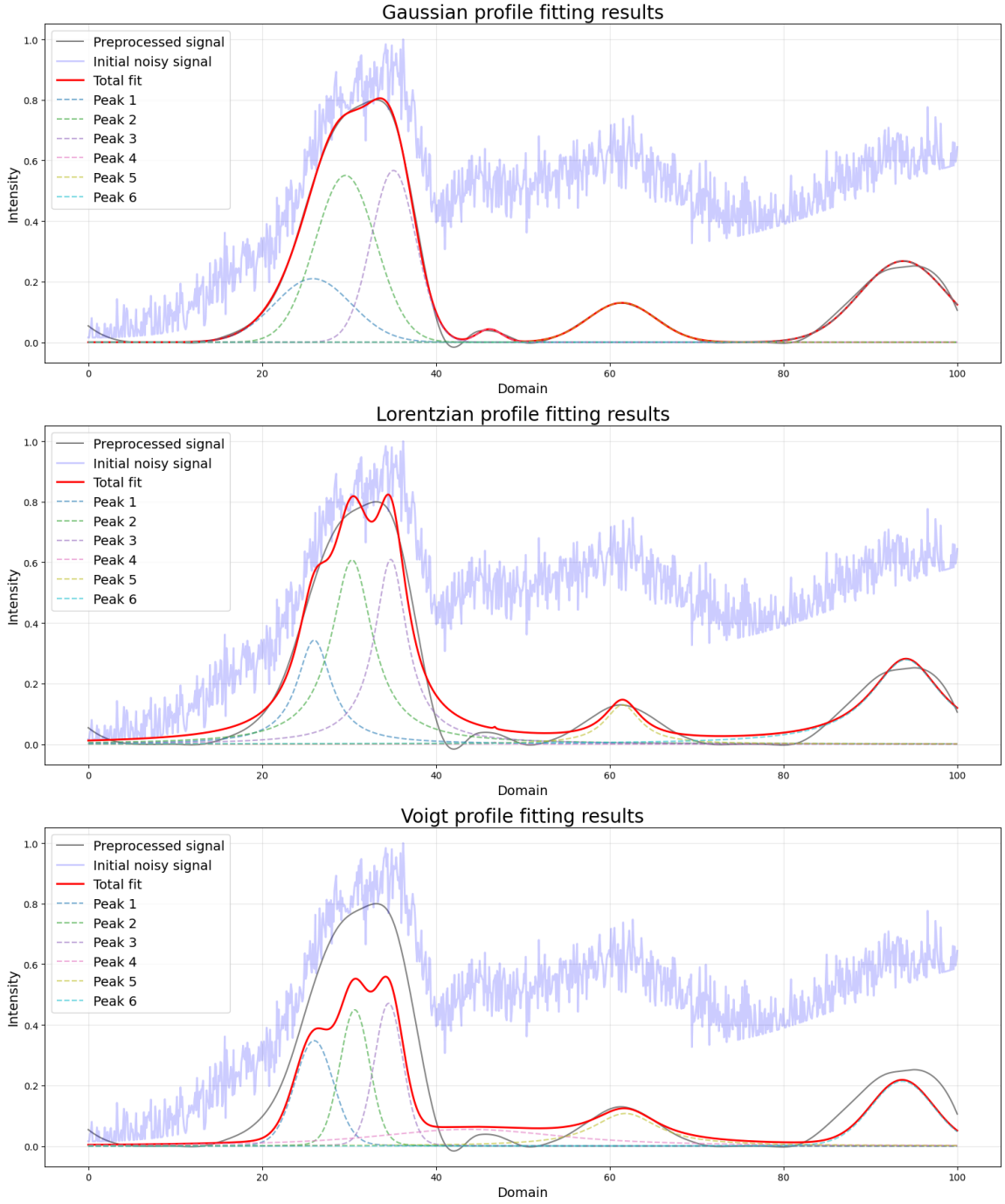


Figure 3.15. *Final Gaussian, Lorentzian and Voigt fitted results.*

3.5 Application to the given vibrational spectral data

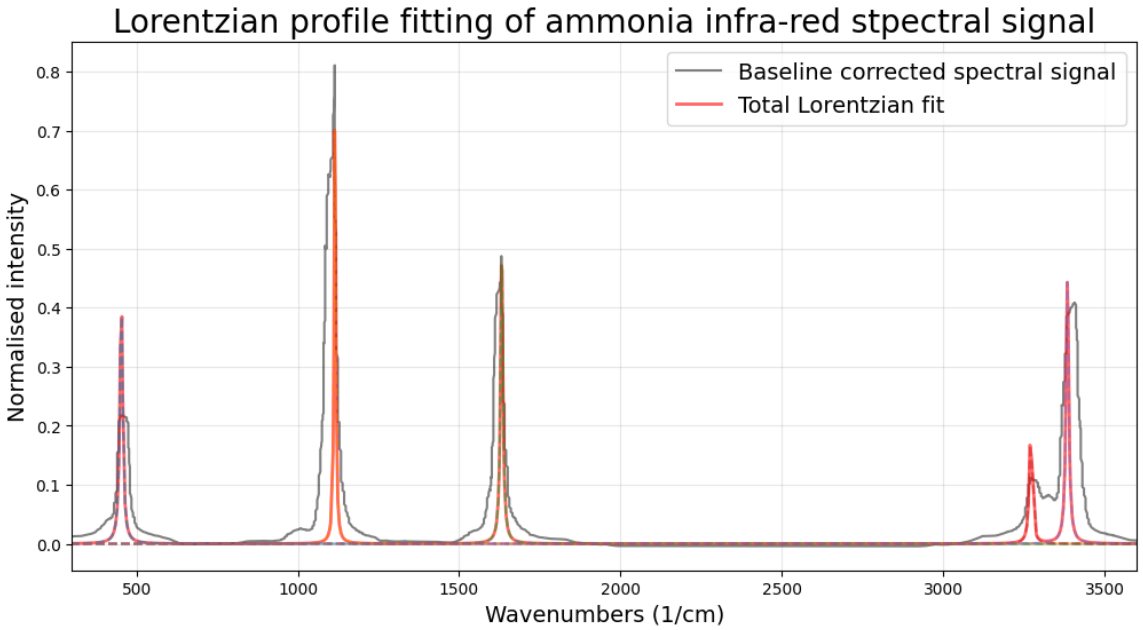


Figure 3.16. *Simulated ammonia infrared spectrum and its Lorentzian decomposition*.

This section illustrates the application of the methods outlined in this chapter for the analysis of vibrational spectral signals derived either from experimental data or computational simulations. First, the infrared spectral signal of ammonia is examined, as computed using the TheSeuSS quantum mechanical calculation package [79]. The package generates the spectrum by broadening each vibrational frequency (corresponding to the identified normal mode frequencies) with a Lorentzian function and summing these distributions. Given this construction method, Lorentzian fitting was naturally employed to decompose the signal back into its constituent distributions.

The spectral peaks exhibited relatively narrow profiles compared to the overall domain size. Given that the data set consisted of 100 points, a window size of 5 was selected for peak detection. The threshold for the peak detection step was set to 10^{-5}, reflecting the significant differences in peak heights.

Additionally, no peaks were filtered based on their magnitudes, owing to both the substantial variation in peak intensities and the cleaness of the signal.

Frequency in sim. data (cm^{-1})	Frequency peaks detected (cm^{-1})
468	467
1112	1111
1631	1631
3287	3286
3405	3412

Table 3.1. *Performance benchmark of the introduced method for simulated IR spectrum of solid ammonia*. *Abbreviation*: *sim.*: *simulated reference data*.

Baseline correction was performed using the arPLS method, not to eliminate the baseline itself, but to enhance the clarity of the signal and facilitate more accurate identification of the maximum spectral intensity. A low smoothing parameter, $\lambda = 200$, was chosen, along with a ratio criterion of 10^{-6}, to achieve optimal baseline adjustment and peak isolation. The Lorentzian decomposition of the given signal is shown in Figure 3.16.

The simulation predicted peaks at 468, 1112, 1631, 3287, and 3405 cm^{-1}, corresponding to lattice mode, symmetric bending, antisymmetric bending, symmetric stretching, and a combination of symmetric stretching and lattice mode vibrations, respectively. Table 3.1 compares the peak information obtained using the method introduced in this chapter with the peak locations provided by the simulation.

As an experimental sample, the Raman spectral signal of abelsonite is analysed. Discovered in the 1970s in Utah, United States, and officially named in 1976 [80], abelsonite was selected due to its distinctive porphyrin structure. This structure exhibits characteristic vibrational modes between 600 and 700 cm^{-1} corresponding to C-H bond vibrations, alongside significant peaks around 1500 cm^{-1}. The signal is further complicated by contributions from nickel-coordinated ligands, producing peaks in the range of 400 to approximately 600 cm^{-1}, as well as complex deformation modes between 700 and 1000 cm^{-1}. [81]

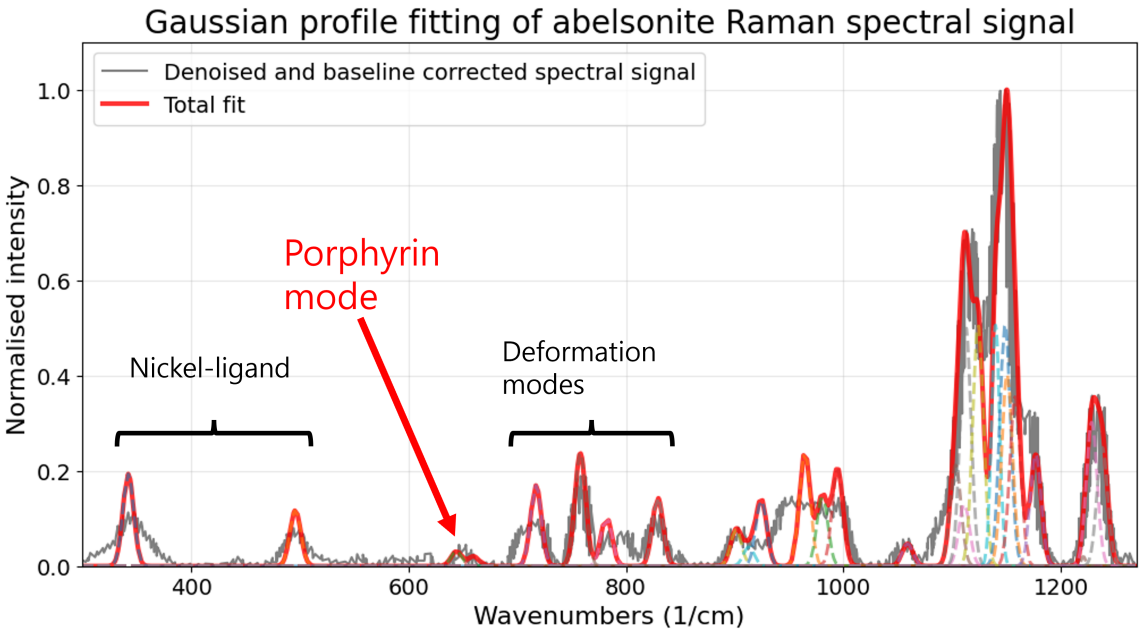


Figure 3.17. *Gaussian decomposition of an abelsonite Raman spectrum from RRUFF datasbase*.

These overlapping features make the spectral output particularly intriguing and challenging, enhancing its suitability for demonstrating the robustness of the proposed methods.

The modified sinc method, with $m = 91$ and a degree of 4, was applied to denoise the signal. Wavelet transformation was not employed since the denoised signal produced by the modified sinc method did not exhibit any irregularities. The resulting signal was then interpolated using cubic splines, generating 10,000 data points. For peak detection, a window size of 10 was used, with a minimum amplitude threshold of 0.01 and a median local peak threshold of 0.05. Baseline correction was deemed unnecessary, as no baseline was detected in the RRUFF dataset from which the signal was obtained. [82]

Figure 3.17 presents the Gaussian-decomposed and fitted Raman spectrum of abelsonite. All physically significant peaks are correctly identified, with a distinct, albeit smaller, peak at 657 cm^{-1} that highlights the porphyrin

structure characteristic of the abelsonite crystal.

CHAPTER 4

Fourier integrator molecular dynamics

當時에 녀던 길을 몇 해를 버려두고
어디 가 다니다가 이제사 돌아온고
이제나 돌아오나니 딴 데 마음 마로리

퇴계 이황 (退溪 李滉)
陶山十二曲 第十曲

Years of pursuits, *I let slip away*
After wandering, *I finally returned*
Returned at last, *I shall remain still*.

TOEGYEO YI HWANG
Twelve Songs from Dosan, *No*. *10*

Molecular dynamics simulations have become an essential tool for understanding molecular behaviour across various time and length scales.[83] Traditional approaches often struggle to directly isolate and analyse specific vibrational modes or frequency bands, particularly when studying molecular spectroscopy. Such capability would be valuable for investigating energy transfer mechanisms [84] and for efficient investigation targeted to specific vibrational modes and frequencies. To address this, the present framework begins with a harmonic ansatz and is subsequently extended to incorporate anharmonic mode couplings, which are then analysed quantitatively. Let A be an arbitrary molecule under study, and let its m^{th}-mode vibrational state as $A(\text{mode } m)$ Energy transfer between the vibrational modes of A, can be analysed by examining the energy difference ΔE_{mode} in a reaction of the form:

$$A(\text{mode } i) + \cdots + A(\text{mode } j) \rightarrow A(\text{mode } k) + \cdots + A(\text{mode } l) + \Delta E_{\text{mode}} \quad (4.1)$$

where the mode index indicates the vibrational mode of the molecule A. [85] To study the energy transfer between modes and mode-specific energy contributions, it is essential to analyse the dynamics starting from specific initial molecular modes. However, this poses a significant challenge, as conventional

molecular dynamics simulations treat all vibrational modes simultaneously and operate in a non-frequency-specific coordinate system, making it difficult to isolate and analyse the behaviour of individual modes.

This chapter introduces a novel approach to molecular dynamics that operates directly in the Fourier space spanned by molecular vibrational mode-specific phase space coordinates. This enables selective study of specific vibrational modes while maintaining approximately the same amplitudes and phase couplings they would have in conventional time-domain molecular dynamics simulations. The method builds upon the mathematical foundation of symplectic integrators while exploiting the natural structure of molecular vibrations in frequency space. This approach offers three key advantages over conventional methods.

1. It enables precise isolation of dynamical components corresponding to specific vibrational modes, allowing detailed analysis of mode-specific behaviour following cues from experiment or from fundamental calculations.
2. It provides a computationally efficient framework for studying vibrational dynamics, reducing the number of energy evaluations required to obtain spectroscopic information in the frequency window of interest.
3. It increases numerical stability even with simple (modified) Verlet-type integrators, as the equations of motion in Fourier space naturally preserve the symplectic structure of the dynamics, potentially allowing larger maximum time-steps when probing lower frequency bands.

The following sections develop this framework systematically, beginning with the mathematical foundations of Fourier integrators, proceeding through their implementation in molecular dynamics, analysing the treatment of harmonic and anharmonic contributions, and concluding with practical applications to molecular systems with nontrivial number of atoms. This progression

aims to demonstrate both the theoretical rigour and the practical utility of this approach to molecular dynamics simulations.

4.1 Integrators on Fourier-expanded variables

Executing molecular dynamics simulations focused on vibrational modes requires fundamental modifications to the integration algorithm to operate directly on modal coordinates. The dynamical evolution is performed on Fourier-transformed variables, necessitating adaptation of the integrator to handle these transformed coordinates effectively. The integration scheme must maintain fidelity to mode-specific dynamics while accurately capturing anharmonic behaviour and inter-mode coupling phenomena beyond the harmonic approximation. These requirements intrinsically demand that the integrators exhibit nonlinear characteristics, possess mode specificity, and operate directly within the Fourier-expanded coordinate space.

For simplicity, consider first a 1-dimensional problem with phase-space coordinates given by position $r = r(t)$ and momentum $p = p(t)$. The position variable can then be expanded in a discrete Fourier series

$$r(t) = \sum_k r_k e^{i\omega_k t} \tag{4.2}$$

for some amplitude variables r_k and frequencies ω_k. Given a Lagrangian $\mathcal{L} = \mathcal{L}(r, p)$ of the problem, the change in momentum variable in the Fourier-space can be written as:

$$\dot{p} = \frac{\delta \mathcal{L}}{\delta r} = \sum_k \frac{\partial \mathcal{L}}{\partial r_k}\frac{\partial r_k}{\partial r} = \sum_k e^{-i\omega_k t}\frac{\partial \mathcal{L}}{\partial r_k}. \tag{4.3}$$

Recalling the Liouville operator formulation of velocity Verlet introduced in Chapter 2.1, the radial Liouville operator given in Definition 2.3 of the one-dimensional problem can be written as

$$\hat{L}_r f = \left[\sum_k (\dot{r}_k + i r_k \omega_k) e^{i\omega_k t}\right] \sum_j e^{-i\omega_j t}\frac{\partial f}{\partial r_j} = \sum_{k,j} e^{i(\omega_k - \omega_j)t}(\dot{r}_k + i r_k \omega_k)\frac{\partial f}{\partial r_j} \tag{4.4}$$

for an arbitrary function f defined over the phase space. Due to the nature of the Fourier expansion, the following lemma holds.

Lemma 4.1. $\left\{\hat{L}_r, \hat{L}_r\right\} = 0$, *where* $\{\cdot,\cdot\}$ *is the Poisson bracket*.

Proof. By writing $\hat{L}_r = \sum_{k,j} W_{k,j} \frac{\partial}{\partial r_j} = W_{k,j} \frac{\partial}{\partial r_j}$, one can immediately see that

$$\begin{aligned}
\left\{\hat{L}_r, \hat{L}_r\right\} f &= \frac{\partial \hat{L}_r}{\partial r_\ell}\frac{\partial \hat{L}_r}{\partial p_\ell} f - \frac{\partial \hat{L}_r}{\partial p_\ell}\frac{\partial \hat{L}_r}{\partial r_\ell} f \\
&= \sum_{k,j,\ell} \left(\frac{\partial W_{k,j}}{\partial r_\ell}\frac{\partial}{\partial r_j} + W_{k,j}\frac{\partial^2}{\partial r_\ell \partial r_j}\right)\left(\frac{\partial W_{k,j}}{\partial p_\ell}\frac{\partial}{\partial r_j} + W_{k,j}\frac{\partial^2}{\partial p_\ell \partial r_j}\right) f \\
&- \sum_{k,j,\ell} \left(\frac{\partial W_{k,j}}{\partial p_\ell}\frac{\partial}{\partial r_j} + W_{k,j}\frac{\partial^2}{\partial p_\ell \partial r_j}\right)\left(\frac{\partial W_{k,j}}{\partial r_\ell}\frac{\partial}{\partial r_j} + W_{k,j}\frac{\partial^2}{\partial r_\ell \partial r_j}\right) f \\
&= 0
\end{aligned} \tag{4.5}$$

as stated. Note that this statement holds only if

$$\frac{\partial^\ell f}{\partial r_{a_1} \cdots \partial r_{a_m} \partial p_{b_1} \cdots \partial p_{b_n}} = \frac{\partial^\ell f}{\partial r_{\sigma(a_1)} \cdots \partial r_{\sigma(a_m)} \partial p_{\sigma(b_1)} \cdots \partial p_{\sigma(b_n)}} \tag{4.6}$$

for any permutation on the indices, σ, and $m + n = \ell \in \mathbb{N}$. This statement holds because the Fourier space basis functions form a linearly independent set, meaning that any derivative with respect to the coordinates

$$r_{a_1}, \cdots, r_{a_m}, p_{b_1}, \cdots, p_{b_n} \tag{4.7}$$

can be expressed as a sum over all possible permutations of these indices, each corresponding to a direct combination of basis sets. [86] □

Lemma 4.1 then immediately gives the following for some phase-space function $f(r,p)$:

$$e^{\hat{L}_r \Delta t} f(r,p) = \sum_{n=0}^{\infty} \sum_{k,j=-\infty}^{\infty} \frac{1}{n!} e^{in(\omega_k - \omega_j)\Delta t} (\dot{r}_k + i\omega_k)^n \Delta t^n \frac{\partial^n f}{\partial {r_j}^n} \tag{4.8}$$

Similarly, as

$$\hat{L}_p f(r,p) = \sum_k e^{-i\omega_k t} \frac{\partial \mathcal{L}}{\partial r_k}\frac{\partial f}{\partial p_k} = \sum_{k,j} e^{-i\omega_k t} \frac{\partial r_j}{\partial p_k}\frac{\partial \mathcal{L}}{\partial r_k}\frac{\partial f}{\partial r_j} \tag{4.9}$$

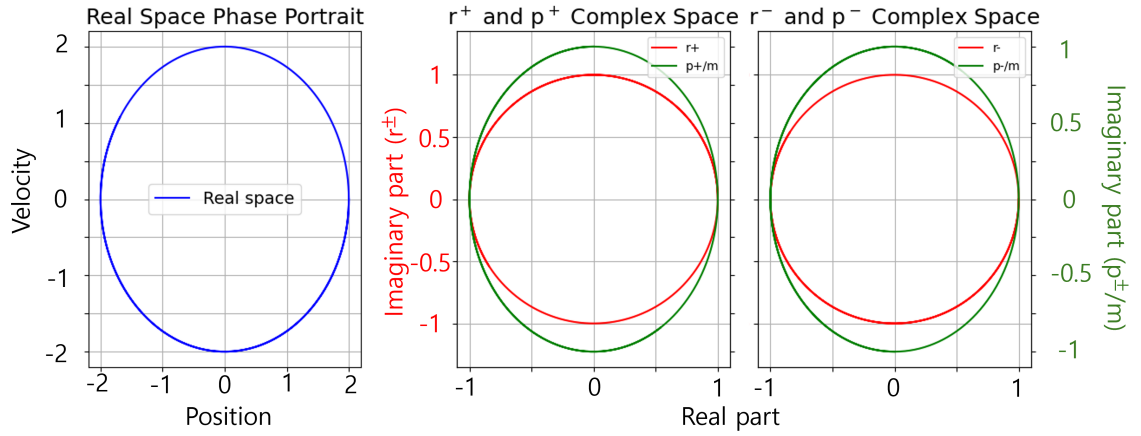


Figure 4.1. *Trajectory information of a simple harmonic oscillator obtained by the Fourier integrators. All parameters, A, ω, and m are set to 1. a) Phase portrait of the simple harmonic oscillator. b) Real and imaginary trajectories of position and momentum of the + mode. c) Real and imaginary trajectories of position and momentum of the - mode. The y-axes for $r^{\pm}$ (red) and $p^{\pm}/m$ (green) are labelled differently so that the two circles, which otherwise overlap, can be clearly distinguished.*

the half time-step on $f(r,p)$ can be written as

$$e^{\hat{L}_p \Delta t/2} f(p) = \sum_n^{\infty} \sum_k \frac{1}{n!} e^{-in\omega_k \Delta t} \left(\frac{\partial r_j}{\partial p_k} \frac{\partial \mathcal{L}}{\partial r_k} \right)^n \frac{\partial^n f}{\partial {r_j}^n} \frac{\Delta t^n}{2^n}. \tag{4.10}$$

The quantity $\partial r_j / \partial p_k$ can be obtained by the Hamilton's equations of motion. As shown in Lemma 4.1, the following can be shown to be true:

$$\left\{ \hat{L}_p, \hat{L}_p \right\} = 0. \tag{4.11}$$

In general, $\left\{ \hat{L}_p, \hat{L}_r \right\} \neq 0$. With these information, it becomes immediate that one can use Trotter's theorem to obtain the velocity Verlet integrators in Fourier basis as shown in Chapter 2.1.2.

4.1.1 Simple harmonic oscillator and Fourier integrators

Consider a one-dimensional harmonic oscillator with single vibrational mode. The position variable can be written as

$$r(t) = r_+ + r_- = Ae^{-i\omega t} + Ae^{i\omega t} = 2A\cos\omega t \tag{4.12}$$

where A and ω are the amplitude and frequency, respectively. For all continuously differentiable f, the positional Liouville operator reads

$$\hat{L}_r[f] = i\omega\left(\frac{\partial f}{\partial r_+} - \frac{\partial f}{\partial r_-}\right)\left(Ae^{i\omega t} - Ae^{-3i\omega t} - e^{-2\omega i t} + 1\right) \tag{4.13}$$

which yields

$$\begin{aligned} \hat{L}_r[r_\pm] &= \pm i\omega A\left[e^{i\omega t} - e^{-3i\omega t} - e^{-2i\omega t} + 1\right] \\ \implies r_\pm(t+\Delta t) &\approx r_\pm(t) \pm i\omega A\left[e^{i\omega \Delta t} - e^{-3i\omega \Delta t} - e^{-2i\omega \Delta t} + 1\right] \end{aligned} \tag{4.14}$$

where the single time-step evolution is obtained by keeping the lowest term in Equation 4.8. With α as the Hooke's constant,

$$\frac{\partial \mathcal{L}}{\partial r_\pm} = -\alpha(Ae^{i\omega t} + Ae^{-i\omega t}) \tag{4.15}$$

and from $p_\pm = \pm mi\omega Ae^{\pm i\omega t}$,

$$\begin{aligned} \frac{\partial r_+}{\partial p_+} &= \frac{1}{mi\omega}, \quad \frac{\partial r_+}{\partial p_-} = 0 \\ \frac{\partial r_-}{\partial p_+} &= 0, \quad \frac{\partial r_-}{\partial p_-} = -\frac{1}{mi\omega} \end{aligned} \tag{4.16}$$

The half time-step evolution in momenta is thus

$$e^{\hat{L}_p \Delta t/2} p_\pm \approx p_\pm \mp \frac{\Delta t}{2}\frac{i\alpha}{m\omega}\left[e^{-i\omega\Delta t/2}(Ae^{i\omega t} + Ae^{-i\omega t}) + e^{i\omega\Delta t/2}(Ae^{i\omega t} + Ae^{-i\omega t})\right]. \tag{4.17}$$

Note that, if this is done for a molecular system, the force field parameters are enter via the Hooke's constant, α.

Figure 4.1 shows the real phase space portrait and the mode trajectory information in complex plane. The real phase space portrait is circular as

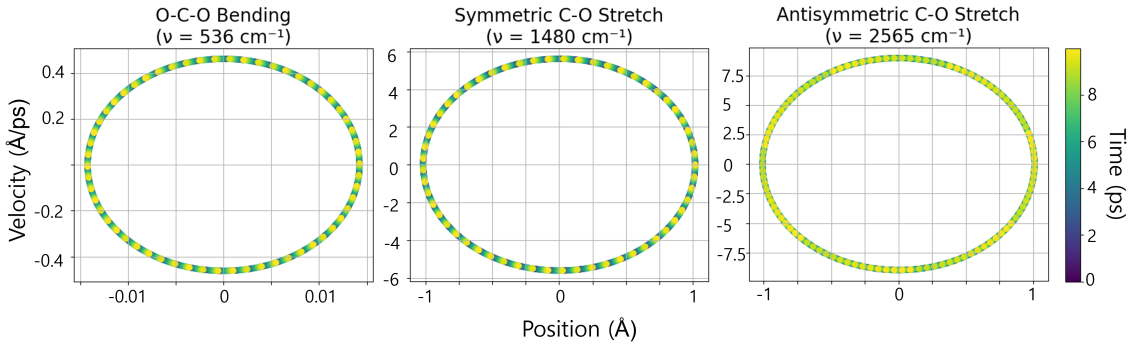


Figure 4.2. *Phase space trajectories of carbon dioxide vibrational modes*: *a*) *O-C-O bond angle vibrational mode*, *b*) *C-O symmetric bond vibrational mode*, *and c*) *C-O anti-symmetric vibrational mode*.

expected by traditionally solved simple harmonic oscillator, showing that the method works perfectly for simple harmonic oscillators.

To facilitate a deeper understanding of the theoretical framework, the Fourier integrator was employed to simulate the dynamics of carbon dioxide. As discussed in Section 2.2, carbon dioxide exhibits three primary vibrational modes: the symmetric stretching of the C-O bonds, the asymmetric stretching of the C-O bonds, and the bending of the O-C-O angle.

The frequencies and normal modes were manually encoded based on known experimental values for these modes. The initial conditions for the simulation were set with positions determined from the equilibrium positions plus random values in the range of $(0, 0.1)$ Å, while velocities were initialised with random values in the range of $(0, 0.01)$ Å/ps. Figure 4.2 demonstrates that the phase-space trajectories of these vibrational modes closely resemble the phase-space diagrams of simple harmonic oscillators. This outcome is anticipated, as the integrators operate independently on each mode, assuming harmonic motion for each vibrational mode.

The evolution of the bond lengths and bond angle is illustrated in Figure 4.3. As expected, the bond lengths oscillate around their equilibrium values,

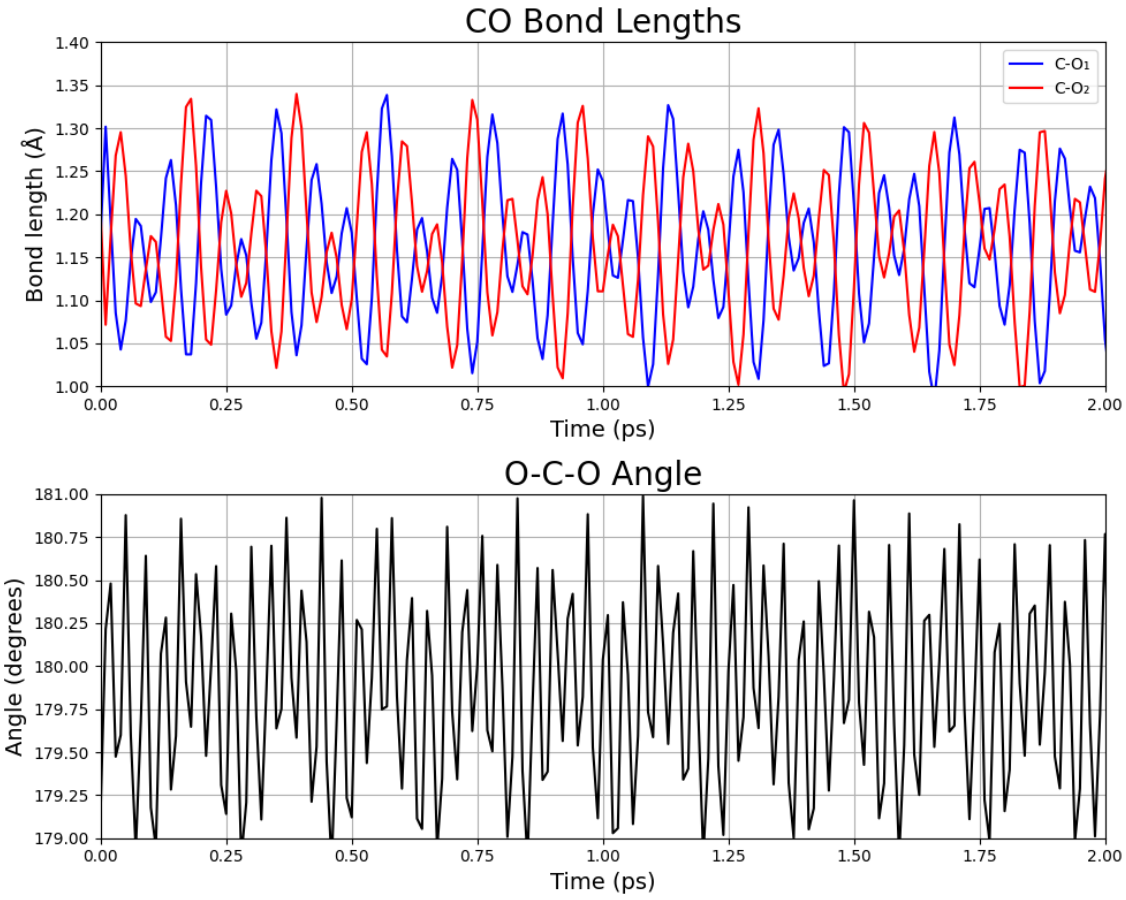


Figure 4.3. *First two picoseconds of changes in a) bond-length and b) bond-angle for an artificial all-harmonic potential of carbon dioxide. The Hamiltonian of the system is Equation* (2.4).

while the bond angle fluctuates around 180 degrees.

The observations from Figures 4.1 and 4.2 immediately allow us to hypothesize that the following theorem holds.

> **Theorem 4.2.** *Let $\mathcal{P}$ be the phase space of an n-body system and denote $\mathcal{P}^{n^v}$ as the subspace of $\mathcal{P}$ with n^v distinct vibrational modes obtained by the Fourier integrator defined in this section. Suppose the r_k are time-independent in $r(t) = \sum_k r_k e^{i\omega_k t}$. The space $\mathcal{P}^{n^v}$ is isomorphic to the product space of n^v number of*

circles, S^1.

$$\mathcal{P}^{n^v} \cong \bigtimes_{i=1}^{n^v} S^1 \tag{4.18}$$

Proof. This statement is proved inductively. For the case where $n^v = 1$, Equation 4.8 becomes

$$e^{\hat{L}_r \Delta t} r_1 = \sum_{n=0}^{\infty} \frac{1}{n!} (i r_1 \omega_1)^n \Delta t^n = e^{i r_1 \omega_1 \Delta t} \tag{4.19}$$

for some $r_1, \omega_1 \in \mathbb{R}$, and Equation 4.10 becomes

$$e^{\hat{L}_p \Delta t/2} p_1 = e^{i\omega_1 \Delta t/2} \sum_{n=0}^{\infty} \frac{1}{n!} (i r_1 \omega_1)^n \frac{\partial^{n-1} p_1}{\partial {r_1}^{n-1}} \frac{\Delta t^n}{2^n} = e^{i\omega_1 \frac{\Delta t}{2} \frac{\partial}{\partial r_1}} p_1. \tag{4.20}$$

As $\Delta t \approx 0$ is expected, letting $\Delta t \to 0$ gives

$$\begin{aligned} \frac{dr_1}{dt} &= i r_1 \omega_1 \\ \frac{dp_1}{dt} &= i p_1 \omega_1 \end{aligned} \tag{4.21}$$

which gives

$$\begin{aligned} \frac{d}{dt}(|r_1|^2 + |p_1|^2) &= \frac{d}{dt}(r_1 \bar{r_1} + p_1 \bar{p_1}) \\ &= r_1 \frac{d\bar{r_1}}{dt} + \bar{r_1} \frac{dr_1}{dt} + p_1 \frac{d\bar{p_1}}{dt} + \bar{p_1} \frac{dp_1}{dt} \\ &= r_1(-i\omega_1 \bar{r_1}) + \bar{r_1}(i\omega_1 r_1) + p_1(-i\omega_1 \bar{p_1}) + \bar{p_1}(i\omega_1 p_1) \\ &= 0 \end{aligned} \tag{4.22}$$

Hence, the phase space trajectory is expected to give a circular motion for $n^v = 1$. Now, suppose that the statement holds up to the case with $n^v - 1$ normal modes, i.e. $\mathcal{P}^{n^v - 1} \cong \times_{i=1}^{n^v - 1} S^1$. Adding another pair of modes (r_{n^v}, p_{n^v}), the phase space becomes $\mathcal{P}^{n^v} = \mathcal{P}^{n^v - 1} \times \mathcal{P}^1$. From Lemma 4.1, the Liouville operators commute, hence the evolution of the n^v-th mode is independent of the previous $n^v - 1$ modes. Therefore, by the same argument as the base case, the additional mode traces a circular trajectory in its phase space, giving $\mathcal{P}^1 \cong S^1$.

The total phase space can then be written as

$$\mathcal{P}^{n^v} = \mathcal{P}^{n^v-1} \times \mathcal{P}^1 \cong \left(\bigtimes_{i=1}^{n^v-1} S^1 \right) \times S^1 = \bigtimes_{i=1}^{n^v} S^1 \tag{4.23}$$

where the first isomorphism follows from the inductive hypothesis and the base case. Hence, by induction, the statement holds for all $n^v \in \mathbb{N}$. □

While the Fourier integrator successfully captures harmonic vibrational motions, as demonstrated in the CO_2 example, real molecular systems often exhibit more complex dynamical behaviours. The phase space trajectories shown in Figures 4.2 and 4.3 represent idealized harmonic oscillations, but deviations from this behaviour can occur due to various physical effects. This observation naturally leads us to examine how the Fourier integration framework extends to non-harmonic molecular motions.

4.2 Molecular dynamics with Fourier integrators

The implementation of molecular dynamics simulations in Fourier space requires modified integration schemes that can simultaneously handle both harmonic oscillations and anharmonic effects. This formulation enables selective investigation of specific frequency bands while maintaining the symplectic nature of the dynamics, ultimately allowing for accurate computation of vibrational spectra and mode-specific molecular motion.

In general, molecular motion comprises both vibrational and non-vibrational components. The latter, non-quadratic anharmonic contributions to the vibrational modes, exist outside the space $\mathcal{P}^{n^v}$, even for a molecule with n^v vibrational modes, as the anharmonic contributions are not on the tori describing the purely harmonic modes. This follows directly from Theorem 4.2. Anharmonic motions arise when molecular vibrations deviate from ideal harmonic oscillator behaviour, typically occurring at higher energies where the potential energy surface can no longer be approximated by a quadratic function.

These anharmonic effects manifest through mode coupling and combination frequencies in the Fourier spectrum.

The separation of harmonic and anharmonic contributions in the molecular Lagrangian is possible in a perturbative treatment of anharmonic effects. Near the equilibrium configuration, the potential energy surface can be expanded in a Taylor series around the minimum:

$$V(\mathbf{r}) = V_0 + \sum_i \left(\frac{\partial V}{\partial r_i}\right)_0 r_i + \frac{1}{2}\sum_{i,j}\left(\frac{\partial^2 V}{\partial r_i \partial r_j}\right)_0 r_i r_j + \frac{1}{6}\sum_{i,j,k}\left(\frac{\partial^3 V}{\partial r_i \partial r_j \partial r_k}\right)_0 r_i r_j r_k + \cdots \quad (4.24)$$

where the subscript 0 denotes evaluation at the equilibrium geometry. The first derivative terms vanish at the minimum, the quadratic terms give rise to harmonic vibrations, and the higher-order terms represent anharmonic corrections.

In Fourier space, these anharmonic terms introduce coupling between different modes. For instance, the cubic term transforms as

$$r_i r_j r_k = \sum_{n,m,l} r_{i,n} r_{j,m} r_{k,l} e^{i(\omega_n + \omega_m + \omega_l)t} \quad (4.25)$$

leading to a natural decomposition of the Lagrangian into

$$\mathcal{L} = \mathcal{L}^{\text{vib}} + \mathcal{L}^{\text{non-vib}} \quad (4.26)$$

where $\mathcal{L}^{\text{vib}}$ describes the harmonic vibrational motions (containing terms up to quadratic order), and $\mathcal{L}^{\text{non-vib}}$ encompasses all other contributions including the mode-coupling terms.

For the purely vibrational components, which are inherently conservative, the Euler-Lagrange equation for $\mathcal{L}^{\text{vib}}$ follows the standard form:

$$\frac{\partial \mathcal{L}^{\text{vib}}}{\partial r_\mu} - \frac{d}{dt}\frac{\partial \mathcal{L}^{\text{vib}}}{\partial \dot{r}_\mu} = 0, \quad (4.27)$$

where $\{r_\mu\}_\mu$ are the generalized coordinates in Fourier space.

However, when including anharmonic contributions, the Euler-Lagrange equation becomes modified by additional source terms:

$$\frac{\partial \mathcal{L}^{\text{non-vib}}}{\partial r_\mu} - \frac{d}{dt}\frac{\partial \mathcal{L}^{\text{non-vib}}}{\partial \dot{r}_\mu} = S(r_\mu, \dot{r}_\mu), \quad (4.28)$$

where the source term S can be expanded in Fourier series as

$$S(r_\mu, \dot{r}_\mu) = \sum_k S_k e^{i\omega_k t} + \sum_{k,l} S_{k,l} e^{i(\omega_k+\omega_l)t} + \sum_{k,l,m} S_{k,l,m} e^{i(\omega_k+\omega_l+\omega_m)t} + \cdots \tag{4.29}$$

with coefficients S_k, $S_{k,l}$, and $S_{k,l,m}$ determined by the specific form of the anharmonic potential.

The source term can be computed by transforming the anharmonic components of the Lagrangian into Fourier space. This is achieved through a two-step process: first, isolating the non-harmonic terms by subtracting $\mathcal{L}^{\text{vib}}$ from the total Lagrangian, then projecting these terms onto the Fourier basis using the mode frequencies ω_k.

While the theoretical framework establishes the basis for molecular dynamics in Fourier space, practical implementation requires careful numerical considerations. The step-by-step procedure is formalized in Algorithm 5.

The algorithm efficiently updates normal modes through subspace iteration, projecting onto previous modes to avoid full Hessian re-diagonalization. While the initial diagonalization can be computationally expensive for large molecules, subsequent updates become more efficient by leveraging existing vibrational modes and the Fourier basis, reducing the cost of force propagation to inexpensive mode-wise scaling operations instead of repeated full Hessian re-diagonalizations.

The algorithm implements the modified velocity Verlet integration scheme in the Fourier basis (Section 4.1), separating harmonic motion in the normal mode space from anharmonic contributions through the source term. A Langevin thermostat with friction coefficient $\gamma = 0.1/\text{ps}$ maintains temperature control, where ξ represents complex Gaussian white noise with zero mean and unit variance. The Langevin thermostat was used since it ensures accurate temperature regulation while introducing local stochastic forces and dissipative damping, enabling realistic thermal fluctuations without significantly disrupting the intrinsic vibrational dynamics. For large amplitude motions that exceed an energy threshold, the algorithm updates the reference geometry

Algorithm 5 Fourier Basis Molecular Dynamics

Initialize reference geometry $\mathbf{r}_0$ and compute Hessian $\mathbf{H}$
Obtain the eigenmodes, $\{v_k\}_k$ of mass-weighted Hessian $\mathbf{M}^{-1/2}\mathbf{H}\mathbf{M}^{-1/2}$
Initialize mode coordinates r_k and momenta p_k with thermal energy k_BT
Set energy difference threshold $\epsilon = 5\%$ and initialize $E_{\text{prev}} = 0$
for each time step t **do**
    Transform to Cartesian coordinates: $\mathbf{r}(t) = \mathbf{r}_0 + \sum_k r_k(t)e^{i\omega_k t}\mathbf{M}^{-1/2}\mathbf{v}_k$
    Compute total force $\mathbf{F}_{\text{total}}$ and harmonic force $\mathbf{F}_{\text{harm}} = -\mathbf{H}(\mathbf{r}(t) - \mathbf{r}_0)$
    Calculate source term $\mathbf{S} = \mathbf{F}_{\text{total}} - \mathbf{F}_{\text{harm}}$
    Project source onto modes: $S_k = \mathbf{v}_k^\dagger \mathbf{M}^{-1/2}\mathbf{S}$
    Update momenta: $p_k(t + \frac{\Delta t}{2}) = p_k(t) + \frac{\Delta t}{2}(S_k e^{-i\omega_k t})$
    Update positions: $r_k(t + \Delta t) = r_k(t)e^{i\omega_k \Delta t} + \frac{p_k(t+\frac{\Delta t}{2})}{\omega_k}e^{i\omega_k \Delta t}$
    Complete momentum update: $p_k(t + \Delta t) = p_k(t + \frac{\Delta t}{2}) + \frac{\Delta t}{2}(S_k e^{-i\omega_k (t+\Delta t)})$
    Apply Langevin thermostat: $p_k \to p_k e^{-\gamma \Delta t} + \sqrt{2\gamma k_B T(1 - e^{-2\gamma \Delta t})}\xi$
    Calculate current total energy $E_{\text{curr}} = \sum_k (\frac{1}{2}p_k^2 + \frac{1}{2}\omega_k^2 r_k^2)$
    Compute energy difference: $\Delta E_{\%} = |E_{\text{curr}} - E_{\text{prev}}|/E_{\text{prev}} \times 100\%$
    if $\Delta E_{\%} > \epsilon$ **then**
        Update reference geometry and normal modes
        Recompute the Hessian of the potential, and obtain new normal modes
        Project state onto new modes preserving energy
    end if
    Set $E_{\text{prev}} = E_{\text{curr}}$
end for

and projects the system onto new normal modes obtained through subspace iteration of the Hessian. Note that since the normal mode basis is redefined after each update, care must be taken in enumerating the degrees of freedom to ensure consistent tracking of vibrational modes across iterations, particularly in the presence of near-degenerate frequencies.

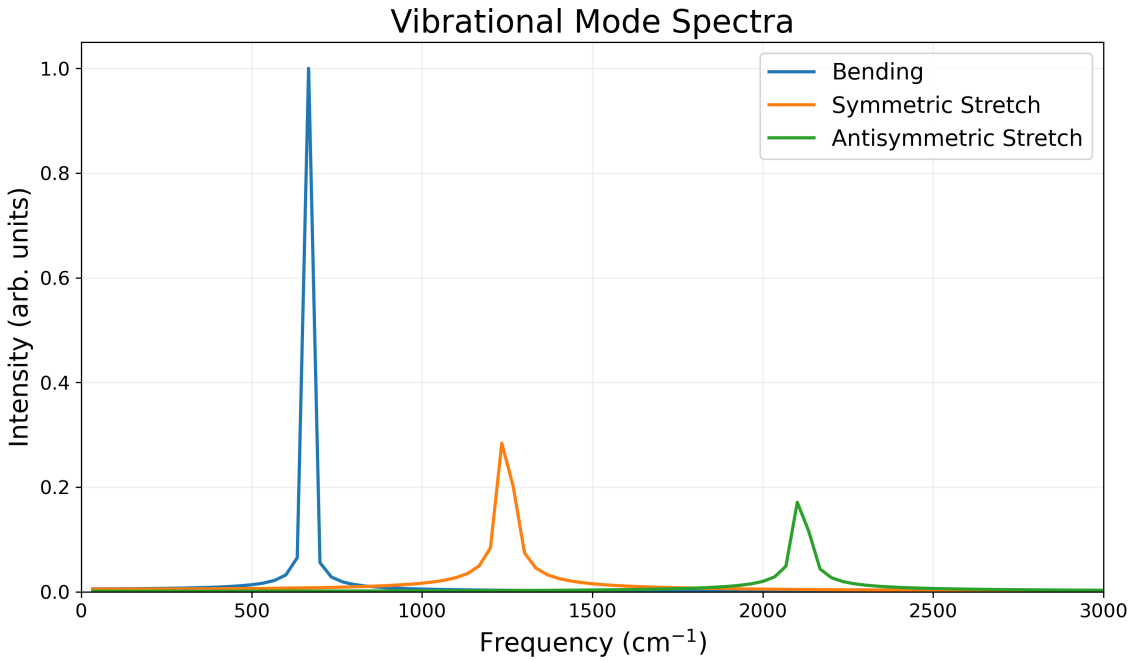


Figure 4.4. *Carbon dioxide vibrational spectra obtained using the methods introduced in this section with the harmonic potential*.

4.2.1 Frequency band selective dynamics

The Liouville operator formalism incorporates anharmonic contributions through source terms in the modified Euler-Lagrange equations. Analysis of the resulting trajectories in the frequency domain reveals these anharmonic effects and guides mode selection. For a trajectory $\mathbf{q}(t)$ evolved with the source-modified integrator over time window $[0, T]$, its Fourier transform:

$$\tilde{\mathbf{q}}(\omega) = \int_0^T \mathbf{q}(t) e^{-i\omega t} dt \tag{4.30}$$

and the conjugate momentum in frequency space:

$$\tilde{\mathbf{p}}(\omega) = i\omega \tilde{\mathbf{q}}(\omega) \tag{4.31}$$

provide a basis for analyzing the actual distribution of motion across frequencies. For a specific frequency band $[\omega_1, \omega_2]$, construction of new normal modes

proceeds through integration:

$$^{\omega_1,\omega_2}\mathbf{u} = \int_{\omega_1}^{\omega_2} \left[\tilde{\mathbf{q}}(\omega) + \frac{i}{\omega} \tilde{\mathbf{p}}(\omega) \right] d\omega \tag{4.32}$$

Within the specified frequency band [ω_1, ω_2], a restricted Lagrangian is constructed:

$$\mathcal{L}^{\text{vib}}_{\text{restricted}} = \sum_{\omega \in [\omega_1, \omega_2]} \frac{1}{2} (\dot{\tilde{\mathbf{q}}}(\omega) \dot{\tilde{\mathbf{q}}}^{(}\omega) - \omega^2 \tilde{\mathbf{q}}(\omega) \tilde{\mathbf{q}}^{(}\omega)) \tag{4.33}$$

By Theorem 4.2, this restricted dynamics preserves the symplectic structure, as each Fourier mode evolves independently on its harmonic toric space. The anharmonic contributions from coupling and external forces appear as projections onto these modes, leading to the equations of motion:

$$\frac{d}{dt} \tilde{\mathbf{q}}(\omega) = \frac{\partial \mathcal{L}^{\text{vib}}_{\text{restricted}}}{\partial \tilde{\mathbf{p}}^*(\omega)} \tag{4.34}$$

$$\frac{d}{dt} \tilde{\mathbf{p}}(\omega) = -\frac{\partial \mathcal{L}^{\text{vib}}_{\text{restricted}}}{\partial \tilde{\mathbf{q}}^*(\omega)} \tag{4.35}$$

The orthogonality of the Fourier basis [86, 87] ensures these equations evolve independently at each frequency, maintaining the symplectic structure of the original system within the selected band. Unlike Algorithm 5, this approach eliminates Hessian computations by selecting frequency bands and modes directly from trajectory data.

Combining Algorithms 5 and 6 enables sequential analysis of vibrational spectra: first, Algorithm 5 generates trajectory data, which Algorithm 6 then refines within specific frequency bands. Figure 4.4 demonstrates this method applied to carbon dioxide, where Algorithm 5 ran for 1000 steps followed by Algorithm 6 for another 1000 steps, both using a 1 fs time step. The vibrational spectra show excellent agreement with experimental values [88, 89], capturing the bending (667 cm^{-1}), symmetric stretching (1,388 cm^{-1}), and antisymmetric stretching modes (2,349 cm^{-1}). While this validates the proposed method, it also motivates a deeper investigation into molecular anharmonicity, which is undertaken in the following section.

Algorithm 6 Frequency Band Selective Molecular Dynamics

Load trajectory data $\{\mathbf{q}(t_i)\}_{i=1}^N$ and velocities $\{\mathbf{v}(t_i)\}_{i=1}^N$
Select frequency band $[\omega_1, \omega_2]$ of interest
Apply window function $w(t)$ to trajectory data (e.g., Hann window)
Compute discrete Fourier transforms:

$$\tilde{\mathbf{q}}(\omega_k) = \frac{1}{N}\sum_{i=1}^{N} w(t_i)\mathbf{q}(t_i)e^{-i\omega_k t_i}, \quad \tilde{\mathbf{p}}(\omega_k) = m\tilde{\mathbf{v}}(\omega_k)$$

where $\omega_k = 2\pi k/T$ for k such that $\omega_k \in [\omega_1, \omega_2]$
Initialize band-restricted coordinates:

$$\tilde{\mathbf{q}}_{\text{band}}(\omega_k) = \begin{cases} \tilde{\mathbf{q}}(\omega_k) & \text{if } \omega_k \in [\omega_1, \omega_2] \\ 0 & \text{otherwise} \end{cases}$$

Construct restricted Lagrangian: $\mathcal{L}^{\text{vib}}_{\text{restricted}} = \sum_{\omega_k \in [\omega_1,\omega_2]} \frac{1}{2}(|\dot{\tilde{\mathbf{q}}}(\omega_k)|^2 - \omega_k^2|\tilde{\mathbf{q}}(\omega_k)|^2)$
for each time step t **do**
    Force computation in restricted space:
        Transform to real space: $\mathbf{q}_{\text{band}}(t) = \sum_{\omega_k \in [\omega_1,\omega_2]} \text{Re}[\tilde{\mathbf{q}}(\omega_k)e^{i\omega_k t}]$
        Compute real-space forces: $\mathbf{F}_{\text{band}} = -\nabla V(\mathbf{q}_{\text{band}})$
        Project forces to frequency space: $\tilde{\mathbf{F}}(\omega_k) = \int_0^T \mathbf{F}_{\text{band}}(t)e^{-i\omega_k t}dt/T$
    Update positions and momenta for each $\omega_k \in [\omega_1, \omega_2]$:
        $\tilde{\mathbf{p}}(\omega_k, t + \frac{\Delta t}{2}) = \tilde{\mathbf{p}}(\omega_k, t) + \frac{\Delta t}{2}\tilde{\mathbf{F}}(\omega_k)$
        $\tilde{\mathbf{q}}(\omega_k, t + \Delta t) = \tilde{\mathbf{q}}(\omega_k, t)e^{i\omega_k \Delta t} + \frac{\tilde{\mathbf{p}}(\omega_k, t+\frac{\Delta t}{2})}{m\omega_k}(e^{i\omega_k \Delta t} - 1)$
        $\tilde{\mathbf{p}}(\omega_k, t + \Delta t) = \tilde{\mathbf{p}}(\omega_k, t + \frac{\Delta t}{2}) + \frac{\Delta t}{2}\tilde{\mathbf{F}}(\omega_k)$
    Apply Langevin thermostat to each mode ω_k:
        $\tilde{\mathbf{p}}(\omega_k) \to \tilde{\mathbf{p}}(\omega_k)e^{-\gamma \Delta t} + \sqrt{2m\gamma k_B T(1 - e^{-2\gamma \Delta t})}\tilde{\xi}(\omega_k)$
        where $\tilde{\xi}(\omega_k)$ is complex Gaussian noise with $\langle \tilde{\xi}(\omega_k) \rangle = 0$, $\langle |\tilde{\xi}(\omega_k)|^2 \rangle = 1$
end for
Reconstruct real-space trajectory via discrete Fourier synthesis:

$$\mathbf{q}_{\text{band}}(t_j) = \sum_{\omega_k \in [\omega_1,\omega_2]} 2\text{Re}[\tilde{\mathbf{q}}(\omega_k)e^{i\omega_k t_j}]$$

for output time points $\{t_j\}$

4.3 Understanding anharmonicities through phase space analysis

Characterizing anharmonicities and harmonic mode couplings in molecular dynamics remains a significant challenge. While harmonic modes are fundamental to analyzing molecular vibrations, understanding their coupling and energy exchange through anharmonic interactions is essential for explaining energy redistribution and spectroscopic properties. Traditional approaches require detailed knowledge of the potential energy surface—often computationally intractable for complex systems.

This section presents a geometric approach that leverages Fourier integrator-based molecular dynamics to circumvent these limitations. By analysing the phase-space manifold in the Fourier mode basis, molecular motion is separated into harmonic and anharmonic components without requiring explicit knowledge of the potential energy surface. The phase space naturally decomposes into two structures: a product of harmonic tori (each representing a pure harmonic mode's closed path) and a quotient space capturing anharmonic motion. This decomposition enables quantification of inter-mode energy transfer directly from molecular trajectories.

This geometric approach is demonstrated using carbon dioxide as the model system.

4.3.1 Phase space manifold structure in Fourier mode basis

The phase space structure of carbon dioxide in the Fourier mode basis can be visualized in $\mathbb{R}^3$ through dimensionality reduction, as shown in Figure 4.5. The resulting manifold illustrated in the figure captures 99.8% of the total variance, with the first two principal components accounting for 54.4% and 45.4% respectively. The colour gradient on the manifold represents the total energy distribution, providing insight into the energetic landscape of the molecular

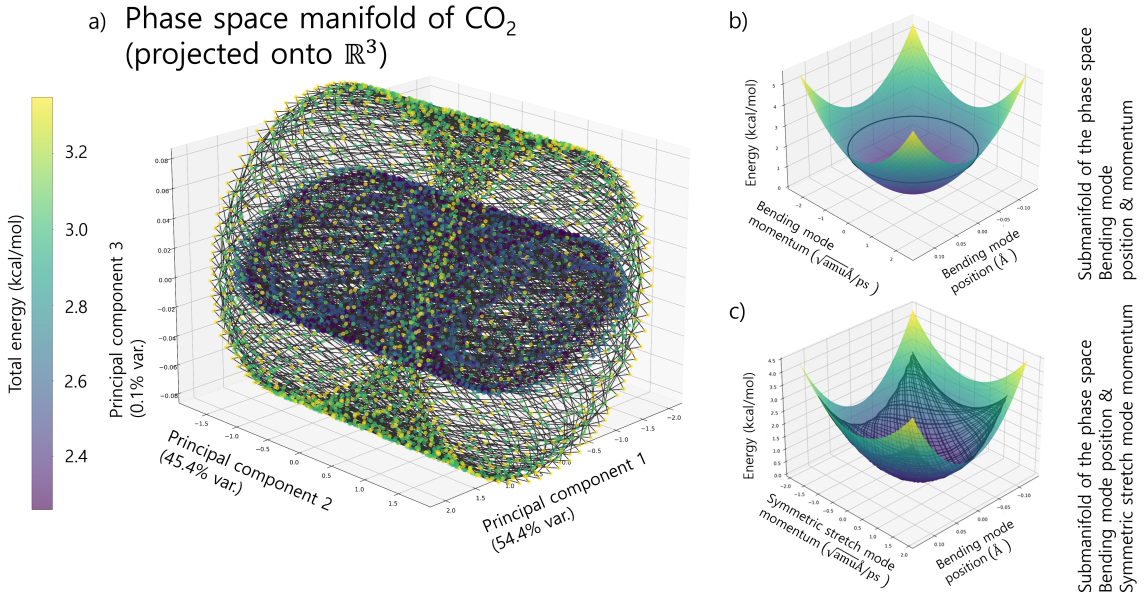


Figure 4.5. *The phase space manifold of CO_2 in Fourier mode basis. a) Phase-space manifold of CO_2 in Fourier mode basis projected onto $\mathbb{R}^3$ via PCA. b) Submanifold of the bending mode dynamics in phase space. c) Submanifold showing the coupling between symmetric stretch mode momentum and bending mode position.*

motion.

The manifold decomposes into toric submanifolds representing specific modes of motion, alongside regions capturing anharmonicities. Figure 4.5b shows the phase space structure of the bending mode, where the purely harmonic nature of this mode in the Fourier basis produces a perfect circular trajectory within the potential well. This circular path demonstrates the conservation of the bending mode's energy contribution throughout the motion.

Figure 4.5c illustrates the coupling between symmetric stretch mode momentum and bending mode coordinates. The trajectory on this submanifold exhibits a distinctive sieve-like pattern bounded by energy constraints. This non-circular oscillatory behavior provides direct evidence of energy transfer through the symmetric mode.

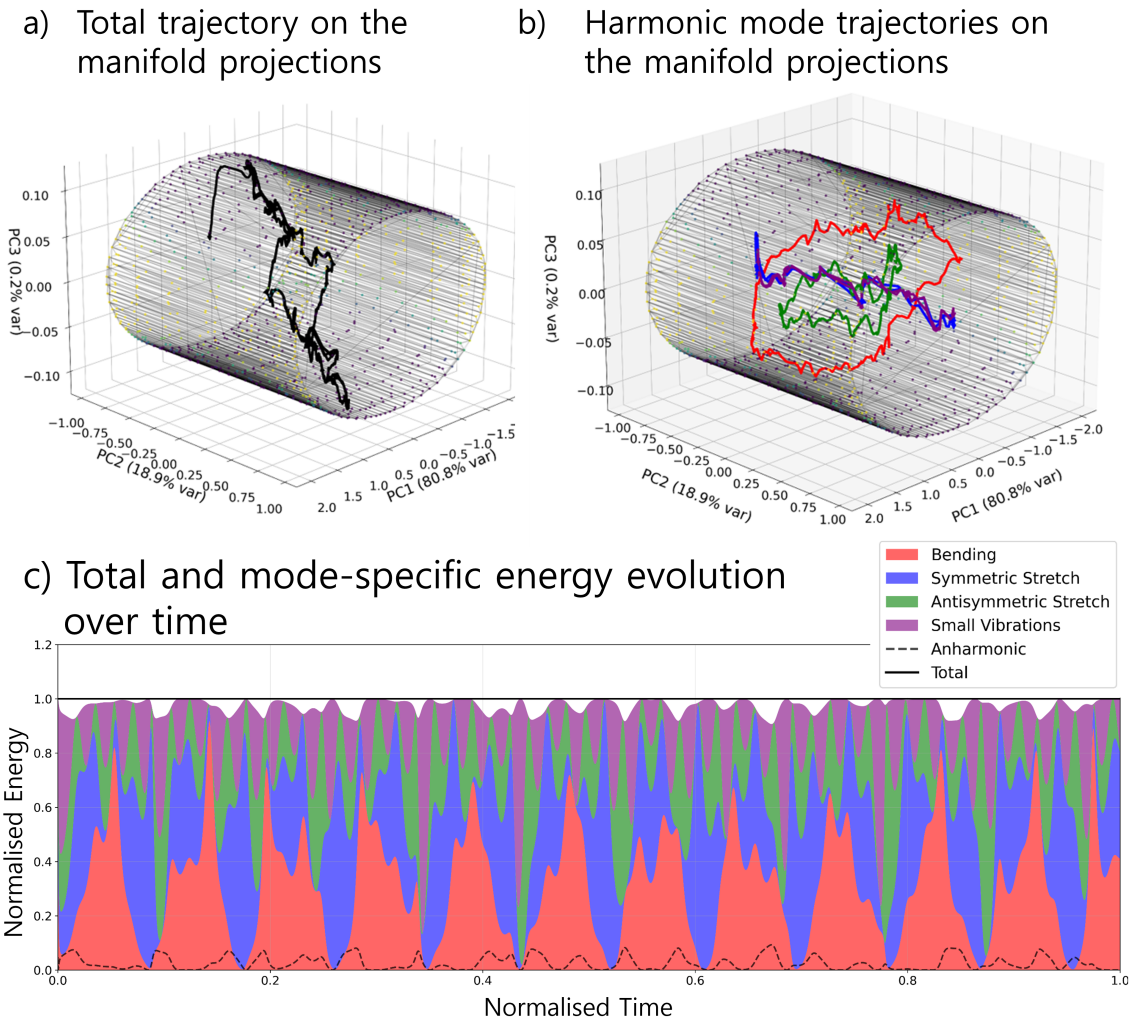


Figure 4.6. *Trajectories on the phase space manifold of CO_2 in Fourier mode basis. a) Total trajectory projected onto the manifold. b) Harmonic mode trajectories on the manifold. c) Energy evolution over time. The constant total energy (solid black) is distributed among mode contributions (coloured regions) and anharmonic contributions (dashed line), showing energy transfer between modes.*

Figure 4.6a) shows the total trajectory (black) on the manifold, which includes both harmonic and anharmonic components of the motion. The irregular pattern of this trajectory reflects the complexity arising from mode

coupling and anharmonic effects.

Figure 4.6b) presents the decomposition into individual harmonic mode trajectories, with each mode shown in a different colour. The bending mode (red), symmetric stretch (blue), antisymmetric stretch (green), and small vibrations in the higher-frequency domain (purple) exhibit more regular patterns compared to the total trajectory, characteristic of harmonic motion. However, their paths still show deviations from perfect periodic orbits, indicating residual anharmonic effects.

The energy evolution plot (Figure 4.6c) quantifies the energy exchange between modes over time. The stacked area plot reveals how energy flows between different vibrational modes while maintaining constant total energy (black line). The thickness of each coloured region represents the instantaneous energy contribution from each mode. The persistent but small anharmonic contribution (dashed line) demonstrates that while anharmonic effects are always present, they remain bounded.

While the total energy remains conserved, individual mode-specific energies exhibit temporal variations due to inter-modal energy transfer driven by anharmonic interactions. The anharmonic origin of these energy transfers is evident in Figure 4.7, which illustrates the temporal evolution of energy distribution. Prior to anharmonic energy input, each mode maintains a constant energy contribution. Following the input, the energy contributions fluctuate across modes, indicating inter-modal energy transfer.

To quantify the energy transfer dynamics, the correlation between modal energy contributions is computed:

$$\mathrm{Corr}(E_i, E_j) = \frac{\mathrm{Cov}(E_i, E_j)}{\sigma_{E_i}\sigma_{E_j}} \tag{4.36}$$

where E_i and E_j denote the energies of modes i and j, respectively, Cov represents the covariance, and σ indicates the standard deviation. Figure 4.8 shows the temporal evolution of energy correlations between harmonic modes after anharmonic energy input at t=0.05 (normalized time). Positive correlations

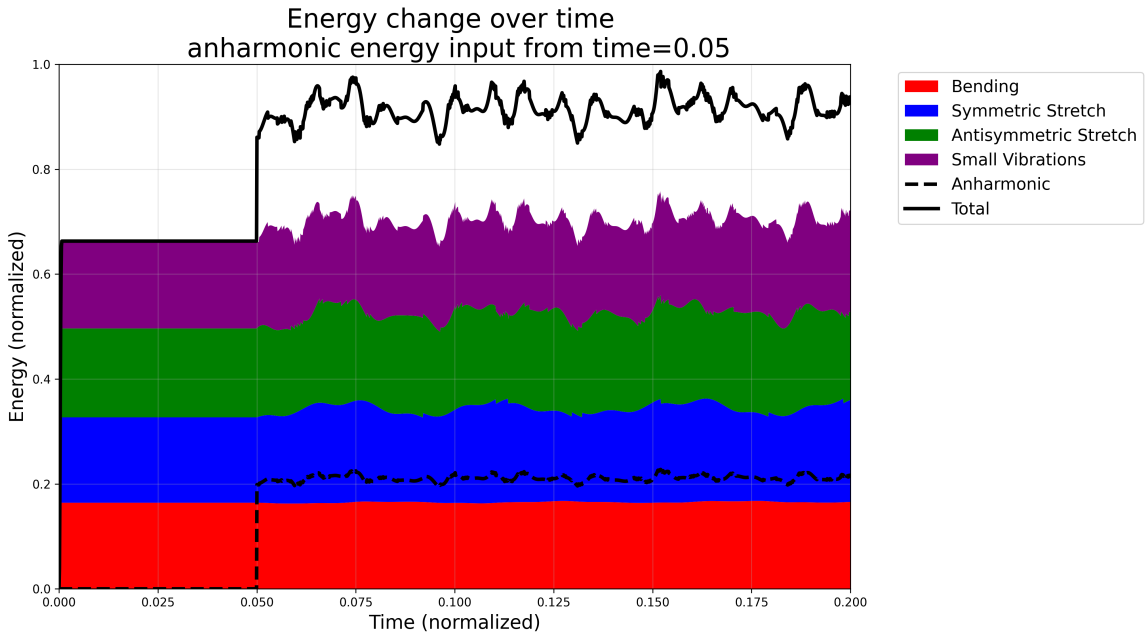


Figure 4.7. *Normalised plot of energy evolution over time, where the anharmonic energy input to the system is introduced from time 0.05 in the normalised scale.*

indicate simultaneous increase or decrease in mode energies, while negative correlations signify opposing energy changes. The bending and symmetric stretch modes exhibit strong positive correlations, frequently approaching unity, demonstrating robust energy coupling. During t=[0.06, 0.09], for example, the mode energy plot confirms this coupling through synchronized energy variations between these modes. In contrast, the symmetric and antisymmetric stretch modes display negative correlation in the later part of this window, evidenced by their opposing energy trajectories. For high-frequency small vibrational modes, the correlation of this mode to the bending mode is negative in this domain, indicating energy transfer from one mode to the other between these two modes.

Figure 4.9 illustrates the transfer rate between harmonic modes, calculated as the time derivative of the energy correlation. The high-frequency small vibrational modes exhibit a strong transfer rate with the antisymmetric

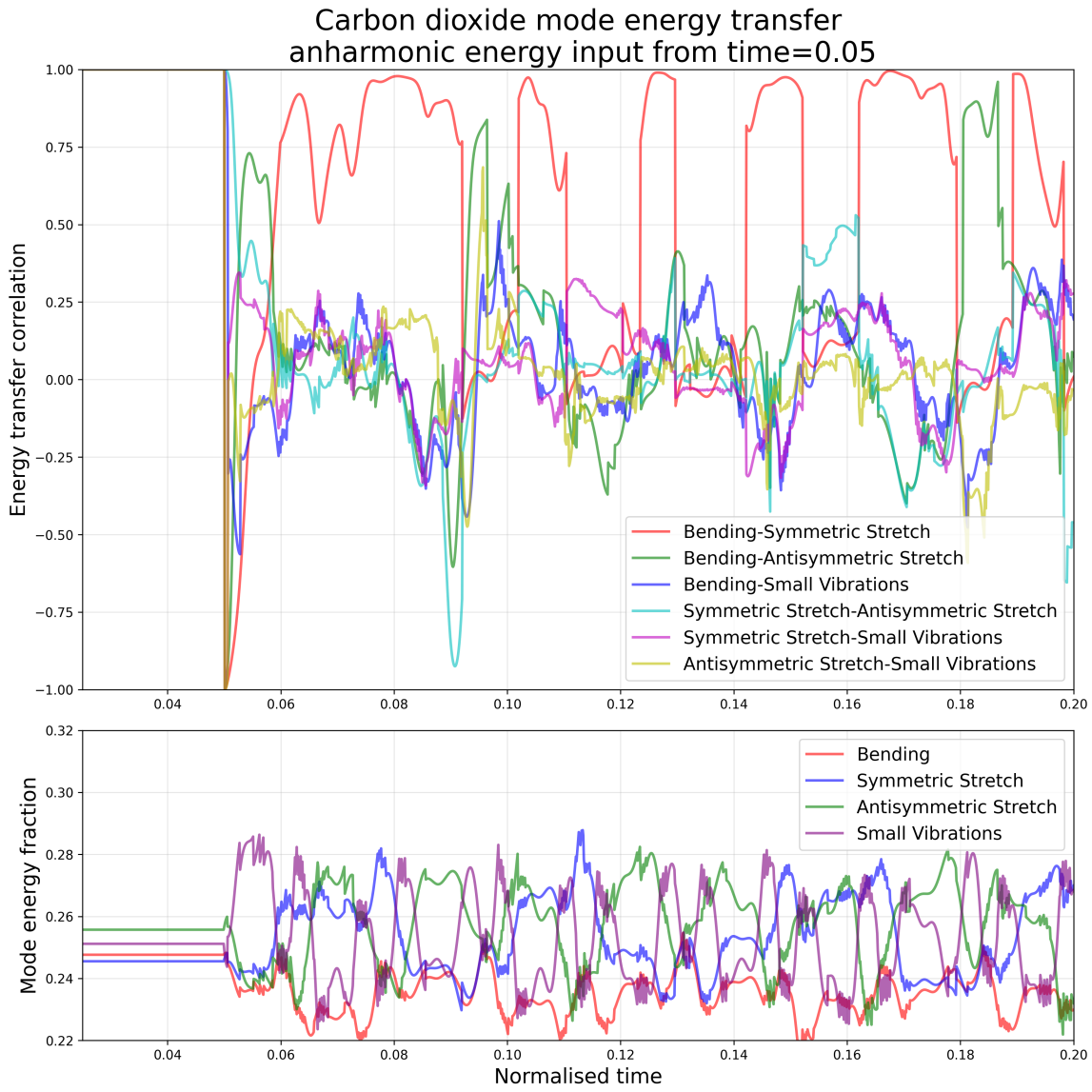


Figure 4.8. *Energy transfer correlations* (*top*) *and mode energy fractions* (*bottom*) *in* CO_2 *after anharmonic excitation at t=0.05, in the normalised scale, showing coupled vibrational dynamics between different normal modes.*

stretch mode, indicating significant energy exchange between these modes. Although the bending and symmetric stretch modes are positively correlated, the exchange rate between these two modes is small in magnitude, indicating

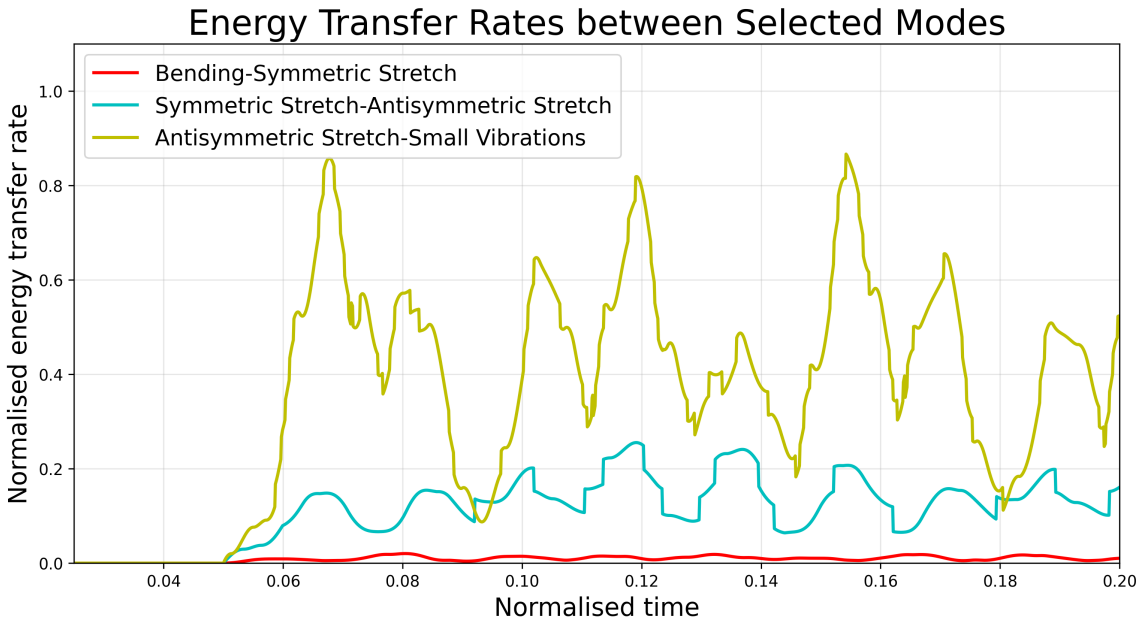


Figure 4.9. *Normalised transfer rate over time for selected mode-couples. Anharmonic excitation is introduced at t=0.05 in the normalised time scale.*

that there is only a small energy exchange between these modes, but they exchange energy indirectly via other modes.

4.4 Pseudo-normal modes for thermostatted molecular system

In the previous sections, the Fourier integration framework, frequency-limiting formulation of the dynamics and explicit consideration of anharmonicities in Fourier integration framework were established. Building upon this foundation, this section turns attention to the changing nature of normal modes due to the thermostats. For carbon dioxide molecule considered in the previous sections, the Fourier integrators worked perfectly as the system size is small and the shift in normal modes due to the dissipative terms is minimal. However, for bigger molecules, even for relatively small molecules like phenol, if thermostats intro-

duce dissipative terms and external energy-sources connected to a heat bath, they can fundamentally alter the nature of vibrational dynamics, particularly when anharmonic effects are present. The pseudo-normal mode approach extends the framework by introducing a perturbative correction to standard normal modes, creating a basis that captures anharmonic contributions to molecular motion while maintaining compatibility with standard thermostats. This method is widely used across various fields of physics concerned with dissipative vibrational modes, including geophysics [90], fluid dynamics [91,92], acoustics [93], and molecular physics [94].

4.4.1 Pseudo-normal modes of molecules under a thermostat

As established in Theorem 4.2, the phase space of a purely harmonic system with n^v vibrational modes is isomorphic to the product of n^v circles. Each mode follows an elliptic trajectory in its own subspace, reflecting the conservative nature of harmonic motion. However, when such a system is coupled to a thermal environment, this idealised structure is no longer preserved. The presence of dissipative forces and stochastic energy exchange–characteristic of thermostatted dynamics–distorts the harmonic geometry, as evidenced by the non-elliptic, noisy trajectories observed in the normal mode representation (see Figure 4.6).

This distortion signals a breakdown of the harmonic mode decomosition: the normal mode basis, which once aligned with the invariant tori of the undisturbed phase space, no longer captures the true geometry of motion under thermal influence. Dissipation alters not just the energies of individual modes, but the very directions along which meaningful vibrational dynamics unfold. In this context, the pseudo-normal mode approach arises as a natural extension–a perturbative redefinition/correction of the modal basis that accomodates the deformation of phase space structure induced by the heat bath.

Let $\{w^1{}_\mu(t), \cdots w^m{}_\mu(t)\}$ be the set of normal mode vectors of the molecule

at time t and m be the total number of normal mode directions of the system. The matrix elements of standard normal modes can then be formed as

$$M_{\mu\nu} = w^{\nu}{}_{\mu}(t) \tag{4.37}$$

where ν indexes the vibrational modes and μ indexes the coordinates. The pseudo-normal modes can then be constructed as a perturbative modification of $M_{\mu\nu}$ as

$$P_{\mu\nu} = M_{\mu\nu} + \tilde{\alpha}\delta M_{\mu\nu} \tag{4.38}$$

where $\tilde{\alpha}$ is a small dimensionless parameter controlling the strength of the perturbation, and $\delta M_{\mu\nu}$ is a correction term derived from anharmonic contributions to the potential energy. [91] This formulation is motivated by the perturbative expansion of the potential introduced in Equation 4.24. The standard normal modes are constructed via the diagonalization of the second-derivative term of this equation, where $\delta M_{\mu\nu}$ is designed to approximately account for contributions from higher-order derivatives.

The quantity $\delta M_{\mu\nu}$ can be obtained via correlations observed in the system's vibrational dynamics under thermal regulation. If one considers the time series of mode-projected coordinates $q_\nu(t) = \sum_\mu M^\dagger_{\nu\mu}\tilde{q}_\mu(t)$ obtained by projecting the mass-weighed Cartesian displacements, $\tilde{q}_\mu$, onto the harmonic normal mode basis via $M^\dagger$. In a purely harmonic system, these coordinates evolve independently, resulting in a diagonal time-averaged covariance matrix

$$C_{\mu\nu} = \langle q_\mu(t) q^*_\nu(t)\rangle_t \approx \frac{1}{L_t}\sum_{a=1}^{L_t} q_\mu(t_a) q^*_\nu(t_a) \tag{4.39}$$

given that one is concerned with L_t past time-steps to evaluate $C_{\nu\mu}$. However, when the system is under an influence of a thermostat, the off-diagonal elements of $C_{\mu\nu}$ are non-zero due to the mode-mixing induced by the dissipation and anharmonicity. With this in mind, the correction matrix $\delta M_{\mu\nu}$ can be defined as

$$\delta M_{\mu\nu} = \sum_{\kappa\neq\nu} M_{\mu\kappa} C_{\kappa\nu} \tag{4.40}$$

which represents a linear mixing of mode vectors weighted by their dynamic correlations.

Once the pseudo-normal mode matrix, $P_{\mu\nu}$, is obtained, one can construct a transformation matrix that maps the normal mode coordinates to pseudo-normal mode coordinates:

$$T_{\mu\nu} = \sum_{\kappa} P^{\dagger}_{\mu\kappa} M_{\kappa\nu}. \tag{4.41}$$

This transformation maps the normal mode coordinates to pseudo-normal mode coordinates as:

$$q^{(P)}_{\mu} = \sum_{\kappa} T_{\mu\kappa} q_{\kappa}, \quad p^{(P)}_{\mu} = \sum_{\kappa} T_{\mu\kappa} p_{\kappa} \tag{4.42}$$

Thermostats can then be applied to each pseudo-mode independently, by evolving them as damped harmonic oscillators with a corrective rescaling step that drives each mode's energy toward the thermal target. Since the dynamics is defined in the normal mode frame, the updated coordinates must be projected back using the inverse of the transformation matrix:

$$q^{\text{new}}_{\mu} = \sum_{\kappa} \left(T^{-1}\right)_{\mu\kappa} q^{(P)}_{\kappa}, \quad p^{\text{new}}_{\mu} = \sum_{\kappa} \left(T^{-1}\right)_{\mu\kappa} p^{(P)}_{\kappa} \tag{4.43}$$

This approach allows thermostatting to operate in a rotated, perturbed basis that captures the influence of dissipation and anharmonic coupling while preserving compatibility with the Fourier mode integrator.

By construction, the pseudo-normal mode transformation $T_{\mu\nu} = \sum_{\kappa} P^{\dagger}_{\mu\kappa} M_{\kappa\nu}$ is a square $(3N-6) \times (3N-6)$ matrix that maps between two bases of the same vibrational subspace. Since $\delta M_{\mu\nu}$ is built from linear combinations of the original normal mode vectors $M_{\mu\kappa}$, the transformation cannot increase the rank or dimensionality of the space. The existence of T^{-1} ensures that the pseudo-normal modes span exactly the same $(3N-6)$-dimensional space as the original normal modes, where 6 degrees of freedom are removed to account for the 3 translational and 3 rotational rigid-body motions.

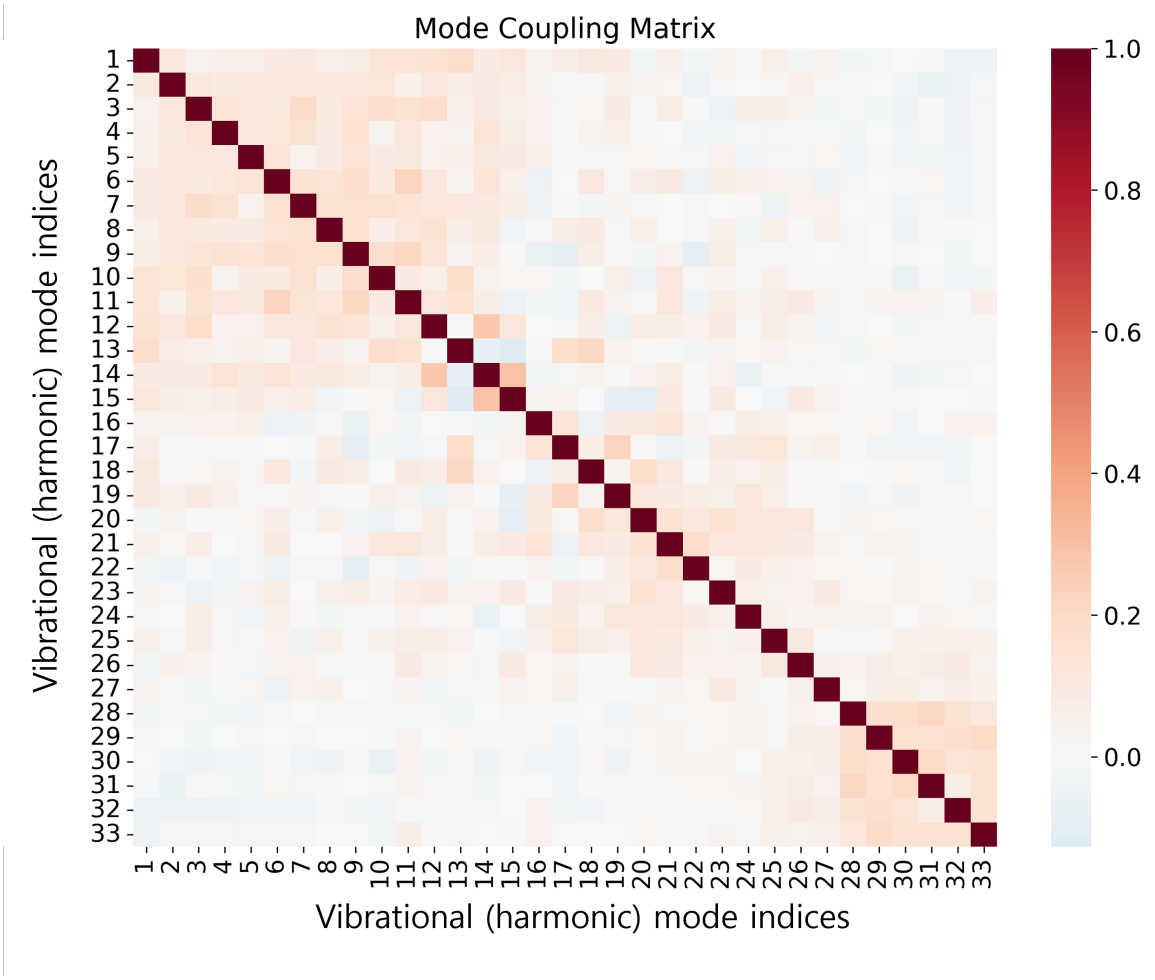


Figure 4.10. *Average coupling matrix for the vibrational modes of phenol. The entries of the matrix are correlation between the modes.*

The overall flow of the procedure can be summarised as:

$$\text{normal modes} \xrightarrow{T} \text{pseudo-normal modes} \xrightarrow{\text{thermostat}} \text{pseudo-normal modes} \xrightarrow{T^{-1}} \text{normal modes} \tag{4.44}$$

4.5 A case study: phenol

In this section, phenol molecule's vibrational density of states from trajectories obtained by traditional velocity Verlet integrators, Fourier method introduced in this chapter and the Fourier method with pseudo-normal mode corrections. The force field parameters are derived from the Generalized AMBER Force Field (GAFF) [60] through the Nucleic Acid Builder (NAB) program [95]. NAB was chosen for its capability to provide analytical expressions for both forces and Hessians and its integrability to the AMBER molecular dynamics package [96].

Phenol, consisting of 13 atoms, has 33 vibrational modes. Table 4.1 lists the vibrational modes of phenol along with their corresponding frequencies. As can be seen from the table, the 33 vibrational modes of phenol exhibit a variety of frequency values and different combinations of internal coordinate contributions, including bond stretching, in-plane bending, out-of-plane bending, and torsion. The vibrational frequencies span from 242 cm^{-1} to 3,655 cm^{-1}, with the highest frequency corresponding to the O-H stretch mode and the lowest to the γCC mode with CO contribution. Many modes show mixed character, particularly in the mid-frequency range, where multiple internal coordinates contribute to single normal modes. For instance, the mode at 1,609 cm^{-1} combines CC stretching (64%), CH bending (18%), and CC bending (10%), demonstrating the complex nature of phenol's molecular vibrations.

Figure 4.11 presents six different vibrational density of states (VDOS) profiles, obtained via the traditional velocity Verlet integration method, the Fourier integration scheme with and without pseudo-normal mode corrections, and several band-limited Fourier integration approaches.

Figure 4.11a) highlights the importance of pseudo-normal mode corrections in the Fourier integration scheme. The VDOS obtained from the standard velocity Verlet method and from the Fourier-based method applied to pseudo-normal modes show excellent agreement in peak positions, widths, and amplitudes. In contrast, the Fourier scheme without pseudo-normal corrections

Table 4.1. *Vibrational Modes and Frequencies of Phenol* (*gas phase*). *The ratios in the vibrational mode descriptions indicate the relative contribution of each internal coordinate to the normal mode*, *where* ν *represents bond stretching*, δ *represents in-plane bending*, γ *represents out-of-plane bending*, *and* τ *represents torsion*. *The values are from the work of Gábor Keresztury et al*. *in* $293\,\mathrm{K}$. *[2]*

Vibrational Mode	Measured Frequencies (cm^{-1})
νOH	3655
νCH	3021, 3046, 3052, 3061, 3074
0.64νCC+0.18βCH+0.1βCC	1609
0.66νCC+0.16βCH+0.9βCC	1604
0.56βCH+0.33νCC+0.06νCO	1501
0.52βCH+0.33νCC+0.08βOH	1472
0.63βCH+0.26βOH+0.07νCC	1361
0.63νCC+0.33βCH	1344
0.5νCO+0.19νCC+0.17βCH+0.1βCC	1261
0.45βOH+0.26νCC+0.23βCH	1197
0.75βCH+0.24νCC	1176
0.23νCC+0.74βCH	1150
0.55νCC+0.38βCH	1070
0.69νCC+0.23βCH+0.06βCC	1026
0.61βCC+0.37νCC	999
0.83γCH+0.17γCC	973
0.88γCH+0.11γCC	881, 956
γCH	823
0.42νCC+0.30βCC+0.24νCO	810
0.65γCH+0.07γCC+0.27γCO	752
0.05βCH+0.91γCC	687
0.82βCC+0.11νCC+0.05νCO	618
0.78βCC+0.11νCC+0.09νCO	526
0.49γCC+0.45γCO+0.06γCC	503
0.88γCC+0.12γCO	420
0.08βCC+0.08νCC+0.79βCO	410
0.98τOH	310
0.87γCC+0.13γCO	242

exhibits significant deviations in both amplitudes and peak locations.

Figures 4.11b)–d) display zoomed-in views of the VDOS under band-limited Fourier integration. Each band-restricted simulation successfully captures vibrational peaks within its targeted frequency window. The low-frequency and high-frequency cases, shown in Figures 4.11b and 4.11d, fully isolate the dynamics within their respective bands—yielding zero amplitude outside their specified ranges.

The intermediate frequency range, targeting 1500–2000 cm^{-1} (Figure 4.11c)), tells a slightly different story. While the method isolates the intended vibrational modes, a small bump appears in the low-frequency region (900–1100 cm^{-1}). This feature is not an unexpected artefact. The 1500–2000 cm^{-1} band corresponds to ring deformations in phenol's structure, and since the ring is connected to other parts of the molecule, its motion couples with lower-frequency vibrations such as those involving C-H and C-O bonds.

This coupling is also supported by the correlation matrix shown in Figure 4.10, where the 11$^{\text{th}}$ to 26$^{\text{th}}$ vibrational modes (which include this ring region) are significantly correlated with lower-frequency modes below the 10$^{\text{th}}$.

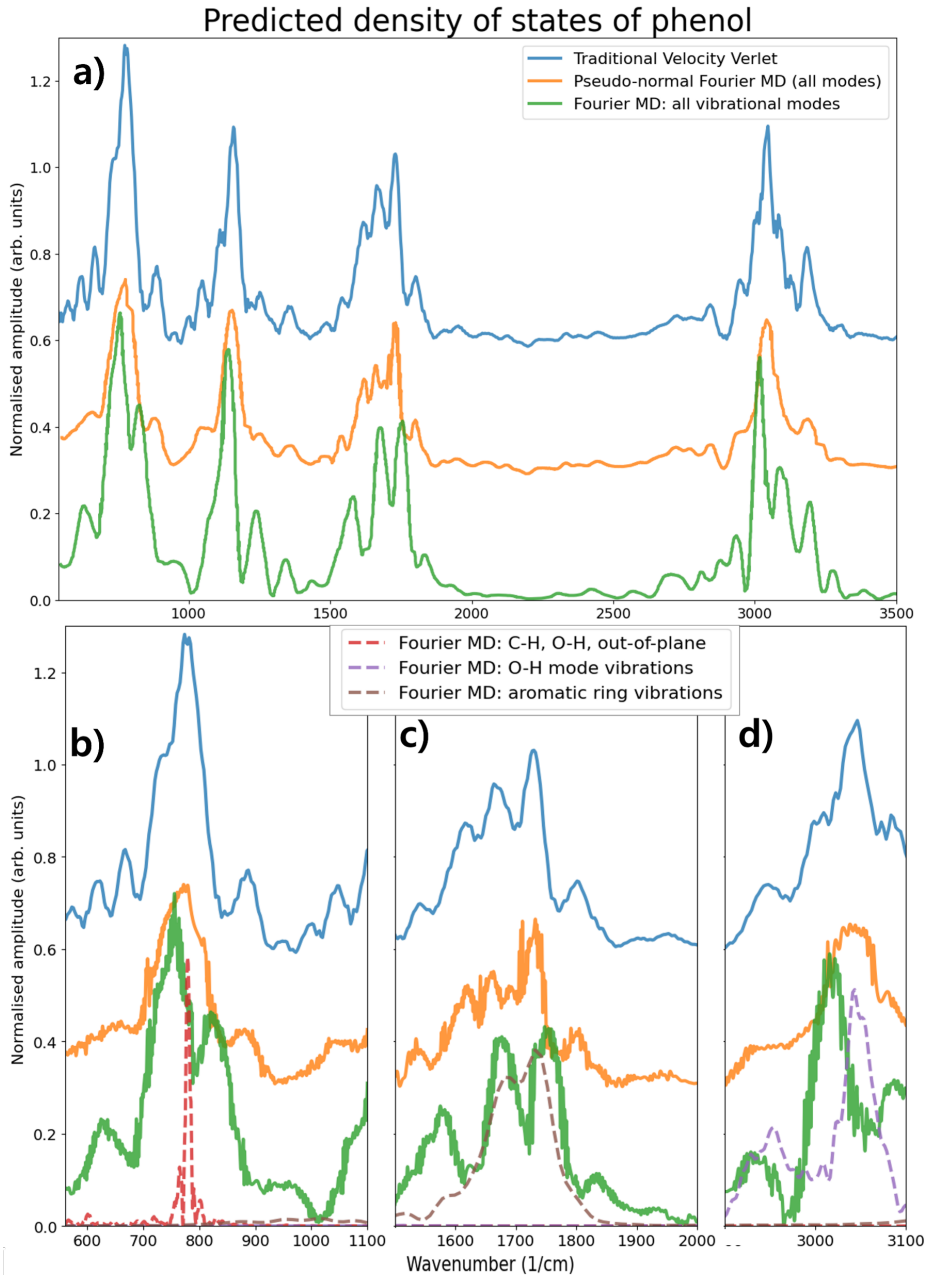


Figure 4.11. *Predicted vibrational density of states (VDOS) of phenol using traditional and Fourier-based integration schemes.* (*a*) *VDOS obtained from trajectories using traditional velocity Verlet integration* (*top*, *blue*), *Fourier integrators including all vibrational modes* (*bottom*, *green*), *and Fourier integrators with pseudo-normal mode corrections* (*middle*, *orange*). (*b*) *Zoomed-in view of the 550–1100* cm^{-1} *region, with band-limited dynamics applied over a 500–1000* cm^{-1} *bandwidth* (*dashed red*). (*c*) *Zoomed-in view of the 1500–2000* cm^{-1} *region, with band-limited dynamics over the same frequency range* (*dashed brown*). (*d*) *Zoomed-in view of the 2900–3100* cm^{-1} *region, with band-limited dynamics over the same range* (*dashed purple*).

CHAPTER 5

Geometrical molecular dynamics and uncertainty-driven thermostat

... Das Verhältnis dieser Voraussetzungen bleibt dabei im Dunkeln; man sieht weder ein, ob und in wie weit ihre Verbindung notwendig, noch a priori, ob sie möglich ist.

GEORG FRIEDRICH BERNHARD RIEMANN
Über die Hypothesen, welche der Geometrie zu Grunde liegen

The relationship between space and the means of construction remains in darkness; one perceives neither whether and to what extent their connection is necessary, nor a priori whether it is possible.

GEORG FRIEDRICH BERNHARD RIEMANN
On the hypotheses which underlie geometry

Classical molecular dynamics simulations are extensively employed to compute both the qualitative and quantitative features of a molecule's vibrational density of states and its associated vibrational spectra. Despite the utility of these simulations, reliably attributing spectral peaks in the vibrational density of states or the power spectrum to specific normal modes or localised vibrational motions remain a nontrivial problem, particularly in systems with strong mode coupling or anharmonic effects.

In the harmonic approximation, where molecular coordinates remain close to their equilibrium configuration on the potential energy surface, the vibrational modes correspond to the normal modes. The potential energy is approximated as a quadratic function of the atomic displacements, reducing the time evolutions to a system of linear differential equations. While the full system remains a many-body problem, the linearity of these differential equations allows for exact solutions via diagonalisation of the mass-weighted Hessian

matrix [97]. Each normal mode behaves as an independent harmonic oscillator, yielding a molecular trajectory that can be expressed as a superposition of sinusoidal motions.

When the harmonic approximation is no longer valid–either due to large-amplitude displacements, strong anharmonicity, thermostat, or mode coupling–the vibrational dynamics become intrinsically non-linear and cannot be described by a fixed set of normal modes [47]. In such anharmonic regime, vibrational dynamics deviate significantly from the harmonic model: oscillation frequencies become amplitude-dependent [98] and intramolecular vibrational energy redistribution leads to rapid energy exchange between modes—phenomena that can result in highly complex and even chaotic molecular trajectories [99]. These can lead to very complex or even chaotic trajectories as there are more than two interacting bodies in general molecular system [100, 101].

These anharmonic complexities render the task of associating spectral features with specific vibrational modes particularly challenging. In highly coupled or nonlinear systems, vibrational energy is not confined to well-defined normal modes but instead delocalises across many degrees of freedom. Consequently, rather than attempting to recover a single mode associated with a spectral frequency, a more fruitful approach is to identify the underlying (quasi-)periodic motions that contribute to sustained vibrational patterns in phase space.

Even in systems exhibiting global anharmonicity and regions of chaos, certain initial conditions can give rise to bounded, recurrent motions, corresponding to low-dimensional invariant tori embedded within the high-dimensional phase space. For instance, early work by Swamy and Hase on formaldehyde revealed that specific vibrational excitations yield periodic or nearly periodic motions, with trajectories confined to islands of stability amidst otherwise chaotic dynamics [102]. More recently, Bach, Hostettler, and Chen showed that highly excited ethyl radicals can become dynamically trapped in long-lived oscillatory states, delaying dissociation by temporarily localising energy within

select vibrational coordinates [103].

These findings suggest the presence of an underlying dynamical structure in vibrational trajectories, intimately connected to their degree of quasi-periodic behaviour, including the strictly periodic case. When trajectories evolve on a compact manifold, such as the symplectic phase space of a Hamiltonian system, long-time boundedness and recurrence can be rigorously established under known conditions [10, 104]. When such periodic orbits are identified, they can be directly associated with distinct vibrational spectral features, serving as classical counterparts to quantum vibrational modes [105, 106]. This chapter aims to demonstrate how tools from Riemannian geometry, together with insights from the study of *monodromy*, can be employed to construct molecular trajectories whose bounded, periodic behaviour corresponds to specific vibrational periods corresponding to the given frequency bands of interest.

5.0.1 Structure of the chapter

Chapter 5.1 introduces basic concepts of Riemannian geometry and how molecular dynamics can be described according to the framework provided by Riemannian geometry via Jacobi metric. At the end of this section, a proof that the manifold which molecular trajectories are defined on is compact, and the interior of this manifold is Riemannian with compact closure.

Chapter 5.2 introduces a method for thermostatting molecular systems based on the quantum mechanical uncertainty principle. The thermostat presented here offers a gentler alternative to conventional approaches such as the Langevin or Nosé–Hoover methods.

Chapter 5.3 demonstrates how the techniques developed in Chapters 5.1 and 5.2 can be applied to compute the trajectory of a carbon dioxide molecule.

Chapter 5.4 presents a generalisation of normal mode analysis to Riemannian manifolds using the Jacobi metric. Building upon this foundation, Chapter 5.5 develops a framework for identifying periodic orbits, and it contains the

main results concerning trajectory periodicity. Readers with a background in differential geometry may wish to proceed directly to Chapter 5.5.

Lastly, Chapter, 5.6 introduces how the tools from fat manifolds can be used to study the coupling between two vibrational modes.

5.1 Geometric description of mechanics

Let r_μ and p_μ be the position and momentum coordinates of the given system with Hamiltonian $\mathcal{H} = T + V$. The Lagrangian of this system is then defined as

$$\mathcal{L} = T - V = p_\mu \dot{r}^\mu - \mathcal{H} \tag{5.1}$$

where $T = \frac{1}{2} m^{\mu\nu} p_\mu p_\nu$ is the kinetic energy, V is the potential energy of the system, and $m_{\mu\nu}$ is the mass matrix, with its inverse being $m^{\mu\nu}$. This matrix represents the effective mass between the position coordinates indexed by μ and ν.

Throughout this manuscript, Einstein summation convention is used to denote the tensors and their contraction. An element of a vector space $v \in \mathcal{V}$ would be written as

$$v^\mu = \left(v^1, \cdots, v^{\dim(\mathcal{V})}\right) \tag{5.2}$$

and elements of n-tensors, $w \in \mathcal{V} \otimes \cdots \mathcal{V}$ are denoted as: $w^{\mu_1 \mu_2 \cdots \mu_n}$. The sum over any index that appears twice—once as a superscript and once as a subscript—is understood to run from 1 to $\dim(\mathcal{V})$, and the summation symbol is omitted. For example, for any $u, v, w \in \mathcal{V}$,

$$u_\mu v^\mu w_\kappa = \sum_{\mu=1}^{\dim(\mathcal{V})} u_\mu v^\mu w_\kappa. \tag{5.3}$$

Here, μ is called a dummy index, while any index that appears only once in a term (a free index) labels the components of the resulting tensor. Contraction via the Einstein convention thus reduces the total tensor rank by two.

One can define the action of the system as

$$\mathcal{S} = \int \mathcal{L} dt. \tag{5.4}$$

Hamilton's principle of least action is defined so that the Hamiltonian flow of the system keeps the action stationary. Formally, this statement can be written as

$$\delta\mathcal{S} = \int \left(\frac{\partial \mathcal{L}}{\partial r_\mu} \delta r_\mu + \frac{\partial \mathcal{L}}{\partial \dot{r}_\mu} \delta \dot{r}_\mu \right) dt = 0, \tag{5.5}$$

where $\delta\mathcal{S}$ represents the variation of the action under small changes in the trajectory given by the Hamiltonian flow, and δr_μ and $\delta \dot{r}_\mu$ are the small changes in the respective variables.

This variational principle bears a striking resemblance to the definition of geodesic, i.e., the shortest path between two points, in differential geometry. Just as mechanical systems follow paths that minimize the action, geodesics follow paths that minimize (or more generally, make stationary) the length functional:

$$s = \int ds = \int \sqrt{g_{\mu\nu} \dot{x}^\mu \dot{x}^\nu} dt, \tag{5.6}$$

where ds is the element of the arclength and $g_{\mu\nu}$ is the metric tensor of the manifold. The metric tensor here encodes all the geometric information about distances and angles in our space–for a flat Euclidean space, it reduces to the identity matrix $\delta_{\mu\nu}$, as it transforms ds as expected in the usual cartesian coordinate system in flat space as follows:

$$ds = \sqrt{\delta_{\mu\nu} \dot{x}^\mu \dot{x}^\nu} dt = \sqrt{\dot{x}_\mu \dot{x}^\mu} dt \tag{5.7}$$

which is the Pythagorean identity giving the usual Euclidean norm on $\mathbb{R}^d$, while for curved spaces it captures the local distortion of distances.

In Riemannian geometry, the inner product ($< \cdot, \cdot >$) between two vectors is naturally defined by the definition of metric g, as:

$$\langle x, y \rangle_g = x^\mu g_{\mu\nu} y^\nu. \tag{5.8}$$

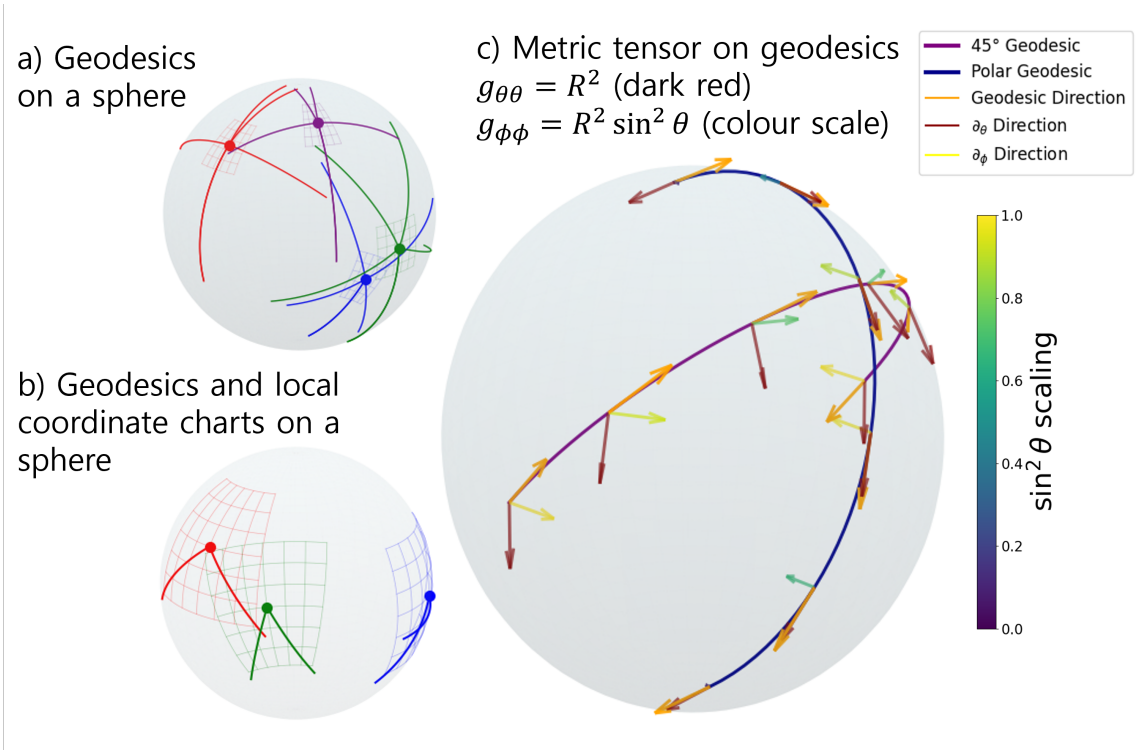


Figure 5.1. *Visualisations of a) geodesics on a spherical surface with radius R, b) distortion of coordinate charts on the sphere, and c) metric tensor components along the geodesic. Note that the colours on ∂_φ directions are changing according to the magnitude of $\sin^2\theta$ in c).*

By the metric being Riemannian, it means that for all vectors x and y, $\langle x, y\rangle_g = \langle y, x\rangle_g \geqslant 0$ with equality holding when $x = y$.

As an example, consider the surface of a sphere with radius R. In spherical coordinates (θ, ϕ), the metric takes the form

$$g_{\mu\nu} = \begin{pmatrix} g_{\theta\theta} & 0 \\ 0 & g_{\phi\phi} \end{pmatrix} = \begin{pmatrix} R^2 & 0 \\ 0 & R^2 \sin^2\theta \end{pmatrix}. \tag{5.9}$$

Unlike in flat space, the shortest path between two points–a geodesic–on this spherically curved surface is itself curved. As shown in Figure 5.1 a), geodesics connecting various points on the sphere do not appear as straight lines but rather as arcs on the sphere.

This curvature arises naturally from the spherical coordinate system as illustrated in Figure 5.1 b). The local coordinate grids shown in the figure reflect how distances and directions vary across the spherical surface.

The metric tensor in Equation (5.9) encodes this geometry. It determines the magnitude and direction of the coordinate basis vectors: changes in the polar angle θ correspond to the ∂_θ direction, while changes in the azimuthal angle φ correspond to the ∂_φ direction.

This is natural as the natural coordinate system on the sphere is curved–illustrated in Figure 5.1 b). The metric tensor as shown in Equation (5.9) naturally incorporates this fact–directions where θ changes, ∂_θ, and where φ changes, ∂_φ, according to Equation (5.9) give the natural direction and magnitude of the direction in which the geodesic is drawn. Figure 5.1 c) visualizes the metric components along the geodesics. These directions are scaled differently depending on location–in particular, $g_{\varphi\varphi}$ scales with $\sin^2\theta$, as illustrated by the colour scale in Figure 5.1 c).

The variation of the action $\delta S = 0$ can be identified with the variation of the length of the arc $\delta s = 0$, such that the extremisation of the length of the path becomes equivalent to the variational principle [10]. In this formulation, the physical trajectory obtained by extremising the action coincides with the geodesic obtained by minimising the arc length, thereby portraying the same physical content through the length functional.

Following Hamilton's principle of stationary action (Equation 5.5), variations δr_μ that vanish at the endpoints of the trajectory are considered:

$$\delta r_\mu(t_{\text{initial}}) = \delta r_\mu(t_{\text{final}}) = 0. \tag{5.10}$$

Since the total energy, $E = T + V$, of the system is conserved, the analysis is carried out on the fixed-energy surface $E = \text{Constant}$ for a natural Lagrangian $\mathcal{L} = T - V$ with T quadratic in velocities. In Maupertuis' principle the configuration endpoints $r(t_{\text{initial}})$, $r(t_{\text{final}})$ are fixed while the times t_{initial}, t_{final} are arbitrarily chosen; only the path in configuration space is varied. By the Euler's theorem

for degree-2 homogeneity, $\dot{r} \cdot \partial\mathcal{L}/\partial\dot{r} = 2T$, hence

$$\mathcal{S} = \int \frac{\partial \mathcal{L}}{\partial \dot{r}^{\mu}} \dot{r}^{\mu}\, dt = \int 2T\, dt = \int p_{\mu} dr^{\mu} \tag{5.11}$$

with $p = \partial\mathcal{L}/\partial\dot{r}$. Under the energy constraint $V(r) < E = \mathrm{Constant}$ along the path, the physical trajectory is characterised by $\delta\mathcal{S} = 0$ [10, 107].

Comparison of Equation 5.11 with Equation 5.6 shows that, under an appropriate metric, the condition $\delta\mathcal{S} = 0$ can be expressed as:

$$0 = \delta \int 2T\, dt = \delta \int \sqrt{g_{\mu\nu}\dot{r}^{\mu}\dot{r}^{\nu}}\, dt = \delta \int ds, \tag{5.12}$$

where the arc length element ds determines the physical trajectory in phase space [104].

5.1.1 Application to classical molecular systems via Jacobi metric

For a classical n_{atoms}-atom system, the potential energy V is taken to be twice continuously differentiable on the collision-free $3n_{\mathrm{atoms}}$-dimensional configuration space, with $V(r) \to +\infty$ as r approaches the collision set and $V(r) \to 0$ as $\|r\| \to \infty$. Let $E_{\mathrm{diss}} < \infty$ denote a dissociation threshold. When time is included as an external parameter, the state space is the extended space $Q \times \mathbb{R} \cong \mathbb{R}^{3n_{\mathrm{atoms}}+1}$, where the "$+1$" accounts for the time axis. For each fixed time t, the accessible configurations form the sublevel set

$$\mathcal{A}(t) \;=\; \big\{\, r \in Q :\; V(r(t)) \leqslant E_{\mathrm{diss}} \,\big\}. \tag{5.13}$$

Equivalently, in the extended space one may write

$$\mathcal{A}_{\mathrm{ext}} \;=\; \big\{\, (r,t) \in Q \times \mathbb{R} :\; V(r(t)) \leqslant E_{\mathrm{diss}} \,\big\}. \tag{5.14}$$

To keep notation light, $V(r)$ will be used when time is either held fixed or when the potential is time-independent; in this notation the "$+1$" in $\mathbb{R}^{3n_{\mathrm{atoms}}+1}$ simply reflects the inclusion of the time coordinate. Above the threshold E_{diss}, the molecule fails to retain its structure and dissociates into nonbonded atoms.

For a molecular system with symmetric, positive-definite mass tensor $m_{\mu\nu}(r)$ and natural Lagrangian $L = T - V$ with $T = \frac{1}{2}\, m_{\mu\nu}(r)\, \dot{r}^{\mu}\dot{r}^{\nu}$, fix a total energy level

$$E := H(r,p) = T + V(r), \tag{5.15}$$

determined by the initial condition. Along physical trajectories E is conserved (hence independent of r), and on the accessible region $V(r) < E$ one has $E - V(r) = T$.

On the fixed-energy surface $H = E$, the configuration-space metric is

$$g_{\mu\nu}(r) \;=\; 2\,[E - V(r)]\, m_{\mu\nu}(r), \tag{5.16}$$

with corresponding line element

$$ds \;=\; \sqrt{2\,[E - V(r)]\, m_{\mu\nu}(r)\, \dot{r}^{\mu}\dot{r}^{\nu}}\, dt. \tag{5.17}$$

Thus the stationarity of Maupertuis' action $S = \int p_{\mu}\, dr^{\mu}$ coincides (for fixed-energy variations) with the stationarity of arc length s—succinctly, $\delta S = 0 = \delta s$ [108].

This formulation embeds energy conservation into the geometry: the conformal factor $2\,[E - V(r)]$ restricts the metric to the accessible region $\{\, V < E \,\}$ and makes the length functional s explicitly energy-dependent. For a fixed energy E, physical trajectories coincide with geodesics of $g_{\mu\nu}$ in configuration space. The tensor $m_{\mu\nu}(r)$ collects position-dependent effective masses and kinetic couplings; diagonal entries encode atomic/effective masses, while off-diagonal terms capture couplings or anisotropic effects (e.g., curvilinear coordinates, constraints). Modifying masses or couplings amounts to perturbing $m_{\mu\nu}$, thereby deforming the metric and, consequently, the geodesic flow (i.e., the trajectories at that energy).

The space equipped with the Jacobi metric constitutes a Riemannian manifold, and the corresponding geodesic equations are expressed as:

$$\frac{dr^{\mu}}{ds^2} + \Gamma^{\mu}{}_{\nu\gamma}(r^{\nu})\frac{dr^{\nu}}{ds}\frac{dr^{\gamma}}{ds} = 0 \tag{5.18}$$

where ${\Gamma^{\mu}}_{\nu\gamma}$ represents the Christoffel symbols, defined by [109]:

$$ {\Gamma^{\mu}}_{\nu\gamma} = \tfrac{1}{2} g^{\mu\lambda} \left(\frac{\partial g_{\lambda\nu}}{\partial r^{\gamma}} + \frac{\partial g_{\lambda\gamma}}{\partial r^{\nu}} - \frac{\partial g_{\nu\gamma}}{\partial r^{\lambda}} \right). \tag{5.19} $$

Substituting the expression for the Jacobi metric into Equation (5.19), one can write

$$ \begin{aligned} {\Gamma^{\mu}}_{\nu\gamma} &= {\Gamma^{\mu}}_{\nu\gamma}[m] + \frac{1}{2T} \left({m^{\mu}}_{\nu} \frac{\partial T}{\partial r^{\gamma}} + {m^{\mu}}_{\gamma} \frac{\partial T}{\partial r^{\nu}} - m^{\mu\lambda} m_{\nu\gamma} \frac{\partial T}{\partial r^{\lambda}} \right) \\ &= {\Gamma^{\mu}}_{\nu\gamma}[m] + \frac{2}{T} m^{\mu\lambda} \left(m_{\nu\gamma} \frac{\partial V}{\partial r^{\lambda}} - m_{\lambda\nu} \frac{\partial V}{\partial r^{\gamma}} - m_{\lambda\gamma} \frac{\partial V}{\partial r^{\nu}} \right) \end{aligned} \tag{5.20} $$

where ${\Gamma^{\mu}}_{\nu\gamma}[m]$ is the Christoffel symbol computed by considering the mass matrix $m_{\mu\nu}$ as a metric using Equation (5.19), explicitly, it is defined as follows.

$$ {\Gamma^{\mu}}_{\nu\gamma}[m] = \tfrac{1}{2} m^{\mu\lambda} \left(\frac{\partial m_{\lambda\nu}}{\partial r^{\gamma}} + \frac{\partial m_{\lambda\gamma}}{\partial r^{\nu}} - \frac{\partial m_{\nu\gamma}}{\partial r^{\lambda}} \right) \tag{5.21} $$

It should be noted that in Equation 5.18, the coordinates including the time-variable are parametrized by the arc length s of the geodesic, as geodesics can invariably be parametrized by arc length [109]. This parametrization implies that the time-variable is now considered as a coordinate within the configuration space, or more precisely, within the Riemannian manifold on which it is defined, in terms of the arc length. While this conceptual framework has been successfully applied in the domains of celestial dynamics and cosmology, it represents a substantial departure from the conventional formulation of molecular dynamics.

On a flat manifold, ${\Gamma^{\mu}}_{\nu\gamma}[m]$ is zero, and noting that $T = \frac{1}{2} m_{\mu\nu} v^{\mu} v^{\nu} = \frac{1}{2} m^{\mu\nu} v_{\mu} v_{\nu}$, it becomes trivial that the geodesic equation is:

$$ \begin{aligned} 0 =& \frac{d^2 t}{ds^2} r^{\mu} + \left(\frac{dt}{ds} \right)^2 \frac{dr^{\mu}}{dt} \\ &+ \frac{2}{T} \left(-\tilde{m}^{\mu}{}_{\gamma} \frac{\partial V}{\partial r^{\nu}} - \tilde{m}^{\mu}{}_{\nu} \frac{\partial V}{\partial r^{\gamma}} + m^{\mu\lambda} m_{\nu\gamma} \frac{\partial V}{\partial r^{\gamma}} \right) \left(\frac{dt}{ds} \right)^2 \frac{dr^{\nu}}{dt} \frac{dr^{\gamma}}{dt} \end{aligned} \tag{5.22} $$

where $\tilde{m}^{\mu}{}_{\nu} = m^{\mu\lambda} m_{\lambda\nu}$. Now, noting that Ω is twice the kinetic energy and that $\frac{dt}{ds}$ is constant in a flat space, the equation reduces to the Newton's second law

once an identification is made that the negative gradient of the potential is the *force*.

While the geodesic formulation may look more abstract, it yields a robust framework for numerically stable integration on a fixed energy surface. The geodesic equations determine the *path* in configuration space up to reparametrisation. Physical time is then recovered by a local quadrature along that path: using

$$ds^2 = 2\left[E - V(r)\right] m_{\mu\nu}(r)\, dr^\mu dr^\nu \tag{5.23}$$

and $T = E - V(r)$ along true motions, one obtains

$$ds \;=\; 2\left[E - V(r)\right] dt \quad \Longrightarrow \quad dt \;=\; \frac{ds}{2\left[E - V(r)\right]}. \tag{5.24}$$

Thus dt is *not* chosen independently but is induced by the arc-length increment ds together with the conserved energy through the factor $E - V(r)$. In practice, integrating geodesics at fixed E enforces energy conservation by construction (within discretisation error), and the physical time parameter is recovered via the above relation.

Conceptually, this formulation is related to adaptive time-stepping algorithms commonly employed in molecular dynamics simulations to ensure stability and accuracy in regions of rapid dynamical change [110, 111]. Such algorithms typically rely on explicit energy evaluation and local error estimators to adaptively control the integration step size, introducing computational overhead and algorithmic complexity. In contrast, the geodesic method bypasses the need for energy-based timestep control altogether. Because the evolution parameter ds is inherently tied to the system's kinetic energy through the $E - V$ term, the method maintains stability and conserves energy intrinsically, without the need for external correction mechanisms.

Unlike conventional adaptive schemes that rescale the timestep based on error estimates or force magnitudes [110], the geodesic formulation embeds the energy constraint into the motion itself by evolving curves at fixed E with

respect to the length element $ds^2 = 2\,[E - V(r)]\,m_{\mu\nu}(r)\,dr^\mu dr^\nu$. It is useful to contrast this with *symplectic* integrators (e.g., velocity–Verlet, Störmer–Cowell), which preserve the symplectic 2-form and therefore exhibit near-conservation of a modified Hamiltonian over very long times (backward-error sense), but do not enforce $H = E$ exactly at each step and require a prescribed Δt.

In the geodesic approach, step control is performed in arc length s rather than in time, and the reconstruction $dt = ds/\big(2\,[E - V(r)]\big)$ ties the effective timestep to the conserved energy, removing the need for separate energy monitors or dynamic timestep heuristics. This can enhance long-time stability in stiff or multiscale regimes by preventing energy drift through construction. A caveat is that the configuration-space geodesic discretisation is not symplectic in phase space unless lifted via a constrained (e.g., RATTLE-type) or variational discretisation; symplecticity yields excellent long-time phase-space fidelity, whereas geodesic marching yields exact (up to quadrature) energy control. In practice the two viewpoints are complementary: one may (i) use a discrete Maupertuis variational scheme to evolve directly on a prescribed energy level E, or (ii) employ a standard symplectic integrator with a lightweight post–step normalisation to the target energy (e.g., a scalar solve for a time–rescaling or a velocity rescaling that restores the prescribed energy to within tolerance). The latter retains the symplectic method's favourable phase–space behaviour while controlling long–time energy drift, acknowledging a slight loss of exact symplecticity due to the correction.

Similarly, constrained algorithms such as RATTLE [112] and SHAKE [113, 114], commonly used to enforce holonomic constraints in molecular dynamics [115], introduce numerical corrections through iterative projection steps. These methods can accumulate constraint errors over time and are sensitive to timestep size and convergence tolerances. The geodesic approach, on the other hand, makes it possible to naturally enforce such constraints by formulating the dynamics on a curved manifold where the constraints are embedded in the geometry of the motion. This ensures that both energy and conformational

constraints are preserved simultaneously to machine precision, without the need for post-integration corrections.

5.1.2 Compactness of the configurational manifold of a classical molecule

Given Equations (5.16) and (5.18), one can learn that the set of all coordinates the molecule under study can take are in the set called *Hill's region* [116, 117]:

$$R_{\text{Hill}} = \{r : V(r) \leqslant E\}, \tag{5.25}$$

where E is the total energy of the system. Hill's region is defined in a $3n_{\text{atoms}} + 1$ dimensional space as each atom contributes 3 dimensions and the time-variable contributes an additional dimension to the system. Note that the Hill's region is not a vector space, unless $E \to \infty$.

If the molecule is globally minimized prior to the main simulation, the system has a finite minimum allowed potential, denoted in this work as $V_{\min}$. This means that the Hill's region for the system can be defined as

$$\mathcal{M}^{cl} = \{r : V_{\min} \leqslant V(r) \leqslant E\}. \tag{5.26}$$

In this section, two statements: one implying that $\overline{\mathcal{M}}^{cl}$ is compact and its interior,

$$\mathcal{M}^{cl}_{\text{int}} = \{r : V_{\min} < V(r) < E\} \tag{5.27}$$

is a Riemannian manifold with compact closure. The superscript cl denotes that the manifold is composed of the coordinates allowed by the classical molecular dynamics simulations.

Lemma 5.1. *$\mathcal{M}^{cl}$ is a compact subset of $\mathbb{R}^{3n_{\text{atoms}}+1}$.*

Proof. Since V is continuous wherever it is finite, and points with $V(r) = +\infty$ are automaticallt excluded by the condition $V(r) < E$, the preimage

$$\mathcal{M}^{cl} = V^{-1}([V_{\min}, E]) \tag{5.28}$$

is closed, as the preimage of a closed interval under a continuous map is closed.

For all $E < E_{\text{diss}}$, there exists $R < \infty$ such that $\|r\| \geqslant R \implies V(r) > E$. Hence, for all r with $V(r) < E$ must satisfy $\|r\| < R$; i.e.,

$$\mathcal{M}^{cl} \subset \overline{B}_R(0) = \left\{x \in \mathbb{R}^{3n_{\text{atoms}}+1} : \|x\| \leqslant R\right\}, \tag{5.29}$$

so $\mathcal{M}^{cl}$ is bounded.

Heine-Borel theorem implies that a subset is compact if and only if it is closed and bounded. Therefore $\overline{\mathcal{M}}$ is compact. □

Recognizing that $\mathcal{M}^{cl}$ is compact is crucial, because this compactness is directly tied to the stability of the trajectories [116]. Using this lemma, Theorem 5.2 can be proved.

Theorem 5.2. *Suppose $\nabla V(r) \neq 0$ when $V \in \{V_{\min}, E\}$, then $\mathcal{M}^{cl}_{\text{int}}$ with the Jacobi metric is a $3n_{\text{atoms}} + 1$ dimensional Riemannian manifold with compact closure*.

Proof. Since V is continuous on $\{V < \infty\}$, the preimage

$$\mathcal{M}^{cl}_{\text{int}} = V^{-1}((V_{\min}, E)) \tag{5.30}$$

is open in $\mathbb{R}^{3n_{\text{atoms}}+1}$. Therefore $\mathcal{M}^{cl}_{\text{int}}$, with the subspace charts inherited from $\mathbb{R}^{3n_{\text{atoms}}+1}$, is a $3n_{\text{atoms}} + 1$ dimensional smooth manifold.

On $\mathcal{M}^{cl}_{\text{int}}$ we have $E - V(r) > 0$. The map $r \mapsto E - V(r)$ is $\mathcal{C}^2$ on $\mathcal{M}^{cl}_{\text{int}}$ because V is $\mathcal{C}^2$ where it is finite and $\mathcal{M}^{cl}_{\text{int}} \subset \{V < \infty\}$ is opne. Since the mass tensor m is smooth and positive definite on $\mathbb{R}^{3n_{\text{atoms}}+1}$, their pointwise product

$$g_{\mu\nu}(r) = (E - V(r))m_{\mu\nu}(r) \tag{5.31}$$

is $\mathcal{C}^2$ on $\mathcal{M}^{cl}_{\text{int}}$. Moreover, for any $r \in \mathcal{M}^{cl}_{\text{int}}$ and $0 \neq v \in T_r\mathbb{R}^{3n_{\text{atoms}}+1}$,

$$g_{\mu\nu}v^\mu v^\nu = (E - V(r))m_{\mu\nu}v^\mu v^\nu > 0, \tag{5.32}$$

so the metric g is positive definite. Hence, $\left(\mathcal{M}^{cl}_{\text{int}}, g\right)$ is a Riemannian manifold.

Now, assume in addition that E and $V_{\min}$ are regular values of V, i.e. $\nabla V \neq 0$ on $V^{-1}(E) \cup V^{-1}(V_{\min})$. The claim to show now is that $\overline{\mathcal{M}^{cl}_{\text{int}}} = \mathcal{M}^{cl}$. The inclusion $\overline{\mathcal{M}^{cl}_{\text{int}}} \subset \mathcal{M}^{cl}$ is trivial since $\mathcal{M}^{cl}_{\text{int}} = V^{-1}((V_{\min}, E)) \subset V^{-1}([V_{\min}, E]) = \mathcal{M}^{cl}$ and the closedness of $\mathcal{M}^{cl}$.

To establish the equality $\overline{\mathcal{M}^{cl}_{\text{int}}} = \mathcal{M}^{cl}$, only thing left to show is $\mathcal{M}^{cl} \subset \overline{\mathcal{M}^{cl}_{\text{int}}}$. If $V_{\min} < V(r) < E$, the result follows immediately from the definition of $\mathcal{M}^{cl}$.

If $V(r) = E$, by the regular-value assumption, $\nabla V(r) \neq 0$. Choose w with $\langle \nabla V(r), w \rangle_g > 0$. For sufficiently small $t > 0$, the Taylor expansion yields

$$V(r - tw) = V(r) - t\langle \nabla V(r), w \rangle + o(t) < E, \tag{5.33}$$

and by the continuity of V, $V(r - tw) > V_{\min}$ for all small $t > 0$. Therefore, $r_t := r - tw \in \mathcal{M}^{cl}_{\text{int}}$ and $r_t \to r$, so $r \in \overline{\mathcal{M}^{cl}_{\text{int}}}$.

If $V(r) = V_{\min}$, then, again, $\nabla V(r) \neq 0$. Choose w with $\langle \nabla V(r), w \rangle_g < 0$. For sufficiently small $t > 0$, as done in the previous case, one can write

$$V(r + tw) = V(r) + t\langle \nabla V(r), w \rangle_g + o(t) > V_{\min}, \tag{5.34}$$

and still $V(r + tw) < E$ for all small $t > 0$. Thus, $r_t := r + tw \in \mathcal{M}^{cl}_{\text{int}}$ and $r_t \to r \in \mathcal{M}^{cl}_{\text{int}}$.

Thus, for all $r \in \mathcal{M}^{cl}$, one can conclude that $r \in \overline{\mathcal{M}^{cl}_{\text{int}}}$, proving that $\mathcal{M}^{cl} \subset \overline{\mathcal{M}^{cl}_{\text{int}}}$. Therefore,

$$\overline{\mathcal{M}^{cl}_{\text{int}}} = \mathcal{M}^{cl}. \tag{5.35}$$

This is the last part to be proved. □

5.1.3 When $V(r) = E$

At the Hill boundary, $V(r) = E$, the Jacobi metric becomes singular as the factor $2(E-V)$ vanishes there [116,117]. For example, for the one-dimensional harmonic oscillator with equilibrium position r_0 and frequency ω,

$$H^{\text{harm}} = \tfrac{1}{2}v^2 + \tfrac{1}{2}\omega^2(r - r_0)^2, \tag{5.36}$$

one finds

$$ds = \sqrt{2\big(E - \tfrac{1}{2}\omega^2(r - r_0)^2\big)}dt \longrightarrow 0 \quad \text{as } r \to r_0 \ \& \ E - V \to 0, \tag{5.37}$$

so that an infinite arc-length s would be required mererly to reach the turning point. This is an artefact of using s as parameter: physically, the oscillator turns around in finite time.

This degeneracy was resolved by Herbert Seifert in 1948 via a *reflection argument* on the Riemannian manifold with compact closure [118]. Rather than trying to parametrize through the singular boundary in s, one works in Newtonian time t and *expands* the trajectory about the turning point. Concretely, let r_0 be a regular boundary point with $V(r_0) = E$ and $\nabla V(r_0) \neq 0$. Introduce local coordinates $\big(r^1, \cdots, r^{d-1}, y\big)$ such that $y = 0 \Leftrightarrow V(r) = E$ so that y measures the distance normal to the Hill surface. A solution that arrives at r_0 with zero velocity satisfies

$$\dot{r}^\mu(0) = 0, \quad \ddot{r}^\mu(0) = -\partial^\mu V(r_0) \tag{5.38}$$

and hence admits the Taylor expansion in t, i.e.

$$r^\mu(t) = {r_0}^\mu + \tfrac{1}{2}\ddot{r}^\mu(0)t^2 + O(t^3) = {r_0}^\mu - \tfrac{1}{2}\partial^\mu V(r_0)t^2 + O(t^3). \tag{5.39}$$

Although higher odd terms may appear at $O(t^3)$ and beyond, the expansion is even up to quadratic order, so for small t

$$r^\mu(t) = r^\mu(-t) + O\big(|t|^3\big). \tag{5.40}$$

Consequently, the segment for $t < 0$ is the time-reflection of the segment for $t > 0$ to second order: the reflected curve $\tilde{r}(t) := r(|t|)$ satisfies $\tilde{r}(t) = r(t) + O(|t|^3)$ and $\dot{\tilde{r}}(0) = 0$. Hence the trajectory *reflects* smoothly off the Hill boundary up to quadratic order.

The trajectory arrives at r_0, comes to rest, and the departs along the same path in reverse, without traversing any finite arc length induced by the Jacobi metric in the singular metric. By reparameterizing each smooth branch of

the trajectory in terms of its Jacobi arclength s, one can cast the reflection construction directly in the s–parameterisation.

As an illustration, consider a simple harmonic model of a fictitious linear triatomic molecule (like carbon dioxide) with three vibrational modes. In mass-weighted normal coordinates, (q_a, q_b, q_c), corresponding to bending, symmetric and antisymmetric stretch modes, respectively, the harmonic potential can be written as

$$V(q_a, q_b, q_c) = \tfrac{1}{2}{\omega_a}^2{q_a}^2 + \tfrac{1}{2}{\omega_b}^2{q_b}^2 + \tfrac{1}{2}{\omega_c}^2{q_c}^2 \tag{5.41}$$

with mode frequencies ω_a, ω_b and ω_c. The Hill's region $\{(q_a, q_b, q_c) : V(q_a, q_b, q_c) \leqslant E\}$ is the interior of the ellipsoid

$${\omega_a}^2{q_a}^2 + {\omega_b}^2{q_b}^2 + {\omega_c}^2{q_c}^2 = 2E. \tag{5.42}$$

On its boundary where $V = E$, the factor $2(E - V)$ vanishes and $ds^2 \to 0$. Now, on this boundary pick a regular turning point $(q_a^0, q_b^0, q_c^0) = \left(0, 0, \sqrt{2E}/\omega_c\right)$ such that $\nabla V \neq 0$. In Newtonian time t, the trajectory satisfies

$$\dot{q}_i(0) = 0, \quad \ddot{q}_i(0) = -\partial_{q_i} V(q^0), \quad i \in \{a, b, c\}. \tag{5.43}$$

Hence, each coordinate admits an even Taylor series about $t = 0$ in the form given by Equation (5.39). Now, for the pure antisymmetric turning point at $(0, 0, \sqrt{2E}/\omega_c)$, one gets

$$q_c(t) = \sqrt{\frac{2E}{{\omega_c}^2}} - \omega_c\sqrt{\frac{E}{2}}t^2 + O(t^3) \tag{5.44}$$

with $q_a(t) = q_b(t) = O(t^3)$.

Now, to complete the reflection construction, one proceeds as follows:

(i) Split into inbound/outbound branches. Define

$$q_i^{\pm}(t) = q_i(\pm t), \qquad t \geqslant 0, \quad i \in \{a, b, c\}, \tag{5.45}$$

so that $q_i^{+}(t)$ is the trajectory after the turning point and $q_i^{-}(t)$ its time–reflection before. Each branch, obtained by the quadratic approximation given in

Equation (5.40), is C^∞ in t and meets the boundary at $t = 0$ with zero velocity. Note that at $t = 0$, $s = 0$ and due to the continuity,

$$q_i^+(0) = q_i^-(0) \tag{5.46}$$

is expected.

(ii) Introduce Jacobi arclength on each branch. On, say, the "$+$" branch, set

$$s(t) = \int_0^t \sqrt{2\big[E - V\big(q^+(\tau)\big)\big]}\, d\tau, \qquad s \geqslant 0. \tag{5.47}$$

Near the turning point one finds

$$E - V\big(q^+(t)\big) \approx \tfrac{1}{2}{\omega_c}^4 \big(q_c^0\big)^2 t^2, \tag{5.48}$$

so that

$$ds \;\approx\; \frac{\omega_c^2\, q_c^0}{\sqrt{2}}\, |t|\, dt \implies s(t) \;=\; \tfrac{1}{2}\, C\, t^2 + O(t^4), \quad C := \tfrac{\omega_c^2\, q_c^0}{\sqrt{2}}. \tag{5.49}$$

Hence $s(t)$ is strictly increasing for $t > 0$, with

$$t(s) \;=\; \sqrt{\frac{2\, s}{C}} + O(s^{3/2}). \tag{5.50}$$

(iii) Reparametrise the coordinates in s. Substitute $t = t(s)$ into the Taylor expansion for the active coordinate (here the c-mode):

$$q_c^+(s) = q_c^0 - \tfrac{1}{2}\omega_c^2 q_c^0 t(s)^2 + O\big(t(s)^4\big) = q_c^0 - \frac{\omega_c^2\, q_c^0}{C} s + O(s^2), \tag{5.51}$$

and similarly for $q_a^+(s)$, $q_b^+(s)$, each analytic in s near $s = 0$. Thus the "$+$" branch becomes a smooth, s–parameterised geodesic emanating from the turning point.

(iv) Glue the two branches at $s = 0$. Define the full reflected trajectory in the Jacobi–arclength parameter by

$$q_i(s) = \begin{cases} q_i^+(t(s)), & s \geqslant 0, \\ q_i^-\big(t(-s)\big), & s \leqslant 0, \end{cases} \tag{5.52}$$

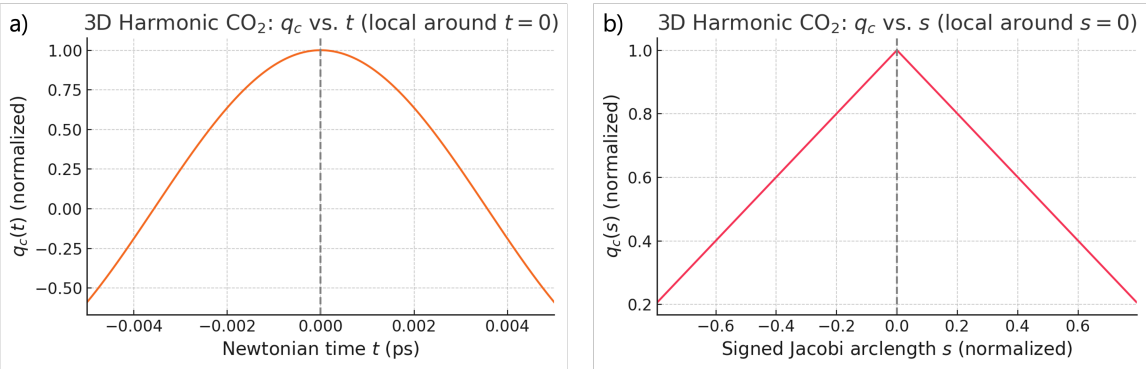


Figure 5.2. (*a*) *Antisymmetric-stretch coordinate* $q_c(t)$ *of the 3-mode* CO_2 *harmonic model, showing a smooth turnaround at the turning point* $t = 0$. (*b*) *The same trajectory re-parameterized by signed Jacobi arclength* s*, with* $s = 0$ *at the Hill boundary. The two linear segments meet at* $s = 0$*, illustrating the geodesic "reflection" in the* q_c *direction.* $\omega_c = 440/\mathrm{ps}$*, derived from the carbon dioxide's antisymmetric strech mode vibrational frequency of* $2349/\mathrm{cm}$ *was used for the simulation.*

so that $q_i(0) = q_i^0$ and dq_i/ds is continuous at $s = 0$. This yields a single, smooth geodesic in the (q_a, q_b, q_c)-ellipsoid which "reflects" off the Hill boundary at $s = 0$, fully bypassing the singularity of the Jacobi metric there.

Figure 5.2 (a) shows the antisymmetric stretch mode $q_c(t) = A\cos(\omega_c t)$, which approaches the turning point at $t = 0$ with zero velocity and then departs along the same path by time-reversal symmetry. A is the amplitude that is normalised to 1 for simplicity. In panel (b) of Figure 5.2, the arc length parameter s is computed via Equation (5.47), assigning $s > 0$ for $t > 0$ and $s < 0$ for $t < 0$. Although $ds \to 0$ at the turning point, the two branches plotted against signed s form straight lines of slope $\pm\frac{{\omega_c}^2 A}{\sqrt{2}}$ meeting at $s = 0$. This linear-in-s behaviour near the boundary confirms Seifert's reflection construction: the geodesic in the Jacobi metric *bounces off* the Hill surface without ever traversing a finite singular arc length.

5.2 Adding quantum mechanical uncertainties as a thermostat

In conventional molecular dynamics, a molecule's trajectory is defined as a Hamiltonian flow over a symplectic manifold. Under such framework, a thermostat is introduced to control the temperature of the system, ensuring that the kinetic energy samples a distribution consistent with a target temperature. However, when the dynamics are reformulated in terms of the geodesic flow induced by the Jacobi metric, these standard approaches encounter significant conceptual and practical challenges as the Hamiltonian flow over a symplectic manifold is now described as a geodesic flow over a Riemannian manifold. Therefore, representing dynamics purely as geodesic motion is of limited use without also prescribing a means of thermostatting to recover canonical sampling. In very large, complex systems one may rely on the bulk of degrees of freedom to act as an implicit heat bath, but a concrete thermostatting mechanism compatible with the Riemannian framework must still be devised.

In geodesic formulation, the dynamics are defined on the tangent bundle $T\mathcal{M}$ of the configuration manifold $\mathcal{M}$ via the Jacobi metric

$$g_{\mu\nu} = 2(E - V)m_{\mu\nu} \tag{5.53}$$

which encodes both the mass tensor and the potential energy landscape. The corresponding relation between the arc-length parameter, s, and the physical time t is then

$$dt = \frac{ds}{\sqrt{2(E - V(r))\, m_{\mu\nu}\, \dot{r}^{\mu}\dot{r}^{\nu}}}\,, \tag{5.54}$$

ensuring that time evolution is consistent with the kinetic energy defined by the metric. This formulation inherently lacks a fixed time scale due to its reparameterisation invariance and energy dependence, presenting challenges for incorporating standard thermostatting methods.

In this framework, the physical velocity v^{μ}_{phys} and the Jacobi (arc-length parametrised) velocity v^{μ}_{jac} are related by

$$v^{\mu}_{\text{jac}} = \frac{v^{\mu}_{\text{phys}}}{\sqrt{2(E - V(r))\, m_{\alpha\beta}\, v^{\alpha}_{\text{phys}}\, v^{\beta}_{\text{phys}}}} \,. \tag{5.55}$$

This ensures that the arc-length parametrisation remains consistent with the kinetic energy defined by the mass tensor and the potential energy landscape.

Thermostats such as Langevin or Nosé-Hoover are designed within a Newtonian framework, where time is fixed and the kinetic energy metric remains constant. Their friction and stochastic forces are defined with respect to a physical time parameter, making them incompatible with the geodesic dynamics governed by the Jacobi metric. In the Jacobi formulation, the metric as shown in Equation (5.53) varies with time as the potential energy evolves (i.e., according to the positions) and the natural parameter is the arc-length rather than physical time.

Naively introducing conventional thermostats into this framework leads to numerical instabilities and incorrect trajectories for two main reasons. Firstly, standard thermostats assume a fixed, flat, time-independent metric, the Jacobi metric's time dependence means that these methods cannot naturally incorporate the arc-length parameter inherent to the geodesic description.

Secondly, the discrete updates applied to the geodesic flow fail to preserve the symplectic properties of the underlying dynamics. As highlighted in Casetti et al. [104], proper numerical integration of geodesic flows must respect the inherent geometric invariants. Standard thermostat algorithms, when applied to the geodesic formulation, break these invariants by introducing corrections that do not conserve the natural structure of the flow. In high-dimensional systems, even small deviations from the correct geometric behavior can accumulate, resulting in significant numerical errors and trajectories that deviate from the true dynamics.

5.2.1 A formulation of an uncertainty-driven thermostat

The central idea behind our uncertainty-driven thermostat is to enforce a finite-temperature condition by comparing the system's instantaneous momentum with an uncertainty threshold derived from quantum considerations. This work employs the finite-temperature uncertainty function introduced by Bei-Lok Hu and Yuhong Zhang in 1993 [119]. At finite temperature and time, the uncertainty relation giving a phase-space spread as a function of time t is written as

$$\Delta x^2 \Delta p^2 \geqslant U_T(t) \tag{5.56}$$

with

$$U_T(t) = \frac{\hbar^2}{4}\Big[\exp(-\gamma_0 t) + \coth(\varepsilon)\Big(1 - \exp(-\gamma_0 t)\Big)\Big]^2, \tag{5.57}$$

where

$$\varepsilon = \frac{\hbar\,\Omega_0}{2\,k_B\,T}\,, \tag{5.58}$$

γ_0 is a damping rate, Ω_0 a characteristic frequency, k_B the Boltzmann constant, and T the target temperature. The plot of U_T is graphically shown in Figure 5.3.

In the present formalism, a typical positional uncertainty Δx_i, for each atom i, is estimated as half the average bond lengths relating to the atom i, and the magnitude of the momentum is computed from the physical velocities (with $p_i = m_i v^i_{\text{phys}}$, Einstein summation convention is not applied in this section). The thermostat is then implemented by comparing the product $\Delta x_i |p_i|$, with the threshold $U_T(t)$. If

$$\Delta x_i |p_i| < \sqrt{|U_T(t)|}, \tag{5.59}$$

a corrective random momentum increment is injected so as to raise the momentum magnitude, thereby ensuring that the system samples a phase space region consistent with the target temperature. Mathematically, if the current momentum magnitude is $|p_i|$, then a momentum increment δp_i is determined

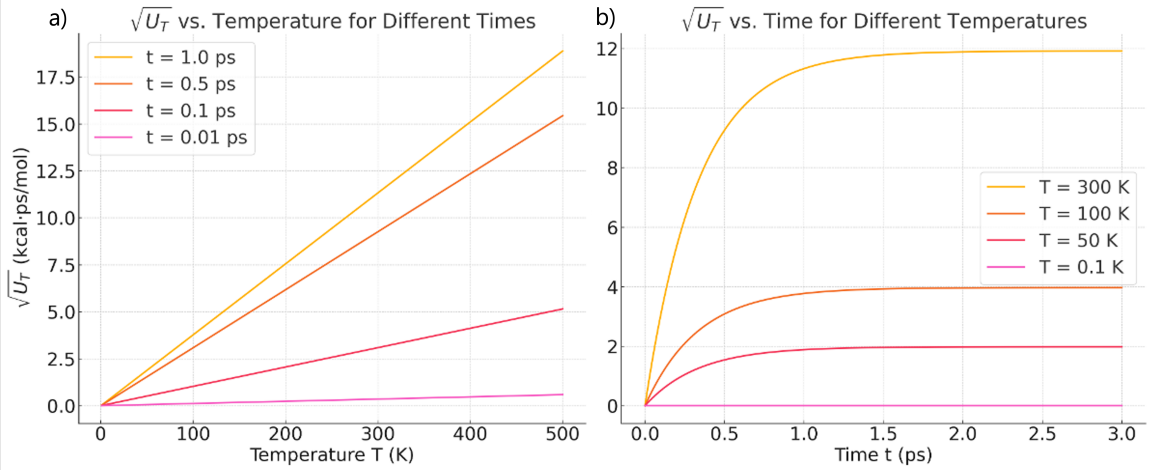


Figure 5.3. (*a*) *Plot of the quantity* $\sqrt{U_T}$ *over different temperatures for fixed times*. (*b*) *Plot of* $\sqrt{U_T}$ *over time with fixed temperatures*. $\gamma_0 = 3.0\,\mathrm{ps}$ *and* $\Omega_0 = 0.05\,\mathrm{ps}$.

such that

$$|p| + \delta p \approx \frac{\sqrt{|U_T(t)|}}{\Delta x_i}, \tag{5.60}$$

with the increment applied along a random direction. This correction is carried out in the physical velocity space and, after adjustment, the new physical velocities are converted back to the Jacobi representation using Equation (5.55).

5.2.2 Adaptive Friction and Energy Regulation

In addition to the uncertainty correction, an adaptive friction term is introduced to remove any excess kinetic energy and thereby enforce the target temperature. The friction coefficient is adjusted based on the instantaneous temperature T_{current}, estimated via equipartition

$$T_{\text{current}} = \frac{2\,\mathrm{KE}}{D\,k_B}, \tag{5.61}$$

where D is the number of degrees of freedom and KE is the kinetic energy of the system. The effective friction coefficient is given by

$$\gamma_{\text{eff}} = \gamma_0 \frac{T_{\text{current}} - T}{T}. \tag{5.62}$$

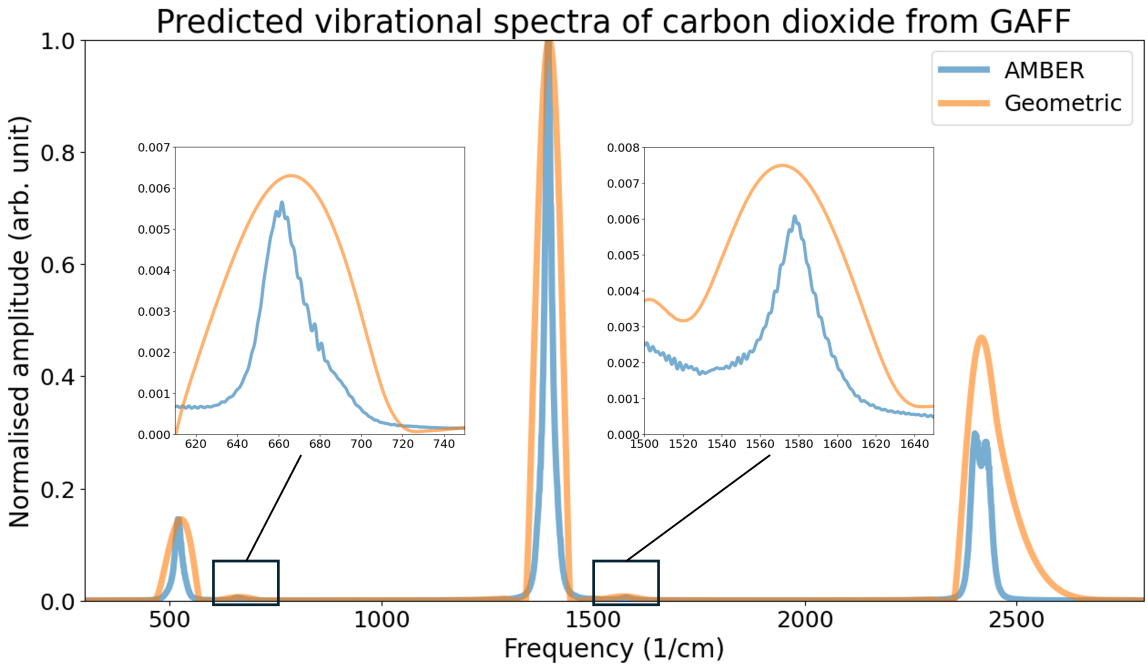


Figure 5.4. *Vibrational spectra of carbon dioxide obtained from the trajectories obtained via the geometric method introduced in this chapter and AMBER molecular dynamics simulation.*

Then, the physical velocities are damped via an exponential decay,

$$v^{i\,\mathrm{new}}_{\mathrm{phys}} = v^{i}_{\mathrm{phys}} \exp(-\gamma_{\mathrm{eff}}\, dt), \tag{5.63}$$

with dt computed from Equation (5.54). The damped velocities are subsequently converted back to the Jacobi frame using Equation (5.55).

The total energy E is updated accordingly to account for both the uncertainty-driven momentum injection and the frictional energy dissipation. This adaptive scheme ensures that the system remains close to the desired thermal state throughout the simulation.

5.3 A case study with carbon dioxide

The carbon dioxide molecule is simulated in vacuum using the geometric method introduced in this work. As a reference, a conventional AMBER molecular dynamics simulation is also performed to benchmark the results of the geometric approach.

In this work, the damping rate is set to $\gamma_0 = 3.0\,\mathrm{ps}^{-1}$, providing a gentle frictional force that prevents excessive energy removal and preserves the near-conservative behavior of the system. Note that if the thermostat were Langevin, this damping coefficient would be considered very high.

The characteristic frequency, Ω_0, is determined by a normal mode analysis of the equilibrium configuration, which yields the intrinsic vibrational frequencies of the molecule. An effective value of $\Omega_0 = 0.05\,\mathrm{ps}$ is selected to capture the dominant high-frequency stretching modes of carbon dioxide, thereby setting the intrinsic timescale of the system. To compute the potential, General AMBER Force Field (GAFF) parameters [60] are used.

A reference AMBER molecular dynamics simulation for carbon dioxide was carried out using GAFF parameters. The system was first minimized to relieve any unfavorable contacts, then gradually heated to the target temperature, and finally equilibrated and simulated in production under an NPT ensemble at $300\,\mathrm{K}$ and $1\,\mathrm{atm}$. Temperature control was achieved via a Langevin thermostat with damping parameter of $\gamma = 1/\mathrm{ps}$. This conventional AMBER trajectory was used as a benchmark for comparison with the trajectories obtained via the geometric method.

Figure 5.4 illustrates the close agreement between the vibrational spectra obtained using the conventional approach—specifically, the AMBER molecular dynamics package—and the geometric method introduced in this chapter. In classical molecular dynamics, the time step is typically chosen based on empirical stability criteria rather than physical principles, and is often kept fixed throughout the simulation. As a result, obtaining a meaningful velocity

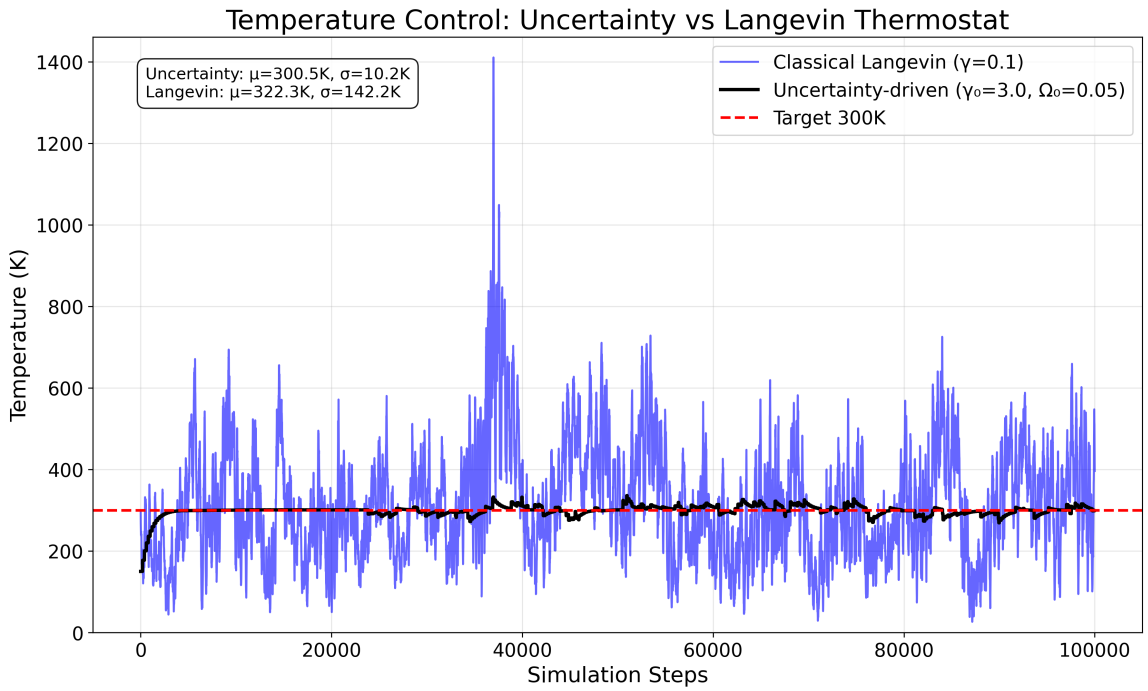


Figure 5.5. *Temperature change over iterations plot. Uncertainty-driven thermostat was applied to a carbon dioxide. The target temeperature is* $300\,\mathrm{K}$.

autocorrelation function generally requires long simulation times to ensure adequate sampling and noise reduction. The AMBER simulation used for Figure 5.4 was run for $10\,\mathrm{ns}$ with a fixed time step of $1\,\mathrm{fs}$. The trajectory obtained using the geometric framework, which produced the spectra in Figure 5.4, employed uniform arc-length steps of $0.001\,\mathrm{amu}\cdot{}^{2}\mathrm{ps}^{-1}$; however, the corresponding physical time steps vary, since ds/dt changes according to the energy landscape traversed by the molecule. Owing to this property, the geometric framework required only 100 ps of simulation to recover accurate vibrational spectra, as demonstrated by the method introduced in this chapter.

The peak locations in Figure 5.4 match between the two methods, as expected, since both simulations use the GAFF force field. However, the peaks obtained from the method introduced in this chapter are broader. This is due to the interpolation error in re-writing the trajectory in the picosecond units

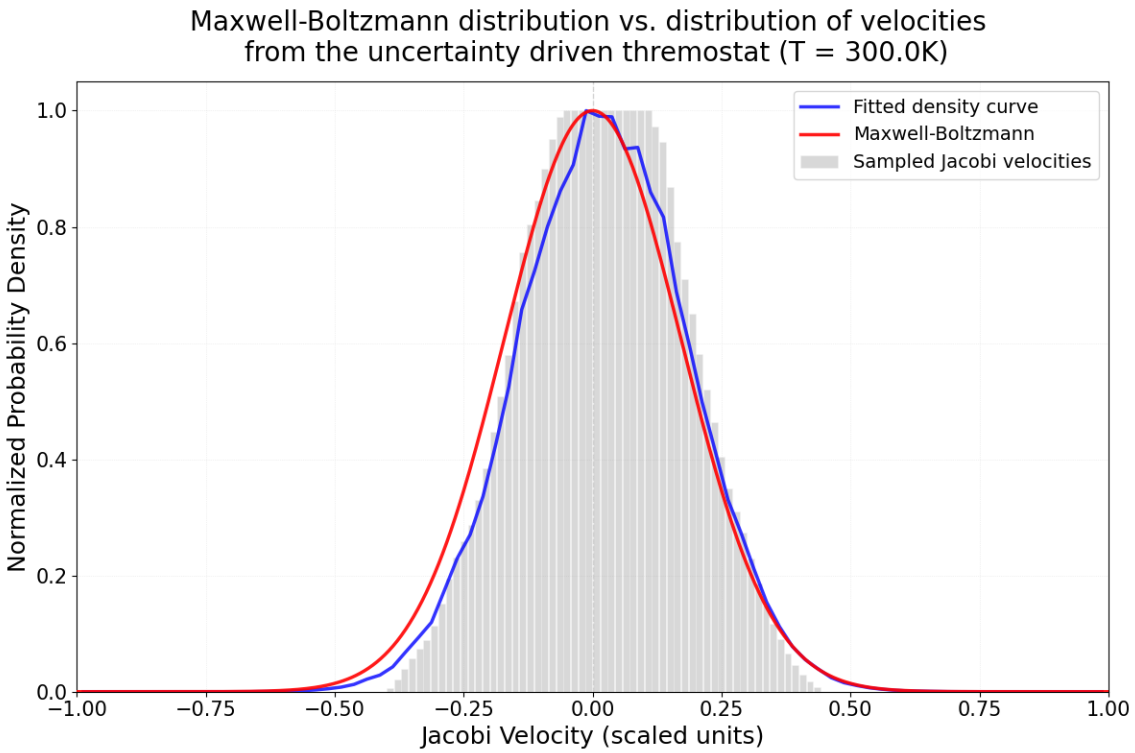


Figure 5.6. *Comparison of the Maxwell–Boltzmann velocity distribution at* $300\,\mathrm{K}$ *with the velocity histogram obtained from the uncertainty-driven thermostat at the same temperature for a carbon dioxide molecule, sampled over* $1\,\mathrm{ns}$ *with a* $0.01\,\mathrm{fs}$ *time step. The histogram comprises 700 bins, and the* x*-axis denotes the Jacobi velocity in units scaled to fit the range* $[-1.0, 1.0]$*. All distributions are normalized to unit maximum for direct comparison.*

from the arc length units.

Figure 5.5 shows how the uncertainty-driven thermostat successfully maintains the temperature of 300K. Unlike Langevin or Nosé-Hoover thermostats, the fluctuation is not violent.

To assess whether the sampled velocities are physically correct, Figure 5.6 shows the Jacobi velocity statistics (grey histogram), together with a fitted density curve (blue) compared against the theoretical Maxwell–Boltzmann distribution (red), which provides the expected equilibrium velocity distribution

at a given temperature [120]. The agreement between the sampled data and the Maxwell–Boltzmann curve demonstrates that the uncertainty-driven thermostat correctly reproduces the equilibrium distribution. Small deviations in the tails are attributable to finite sampling effects and binning resolution.

5.4 Normal mode analysis and the Riemannian framework

Under the traditional formalism of classical molecular dynamics, a system's normal modes can be obtained by diagonalisation of the mass-weighted Hessian matrix at an energy-minimised geometry. These normal modes define the orthogonal vibrational directions of the molecule—being given by the eigenvectors of the mass-weighted Hessian [121, 122]—but this construction is valid only at the equilibrium configuration, demanding that the Taylor expansion of the potential about a point r is given as:

$$V(r+\eta) = V(r) + \tfrac{1}{2}\eta^{\nu}\mathrm{Hess}_{\nu\kappa}(r)\eta^{\kappa} + \mathcal{O}\big(|\eta|^{3}\big), \tag{5.64}$$

and this approximation does not deviate too far from the true values of the potential for small perturbations η. At any instantaneous configuration $r \in \mathcal{M}^{cl}$, one can diagonalise the Hessian

$$\mathrm{Hess}_{\mu\nu}(r) = \frac{\partial^{2} V}{\partial r^{\mu}\partial r^{\nu}}(r) \tag{5.65}$$

to obtain the so-called instantaneous normal modes. However, outside a true minimum of the potential, some Hessian eigenvalues become negative–yielding imaginary frequencies–and the resulting eigenvectors do not in general correspond to stable oscillations [123]. Moreover, performing a full diagonalisation at each molecular dynamics snapshot scales as $\mathcal{O}(n_{\mathrm{atoms}}^{3})$ and quickly becomes computationally prohibitive for large systems, so instantaneous normal modes do not offer a robust or efficient route to isolating stable vibrational motions during on-the-fly dynamics.

With Riemannian geometry, however, one can track the normal modes via the eigenvectors of the quantity called the *Ricci tensor*.

5.4.1 Ricci tensor and normal mode analysis

The Ricci tensor of a system in Riemmanian manifold is defined as:

$$\mathrm{Ric}_{\mu\nu}(r) = \frac{\partial \Gamma^{\kappa}{}_{\mu\nu}}{\partial r^{\kappa}} - \frac{\partial \Gamma^{\kappa}{}_{\kappa\mu}}{\partial r^{\nu}} + \Gamma^{\kappa}{}_{\kappa\lambda}\Gamma^{\lambda}{}_{\mu\nu} - \Gamma^{\kappa}{}_{\mu\lambda}\Gamma^{\lambda}{}_{\kappa\nu}, \tag{5.66}$$

and if the Riemannian manifold is endowed with the Jacobi metric, the formula becomes

$$\begin{aligned}\mathrm{Ric}_{\mu\nu} =& \mathrm{Ric}[m]_{\mu\nu} + \frac{3n_{\mathrm{atoms}}-1}{4(E-V)^2}\big(2(E-V)\big(\partial_\mu\partial_\nu V - \Gamma^{\kappa}_{\mu\nu}[m]\partial_\kappa V\big) + 3\partial_\mu V\partial_\nu V\big) \\ &+ \frac{m_{\mu\nu}}{4(E-V)^2}\big(2(E-V)\partial_\gamma\partial^\gamma V - (3n_{\mathrm{atoms}}-3)m^{\alpha\beta}\partial_\alpha V\partial_\beta V\big)\end{aligned} \tag{5.67}$$

where

$$\mathrm{Ric}[m]_{\mu\nu} = \frac{\partial \Gamma^{\kappa}{}_{\mu\nu}[m]}{\partial r^{\kappa}} - \frac{\partial \Gamma^{\kappa}{}_{\kappa\mu}[m]}{\partial r^{\nu}} + \Gamma^{\kappa}{}_{\kappa\lambda}[m]\Gamma^{\lambda}{}_{\mu\nu}[m] - \Gamma^{\kappa}{}_{\mu\lambda}[m]\Gamma^{\lambda}{}_{\kappa\nu}[m] \tag{5.68}$$

for n_{atoms} being the total number of atoms in the molecule of interest and $\Gamma[m]$ being the Christoffel symbol computed on $m_{\mu\nu}$ as defined in Equation (5.21). This quantity gives the directional curvatures at point $r \in \mathcal{M}^{cl}$, [124, 125] while defining a symmetric bilinear form on the tangent space $T_p\mathcal{M}^{cl}$ as, for some $p \in \mathcal{M}^{cl}$,

$$\mathrm{Ric}_p : T_p\mathcal{M} \times T_p\mathcal{M} \to \mathbb{R}, \quad (A, B) \mapsto \mathrm{Ric}_p(A, B) = R_{\mu\nu}A^\mu B^\nu \tag{5.69}$$

for arbitrary A, $B \in T_p\mathcal{M}^{cl}$ [126].

Its eigenvectors $v^{(i)}(r)$ satisfy

$$\mathrm{Ric}_{\mu\nu}v^{(i)\nu} = \lambda_i m_{\mu\nu}v^{(i)\nu}, \tag{5.70}$$

and the corresponding eigenvalues $\lambda_i(r)$ quantify the average curvature in each principal direction. Equivalently, for any unit-length tangent vector v, the scalar called the *Ricci curvature*:

$$\mathrm{Ric}_g(v, v) = v^\mu \mathrm{Ric}_{\mu\nu}v^\nu \tag{5.71}$$

measures how geodesics initially separated along v will tend to converge ($\mathrm{Ric}_g > 0$) or diverge ($\mathrm{Ric}_g < 0$) [127].

At an equilibrium point ($\partial_\mu V = 0$), by noting that

$$\mathrm{Ric}_{\mu\nu} = -\frac{3n_{\mathrm{atoms}} - 1}{2(E - V(r))} \frac{\partial^2 V}{\partial r^\mu \partial r^\nu} + (\text{isotropic shift}), \tag{5.72}$$

one can learn that the Ricci eigendirections coincide with the usual Hessian normal modes and $\lambda_i \propto -\omega_i^2$. Crucially, this eigen-decomposition remains valid at every snapshot along a trajectory, even far from the energy minima–providing a coordinate-invariant generalisation of vibrational modes that continuously tracks the true dynamical stability of the motion.

5.4.2 Finding the eigenvectors of a Ricci tensor

To avoid the $\mathcal{O}({n_{\mathrm{atoms}}}^3)$ cost of full diagonalisation, one can employ a matrix-free power-iteration with deflation directly on the Ricci operator $\mathrm{Ric}_r : T_r\mathcal{M}^{cl} \to T_r\mathcal{M}^{cl}$ for $r \in \mathcal{M}^{cl}$. Starting from a random unit vector $v^{(0)}$, one can recursively find the better eigenvector candidates of Ric_r as:

$$w^{(k+1)} = \mathrm{Ric}_r v^{(k)}, \quad v^{(k+1)} = \frac{w^{(k+1)}}{\|w^{(k+1)}\|}. \tag{5.73}$$

The iteration terminates when the quantity called Rayleigh quotient:

$$\lambda^{(k)} = \frac{v^{(k)\mu}\mathrm{Ric}_{\mu\nu}v^{(k)\nu}}{v^{(k)\kappa}{v^{(k)}}_\kappa} = v^{(k)\mu}\mathrm{Ric}_{\mu\nu}v^{(k)\nu} \tag{5.74}$$

converges within the tolerance ϵ, i.e. $\left|\lambda^{(k)} - \lambda^{(k-1)}\right| < \epsilon$, and the termination is guaranteed. This step allows one to obtain an approximate pair of the largest eigenvalue and its corresponding eigenvector, (λ_1, v_1), of Ric_r [128, 129]. Removing the largest eigenvector contribution from the Ricci operator by

$$\mathrm{Ric}_r \leftarrow \mathrm{Ric}_r - \lambda_1 v_1 \otimes v_1 \tag{5.75}$$

and following the recipe introduced above yields another pair (λ_2, v_2). The k-largest eigenvalues and its eigenvectors can be found in $\mathcal{O}(kn_{\mathrm{atoms}}^2)$ complexity using this method without diagonalising the Ricci tensor [130, 131].

5.5 Understanding quasi-periodicity via holonomy angles, monodromy and tubular trajectories

In nonlinear dynamical systems, periodic orbits constitute the skeleton of phase space: they organise nearby quasi-periodic tori and delineate the onset of chaotic motion [105,106]. In the context of molecular vibrations, each stable periodic orbit represents an anharmonic continuation of a normal mode, while bifurcations of these orbits presage the emergence of different vibrational pathways and resonances [132]. This relation links pure vibrational loops in the configuration space with the net rotation of the molecule [133], allowing a mechanical construction of the molecular coordinate space whose curvature can be computed to predict the orientational changes induced by the vibrational cycles [134].

In an idealised treatment, each peak in a vibrational spectrum, characterized by its frequency and amplitude, corresponds to an eigenvalue of the potential-energy Hessian, with the associated eigenvector defining the normal-mode oscillation direction. Consequently, identifying these peaks reduces to finding a closed trajectory:

$$r : [0, P] \to \mathcal{M}^{cl}, \quad r(0) = r(P) \tag{5.76}$$

parameterised by arc length, s. In practice, however, identifying these periodic orbits is challenging, since a complete molecular dynamics study requires finding periodic solutions to the equations of motion for an n_{atoms}-body system [135].

In this section, a method that can quantify the quasi-periodicity score of a quasi-periodic system is introduced in two different approaches–holonomy angle and tubular trajectory's self-intersection. These two methods rely heavily on a tool from differential geometry called *parallel-transport*, quantifying the degree of change in the tangent vector along the trajectory while preserving parallelism relative to the manifold's local curvature.

5.5.1 Brief introduction to parallel-transport

Let $(\mathcal{M}, g)$ be the configuration manifold of an N-particle system under an empirical potential, and let $\Gamma^{\mu}{}_{\nu\kappa}(r)$ be the Christoffel symbols of the induced Levi–Cività connection defined as the following.

$$\nabla F^{\gamma} = \frac{dF^{\gamma}}{ds} + \Gamma^{\gamma}{}_{\mu\nu}\frac{dr^{\mu}}{ds}F^{\nu}, \quad \text{and } \Delta V = \nabla \tag{5.77}$$

For a periodic trajectory $r(s) = r(0)$, consider an infinitesimal deviation in the direction given by a tangent vector $v(s) \in T_{r(s)}\mathcal{M}$. Parallel transport of this tangent vector along the trajectory $r(s)$ is given as

$$\nabla v^{\mu} = \frac{dv^{\mu}}{ds} + \Gamma^{\mu}{}_{\nu\kappa}(r(s))v^{\nu}(s)\frac{dr^{\kappa}}{ds} = 0, \tag{5.78}$$

in space which in local components yields the linear system

$$\frac{dv^{\mu}}{ds} = -\Gamma^{\mu}{}_{\nu\kappa}(r(s))\frac{dr^{\nu}}{ds}v^{\kappa}(s), \tag{5.79}$$

where dr^{ν}/ds is the unit-speed tangent. One can then define an operator

$$A(s)^{\mu}{}_{\nu} = -\Gamma^{\mu}_{\nu\kappa}(r(s))\,\frac{dr^{\kappa}}{ds}, \tag{5.80}$$

so that (5.79) reads $\dot{v} = A(s)v$. By the symmetry of the metric tensor and the symmetry in two lower-indices of Christoffel symbol (see Equation (5.19)), the operator $A(s)$ is skew-symmetric: $A^{T}g + gA = 0$, ensuring preservation of the norm $\|v(s)\|_g = \sqrt{g_{\mu\nu}v^{\mu}(s)v^{\nu}(s)}$. Thus the flow generated by $A(s)$ is an isometry on each tangent space.

5.5.2 Holonomy angle

Given the vibrational mode vector $e(s)$, again s being the arclength parameter, the parallel-transported mode vector, e, in the Riemannian manifold $\left(\mathcal{M}^{cl}, g\right)$ is given as:

$$\frac{de^{\mu}}{ds} + \Gamma^{\mu}{}_{\nu\kappa}\frac{dr^{\nu}}{ds}e^{\kappa} = 0 \tag{5.81}$$

where r^μ is the coordinates in $\mathcal{M}^{cl}$. Physically, this condition evaluates how e^μ evolves as straight as possible while preserving its length and its inner product with any other vector under the infinitesimal displacements dictated by the manifold's curvature.

When $r(s)$ traces out a closed geodesic loop, C, returning to its initial point r_0 at some arclength parameter s_{return}, the parallel transport defines a linear map

$$\text{Par}_C : T_{r_0}\mathcal{M}^{cl} \to T_{r_0}\mathcal{M}^{cl} \tag{5.82}$$

which satisfies $\text{Par}_C \in \text{SO}(3n_{\text{atoms}} + 1)$, with $\text{SO}(n)$ being a special orthogonal group for an n-dimensional manifold. Restricting the map Par_C to the one-dimensional subspace spanned by the vibrational mode vector $e(s_0)$ yields a pure rotation in this space by an angle

$$\Theta = \arccos(e^\mu(s_0) g_{\mu\nu} e^\nu(s_{\text{return}})). \tag{5.83}$$

This angle is called *holonomy angle*, and is a gauge-invariant measure of the geometric phase. It depends only on the path C not on the intermediate properties of the trajectory, or the nature of the parameterisation.

When the geometric phase is truly zero, the parallel-transport map around the loop reduces to the identity on our frame vector. Equivalently, one finds

$$e(s_0)^\mu g_{\mu\nu} e(s_{\text{return}})^\nu = 1 \implies \Theta = 0. \tag{5.84}$$

Physically, this implies that if the molecule under study traces out its closed path in the configuration manifold, not only has its coordinate vector returned exactly to r_0 , but the vibrational direction e also coincides perfectly with its orientation.

5.5.3 Holonomy angle & carbon dioxide

One can numerically obtain the values of holonomy angles at every step by Algorithm 7 if a trajectory containing a loop is known. As Equation (5.81) is an

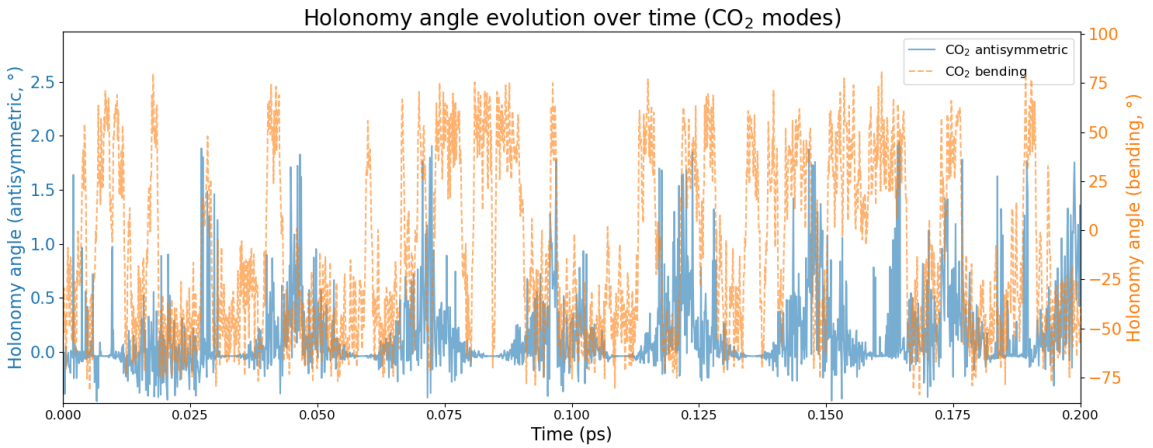


Figure 5.7. *Evolution of holonomy angle over time for carbon dioxide in vacuum at* $300\,\mathrm{K}$. *Blue* (*resp*. *orange*) *curve shows the evolution of the holonomy angle of antisymmetric* (*resp*. *bending*) *mode*.

initial-value problem, a Richardson-style convergence test is used to gauge the error [136].

Figure 5.7 shows the temporal evolution of the holonomy angle for the antisymmetric and bending modes obtained via Algorithm 7. For stability, the embedded Runge–Kutta method RK45 (Dormand–Prince) is used, which yields a fifth-order solution with an embedded fourth-order error estimate for adaptive step-size control [137].

Despite pronounced differences in amplitude, waveform shape, and apparent chaoticity, both curves exhibit oscillatory behaviour and return to near-zero holonomy in a pseudo-periodic fashion. As the geometric-phase accumulation rate is governed by the vibrational motion—faster modes traverse curved regions of the configuration manifold more rapidly—the oscillation frequencies of the holonomy angle closely mirror the underlying normal-mode frequencies. This can be verified once the power spectrum of the angles are computed as in Figure 5.8.

The antisymmetric-stretch holonomy spectrum exhibits a single, well-re-

Algorithm 7 Measurement of holonomy angle by discrete parallel transport

Require: Discrete trajectory $\{(r_i, \dot{r}_i)\}_{i=0}^{N}$ sampled along arc length s with cumulative $\{s_i\}$ and $s_0 = 0$, $s_N = L$. Metric $g(r)$, Christoffel symbols $\Gamma(r)$, initial value of r denoted as r_0, initial vector e_0 with $\|e_0\|_{g(r_0)} = 1$, initial step h, tolerance η, safety $\sigma \in (0,1)$, return threshold ϵ.
($r(s), \dot{r}(s)$ *are treated as piecewise-linear / piecewise-constant from the samples*; *within each* $[s_i, s_{i+1}]$ r *and* $\dot{r}$ *are linearly interpolated*. $\dot{r}(s)$ *indicates the derivative with respect to* s *in this algorithm*.)

$s \leftarrow 0$, $r \leftarrow r_0$, $e \leftarrow e_0$
while $s < L$ **do**
  Limit step to segment:
    Find i with $s \in [s_i, s_{i+1})$ and set $h \leftarrow \min\{h,\ s_{i+1} - s\}$
  Define $A(r, \dot{r})^{\mu}{}_{\nu} \leftarrow -\Gamma^{\mu}{}_{\nu\rho}(r)\, \dot{r}^{\rho}$
  Evaluate RK45 stages:
    For each stage k, compute $(r_k, \dot{r}_k)$ by interpolating at $s + c_k h$
    Update A_k: $A_k \leftarrow A(r_k, \dot{r}_k)$
    Accumulate $e^{(5)} \leftarrow e + \sum b_k\, h\, A_k e_k$, $\quad e^{(4)} \leftarrow e + \sum \hat{b}_k\, h\, A_k e_k$
  Error estimate: $\delta \leftarrow \|e^{(5)} - e^{(4)}\|_{g(r)}$ ▷ use metric at current r
  if $\delta \leqslant \eta$ **then**
    Accept step: $s \leftarrow s + h$; update $r \leftarrow r(s)$ by interpolation; $e \leftarrow e^{(5)}$
    Renormalise: $e \leftarrow e / \sqrt{g(r)(e,e)}$ ▷ use g at the *new* point r
    Adapt $h \leftarrow \min\big(h_{\max},\ \sigma\, h\, (\eta / \max(\delta, \varepsilon_{\mathsf{mach}}))^{1/5}\big)$
  else
    Reject step: $h \leftarrow \max\big(h_{\min},\ \sigma\, h\, (\eta / \max(\delta, \varepsilon_{\mathsf{mach}}))^{1/5}\big)$
  end if
  Return check: if $\|r - r_0\|_{g(r)} < \epsilon$ **and** $s > 0$, **break**
end while
Angle: $c \leftarrow \mathrm{clip}\big(g(r_0)(e_0, e), -1, 1\big)$
$\Theta \leftarrow \arccos(c)$
return $\Theta \times \frac{180}{\pi}$

solved peak that matches the GAFF-predicted vibrational frequency, confirming that the holonomy-angle oscillations faithfully track that mode. By contrast, the symmetric-stretch and bending spectra show extra, spurious peaks in addition to the true vibrational frequencies—an effect of mode coupling and discretisation errors. Thus, the holonomy angles alone cannot unambiguously

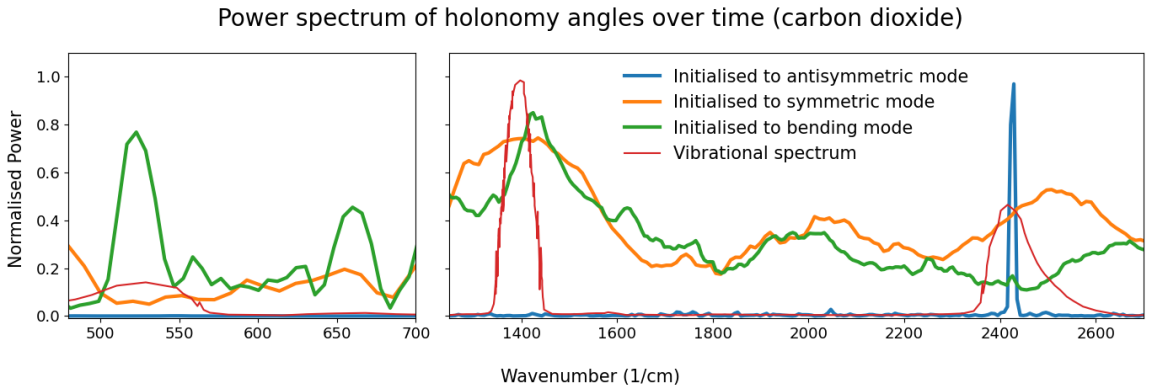


Figure 5.8. *Power spectrum of holonomy angles for respective vibrational modes of carbon dioxide.*

assign a given spectral feature to a particular normal mode; instead, one must isolate the periodic orbit at that frequency. This is where a mathematical tool called *monodromy* becomes useful.

5.5.4 Monodromy matrix and tubular trajectories

Linearised dynamics around a periodic trajectory of period P provide the foundation for understanding both local stability and global volume-evolution properties. The parallel transport equation (5.79) can be written as an ordinary differential equation with matrix value.

$$\frac{d}{ds}\Phi(s) = A(s)\Phi(s), \quad \Phi(0) = I_n, \tag{5.85}$$

where $\Phi(s) \in \mathrm{GL}(3n_{\mathrm{atoms}} + 1)$ is the *fundamental solution matrix* that simultaneously evolves all basis vectors of the tangent space. Component-wise, this satisfies

$$\Phi^{\mu}{}_{\nu}(0) = \delta^{\mu}{}_{\nu}, \quad \dot{\Phi}^{\mu}{}_{\nu} = -\Gamma^{\mu}{}_{\xi\zeta}(q(s))\frac{dr^{\zeta}}{ds}\Phi^{\xi}{}_{\nu}. \tag{5.86}$$

The fundamental solution of this equation can be computed numerically via the exponential update:

$$\Phi(s + \Delta s) = \exp(A(s)\Delta s)\Phi(s), \tag{5.87}$$

where careful attention to the structure of the underlying Lie group ensures the preservation of orthogonality to machine precision [138, 139].

After one complete traversal of the periodic orbit with period P, the *monodromy matrix*:

$$M = \Phi(P) \in \mathrm{O}(3n_{\mathrm{atoms}} + 1) \tag{5.88}$$

encodes how infinitesimal perturbations evolve around the closed loop. Its eigenvalues $\{\lambda_i\}$, known as *Floquet multipliers*, determine the stability and geometric properties of the periodic motion. Floquet multipliers whose moduli lie below unity correspond to exponential stability along their associated eigenvectors, whereas those exceeding unity correspond to exponential instability in the same directions.

Building on the characterization of the monodromy matrix of infinitesimal perturbations after one period, the analysis transitions from a pointwise study of quasi-periodicity and stability to a global, volume-based framework. An infinitesimal neighbourhood volume is transported along the trajectory via the fundamental solution matrix, thereby generating a tubular trajectory in the configuration space. The deformation of this tube over one period encapsulates the quasi-periodic behaviour of the system and reveals the underlying geometric structure.

5.5.4.1 Tubular trajectory volume evolution via parallel transport

Rather than analysing pointwise dynamics, the evolution of finite-volume neighbourhoods around periodic molecular trajectories is characterised by constructing and analysing tubular trajectories in the configuration manifold $\mathcal{M}^{cl}$. The motivation is to carry the entire neighbourhood around the initial point.

Let $\gamma : [0, P] \to \mathcal{M}^{cl}$ be a geodesic. At each point $\gamma(s)$, one can construct the orthogonal complement to the velocity vector:

$$T^{\perp}_{\gamma(s)}\mathcal{M}^{cl} := \big\{ v \in T_{\gamma(s)}\mathcal{M}^{cl} : g_{\mu\nu} v^{\mu} \dot{\gamma}(s)^{\nu} = 0 \big\}, \tag{5.89}$$

where g denotes the Jacobi metric on $\mathcal{M}^{cl}$.

Definition 5.3 (Tubular trajectory)**.** Let $\gamma : [0, P] \to \mathcal{M}$ be a smooth curve in an n-dimensional Riemannian manifold $(\mathcal{M}, g)$, define $T^{\perp}_{\gamma(0)}\mathcal{M}$ as in Equation (5.89), and define denote the $(n-1)$–dimensional orthogonal complement (normal space) to $\dot{\gamma}(s)$.

Choose a nonzero initial normal vector $v_{\perp}(0) \in T^{\perp}_{\gamma(0)}\mathcal{M}$ and parallel transport it along γ to obtain $v_{\perp}(s) \in T^{\perp}_{\gamma(s)}\mathcal{M}$ satisfying $\nabla_{\dot{\gamma}} v_{\perp} = 0$.

Construct a parallel orthonormal frame $\{u_1(s), \ldots, u_{n-1}(s)\}$ of $T^{\perp}_{\gamma(s)}\mathcal{M}$ such that

- *Orthonormality*: $\langle u_i(s), u_j(s) \rangle_g = \delta_{ij}$ for all i, j,
- *Parallel transport*: $\nabla_{\dot{\gamma}} u_i = 0$ for each i,
- *Initial alignment*: $u_1(0) = \dfrac{v_{\perp}(0)}{\|v_{\perp}(0)\|_g}$.

Let $\rho := \|v_{\perp}(0)\|_g$ (equivalently, $\rho = \|v_{\perp}(s)\|_g$). The *tubular trajectory* of radius ρ along γ is

$$\mathcal{T}(\gamma) \;=\; \Big\{ \exp_{\gamma(s)}\Big(\sum_{i=1}^{n-1} \xi_i \, u_i(s)\Big) \;:\; s \in [0, P],\; \sum_{i=1}^{n-1} \xi_i^2 \leqslant \rho^2 \Big\}. \tag{5.90}$$

Definition 5.3 is a generalisation of the definition of cylinder in a curved space with arbitrary dimensions. If the manifold of interest, $\mathcal{M}^{\text{flat}}$, is a 3-dimensional Euclidean manifold and its perpendicular tangent space $T^{\perp}_{\gamma(0)}\mathcal{M}^{\text{flat}}$ is 2-dimensional and connected, a tubular trajectory around a straight line γ in $\mathcal{M}^{\text{flat}}$ can be constructed as below by Definition 5.3:

$$\mathcal{T}(\gamma) = \left\{ \gamma(s) + r\cos(\theta)\frac{v_{\perp}(s)}{|v_{\perp}(s)|_g} + r\sin(\theta)\frac{v^{\perp}_{\perp}(s)}{|v^{\perp}_{\perp}(s)|_g} : s \in [0, \infty), 0 \leqslant r \leqslant \rho, \theta \in [0, 2\pi) \right\} \tag{5.91}$$

for some finite constant $\rho > 0$. Here, $\{v_{\perp}, v^{\perp}_{\perp}\}$ is an orthonormal set of vectors that span $\mathcal{T}^{\perp}_{\gamma(0)}\mathcal{M}^{\text{flat}}$. This is an equation of a cylinder around a straight line γ on a flat Euclidean space with a base radius ρ.

Figure 5.9 shows the tubular trajectories around a geodesic, parameterised by the centre-of-mass coordinates of carbon dioxide, which are chosen as a

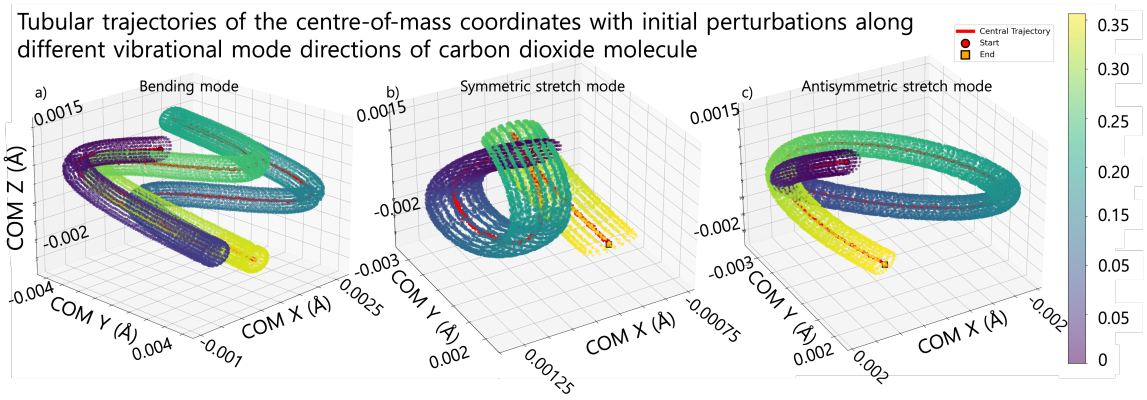


Figure 5.9. *Centre of mass tubular trajectories of a carbon dioxide system with initial perturbations along vibrational modes*: *a*) *bending*, *b*) *symmetric stretch*, *and c*) *anti-symmetric stretch modes*.

visualisation aid to facilitate understanding of the tubular trajectories. Since the system is three-dimensional in terms of centre-of-mass motion, Equation (5.91) can be applied to construct the corresponding tubular trajectories. Although the trajectories are initially perturbed and exhibit markedly different shapes, they self-intersect, illustrating a seemingly periodic nature of these tubular trajectories.

The quantity $\mathcal{T}(\gamma)$, from Definition 5.3, creates a tube whose radius $\rho(s) = |v_\perp(s)|_g$ evolves according to the geodesic deviation equation as the initial perpendicular vector is parallel transported along γ. Note also that the initial magnitude of $v_\perp$ can be arbitrarily chosen at $s = 0$, as long as this initial magnitude is finite.

The perpendicular vector $v_\perp(s)$ satisfies the parallel transport equation:

$$\frac{dv_\perp^\xi}{ds} + \Gamma^\xi{}_{\mu\nu}(\gamma(s))\frac{d\gamma^\mu}{ds}v_\perp^\nu = 0, \tag{5.92}$$

with initial condition $v_\perp(0) = v_0 - \langle v_0, \dot\gamma(0)\rangle_g \frac{\dot\gamma(0)}{\|\dot\gamma(0)\|_g^2}$. In matrix form, this can be written as:

$$\frac{dv_\perp}{ds} = -A(s)v_\perp(s), \quad v_\perp(0) = \text{initial perpendicular component}, \tag{5.93}$$

where the connection matrix $A(s)$ has components:

$$A(s)^{\xi}{}_{\zeta} = \Gamma^{\xi}{}_{\mu\zeta}(\gamma(s))\frac{d\gamma^{\mu}}{ds}, \tag{5.94}$$

with $\Gamma^{\xi}{}_{\mu\nu}$ being the Christoffel symbols of the Jacobi metric.

The evolved radius is given by:

$$\rho(s) = \|v_{\perp}(s)\|_g = \sqrt{g_{\xi\zeta}(\gamma(s))v_{\perp}^{\xi}(s)v_{\perp}^{\zeta}(s)}, \tag{5.95}$$

which changes according to the curvature of the manifold along the geodesic.

The cross-sectional volume of the tube at parameter s is determined by the evolution of vectors in the $3n_{\text{atoms}}$ dimensional space perpendicular to $\dot{\gamma}(s)$.

To construct the complete perpendicular cross-section, complete $v_{\perp}(s)$ to an orthonormal basis $\{u_1(s), u_2(s), \ldots, u_{3n_{\text{atoms}}}(s)\}$ of $T^{\perp}_{\gamma(s)}\mathcal{M}^{cl}$, where $u_1(s) = v_{\perp}(s)/|v_{\perp}(s)|_g$. Each basis vector satisfies the parallel transport equation:

$$\nabla_{\dot{\gamma}} u_i = 0, \quad i = 1, 2, \ldots, 3n_{\text{atoms}}. \tag{5.96}$$

Thus, it becomes natural that the tube cross-section at parameter s is the $3n_{\text{atoms}}$ dimensional ball:

$$\mathcal{C}(s) = \left\{ \gamma(s) + \sum_{i=1}^{3n_{\text{atoms}}} \xi^i u_i(s) : \sum_{i=1}^{3n_{\text{atoms}}} (\xi^i)^2 \leqslant \rho(s)^2 \right\} := \gamma(s) + B_{3n_{\text{atoms}}}(\rho(s)), \tag{5.97}$$

where $\rho(s) = |v_{\perp}(s)|_g$ is the evolved radius and $B_n(W)$ stands for an n-dimensional ball with radius $W > 0$. This definition is natural, as the collection of $\mathcal{C}$ for all s in the interval of interest becomes the entire tubular trajectory as defined in Definition 5.3.

Let $U(s) = [u_1(s)|u_2(s)|\cdots|u_{3n_{\text{atoms}}}(s)]$ be the matrix whose columns are the orthonormal basis vectors of $T^{\perp}_{\gamma(s)}\mathcal{M}^{cl}$. By Equation 5.85, $U(s)$ satisfies

$$U(s) = \Phi^{\perp}(s)U(0) \tag{5.98}$$

where $\Phi^{\perp}(s)$ is the projection of Φ onto $T^{\perp}_{\gamma(s)}\mathcal{M}^{cl}$. With this in mind, the cross-section $C(s)$ can then be rewritten as

$$C(s) = \left\{\gamma(s) + U(s)\xi : \xi \in \mathbb{R}^{3n_{\text{atoms}}}, \|\xi\|^2 \leqslant \rho(s)^2\right\} \tag{5.99}$$

$$= \left\{\gamma(s) + \Phi^{\perp}(s)U(0)\xi : \xi \in \mathbb{R}^{3n_{\text{atoms}}}, \|\xi\|^2 \leqslant \rho(s)^2\right\} \tag{5.100}$$

$$= \gamma(s) + \Phi^{\perp}(s)B_{3n_{\text{atoms}}}(\rho(0)) \tag{5.101}$$

Note that if the Floquet multipliers of $\Phi^{\perp}(s)$ are given as $\lambda_1(s), \cdots, \lambda_{3n_{\text{atoms}}}(s)$, from the stability criteria discussed earlier, one can immediately understand that the cross-section

- stretches along the directions where $\|\lambda_i(s)\| > 1$, and
- contracts along the directions where $\|\lambda_i(s)\| < 1$.

The rotation of the tube is given by the complex nature of the λ_i's, and if the magnitude of the Floquet multiplier is equivalent to 1, the cross-sectional volume will not change in the corresponding direction.

5.5.4.2 Volume overlap between return events

For a trajectory whose return events, as the ones shown in Figure 5.9, have been identified at some arclength domain by holonomy angle analysis $\{s_k\}_{k=0}^{N_{\text{returns}}}$, the quasi-periodicity can be quantified by the volume overlap between any two such returns. Intuitively, if the tube cross-sections at s_i and s_j share a large common volume, then the segment of trajectory between them closely approximates an ideal periodic orbit–hence the motion may be regarded as quasi-periodic to the extent measured by that overlap.

Figure 5.10 illustrates the tubular trajectories that terminate at an intersection volume after a complete period, initiated by perturbations along the direction corresponding to the frequency shared by the vibrational spectrum and the holonomy angle power spectrum. Despite the dissipation introduced

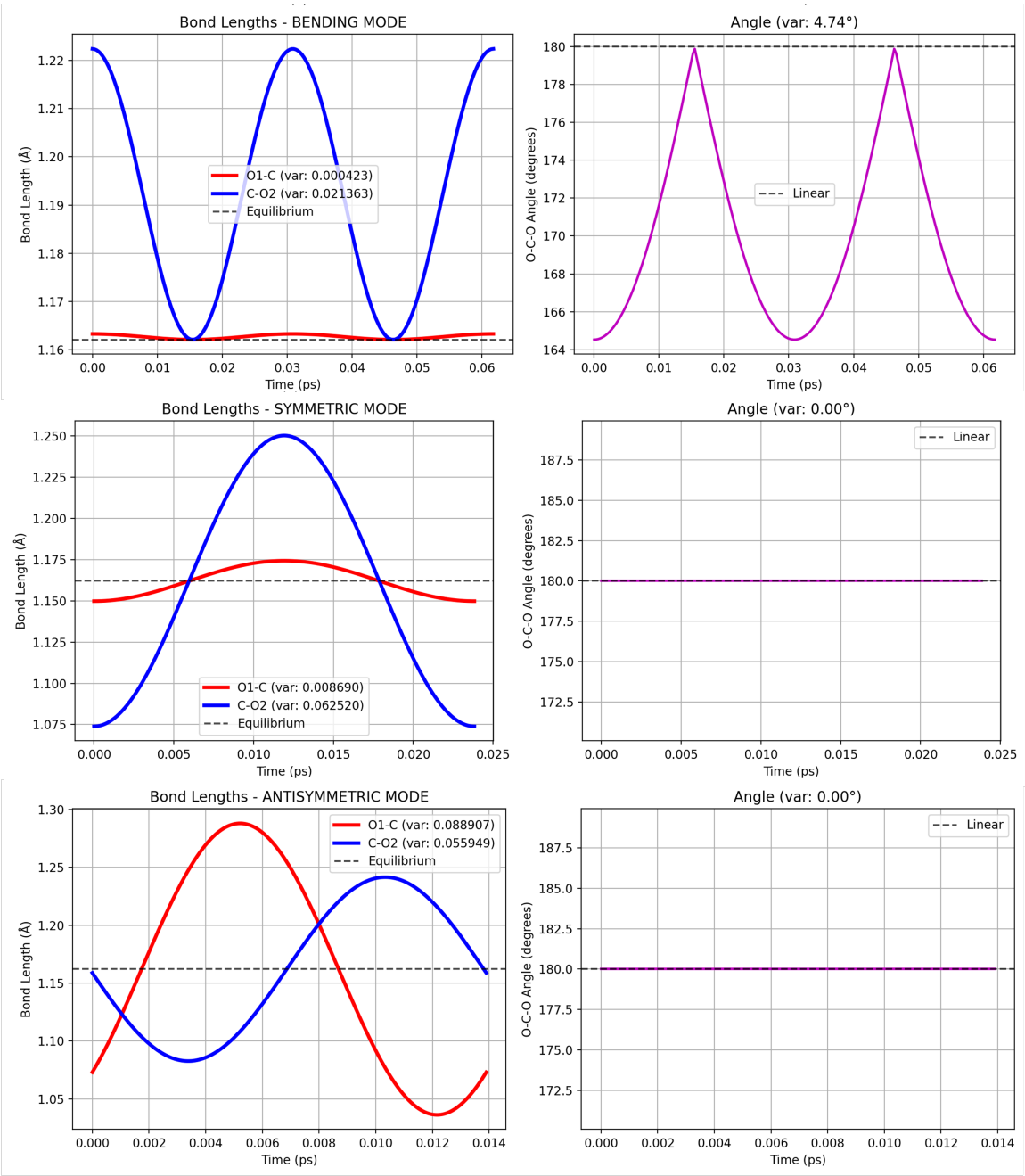


Figure 5.10. *A single period trajectory obtained by identifying a trajectory that ends at the tube-intersection. Different periodic motions are obtained via different initial perturbations in the direction of the Ricci normal modes corresponding to the shared peaks between the holonomy angle power spectrum and the vibrational spectra.*

by the thermostat (300 K), the resulting periodic orbits exhibit the expected physical behaviour and accurately reproduce the total period for each orbit.

The overlap between tube cross-sections at s_i and s_j is defined through the intersection measure:

$$\mathcal{O}(s_i, s_j) = \frac{\mathrm{Vol}(\mathcal{T}(\gamma(s_i)) \cap \mathcal{T}(\gamma(s_j)))}{\mathrm{Vol}(\mathcal{T}(\gamma(s_i)) \cup \mathcal{T}(\gamma(s_j)))}, \tag{5.102}$$

where $\mathrm{Vol}(\cdot)$ denotes the volume measure on $\mathcal{M}^{cl}$. The intersection volume can be computed as

$$\mathrm{Vol}(\mathcal{T}(\gamma(s_i)) \cap \mathcal{T}(\gamma(s_j))) = \int_{\|\xi\| \leqslant \rho(s_i)} \mathbf{1}\left\{\left\|\Phi^{\perp}(s_j)\xi\right\| \leqslant \rho(s_j)\right\} d\xi = I_{ij} \tag{5.103}$$

with ξ being the vectorial notation of ξ_i of Equation (5.90). The notation $\mathbf{1}\{Q\}$ was used to denote the indicator function for a Boolean condition Q. Formally, the indicator function for a condition Q is defined as follows:

$$\mathbf{1}\{Q\}(x) = \begin{cases} 1, & \text{if } Q \text{ is true} \\ 0, & \text{otherwise} \end{cases}. \tag{5.104}$$

The volume of the union is

$$\mathrm{Vol}(\mathcal{T}(\gamma(s_i)) \cup \mathcal{T}(\gamma(s_j))) = C_{3n_{\mathrm{atoms}}}\rho(s_i)^{3n_{\mathrm{atoms}}} + C_{3n_{\mathrm{atoms}}}\rho(s_j)^{3n_{\mathrm{atoms}}} - I_{ij} \tag{5.105}$$

with C_n being the volume of an n-dimensional unit sphere. The expression for $\mathcal{O}(s_i, s_j)$ can be rewritten as:

$$\mathcal{O}(s_i, s_j) = \frac{f_{ij}}{\left(\frac{\rho(s_j)}{\rho(s_i)}\right)^{3n_{\mathrm{atoms}}} + 1 - f_{ij}}, \quad f_{ij} = \frac{I_{ij}}{C_{3n_{\mathrm{atoms}}}\rho(s_i)^{3n_{\mathrm{atoms}}}}. \tag{5.106}$$

If one identifies a random point $\xi \in C(s_i)$, one can find a monodromy matrix M so that $M\xi \in C(s_j)$. Noting that one can identify ξ as a product of two independent uniform sampling processes as $\xi = z\hat{\mathbf{u}}$: one determines the unit vector, $\hat{\mathbf{u}}$, on a unit sphere of $3n_{\mathrm{atoms}}$ dimensional dimensions, and the other determining the magnitude of the vector, $z \in [0, \rho(s_i)]$, by the probability density below.

$$p_z = (3n_{\mathrm{atoms}} + 1)\frac{z^{3n_{\mathrm{atroms}}}}{\rho(s_i)^{3n_{\mathrm{atoms}}+1}}. \tag{5.107}$$

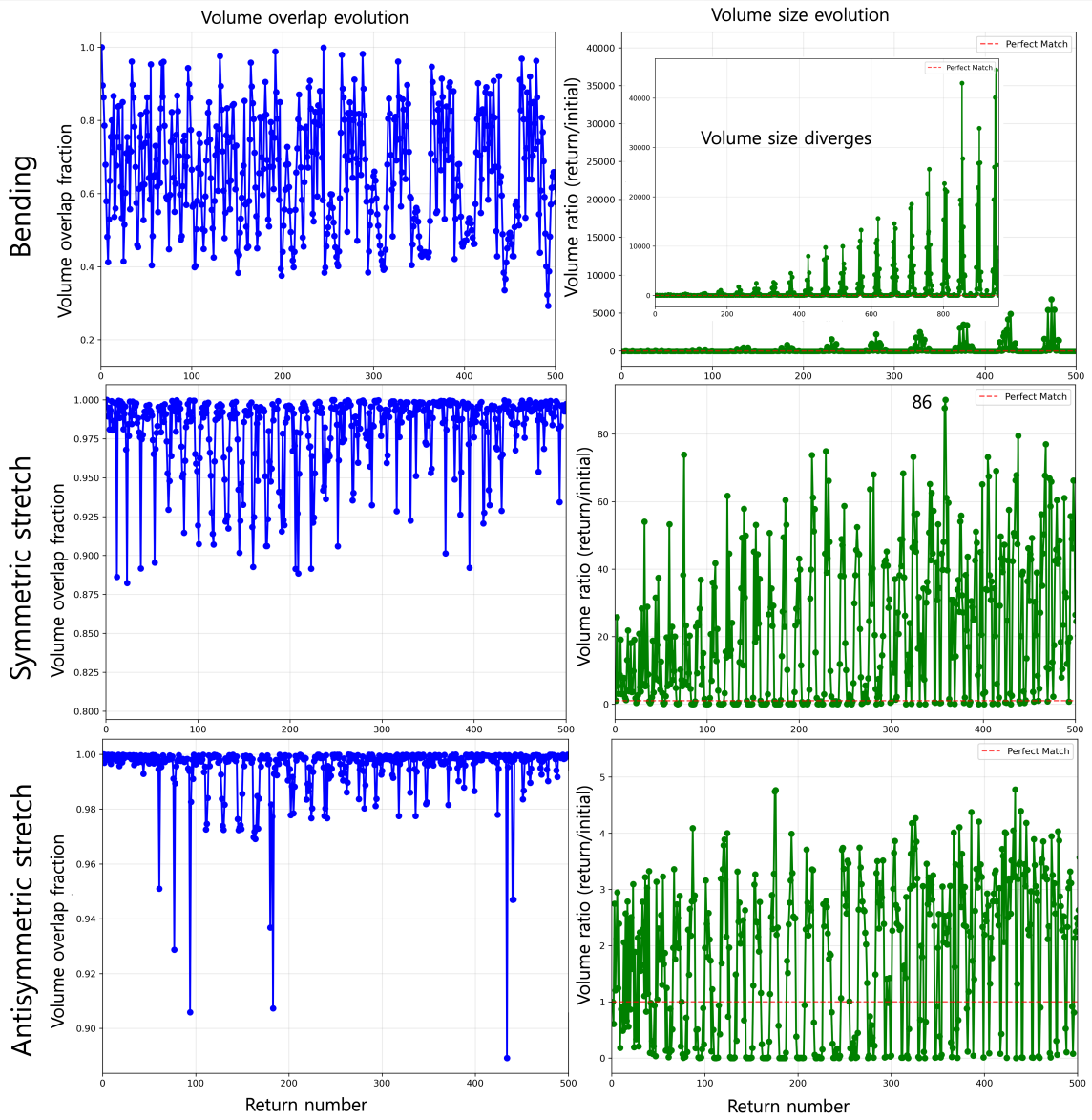


Figure 5.11. *Change of tubular section volume and tube intersection volumes over number of returns to the vicinity of original tubular section* (*for initial perturbations in bending mode* (*TOP*), *symmetric mode* (*MIDDLE*), *and antisymmetric mode* (*BOTTOM*) *directions*).

If the mapping M carried the vector ξ into the region shared by $C(s_i)$ and $C(s_j)$, the following must be satisfied:

$$z\|M\hat{\mathbf{u}}\| \leqslant \rho(s_j) \implies z \leqslant \frac{\rho(s_j)}{\|M\hat{\mathbf{u}}\|}. \tag{5.108}$$

Hence, the fraction of the radial segment that maps ξ into the shared volume is

$$\Pr\left(z \leqslant \min\left(\rho(s_i), \frac{\rho(s_j)}{\|M\hat{\mathbf{u}}\|}\right)\right) = \min\left(1, \frac{\rho(s_j)}{\rho(s_i)\|M\hat{\mathbf{u}}\|}\right)^{3n_{\text{atoms}}}. \tag{5.109}$$

By averaging this over all directions, one can recover f_{ij} as

$$f_{ij} = \mathbb{E}_{\hat{\mathbf{u}}\sim\text{Unif}(S^{3n_{\text{atoms}}})}\left[\Pr\left(z \leqslant \min\left(\rho(s_i), \frac{\rho(s_j)}{\|M\hat{\mathbf{u}}\|}\right)\right)\right] \tag{5.110}$$

$$= \int_{S^{3n_{\text{atoms}}}} \min\left(1, \frac{\rho(s_j)}{\rho(s_i)\|M\hat{\mathbf{u}}\|}\right)^{3n_{\text{atoms}}} \sigma(d\hat{\mathbf{u}}) \tag{5.111}$$

with $\text{Unif}(G)$ signifying uniform sampling from the set G, and σ being the surface measure of the sphere defined as

$$\sigma(d\hat{\mathbf{u}}) = \frac{\Gamma(3n_{\text{atoms}}/2)}{2\pi^{3n_{\text{atoms}}/2}} \prod_{k=1}^{3n_{\text{atoms}}-1} \sin^{3n_{\text{atoms}}-1-k}(\varphi_k) d\varphi_1 \cdots d\varphi_{3n_{\text{atoms}}-1} \tag{5.112}$$

for angular variables on the unit sphere φ_is. $\Gamma(\cdot)$ in this equation was used to represent a Gamma function.

Figure 5.11 shows the progressive divergence of volume sizes at each return for the system initially perturbed in the bending mode, accounting for the multiple, broad peaks observed in the holonomy-angle power spectrum. This divergence indicates that, under thermostat conditions, natural trajectories perturbed along these modal directions struggle to maintain pseudo-periodic oscillations around the ideal vibrational frequency. Although the cross-sectional volume grows without bound, the volume overlap oscillates widely between approximately 100 % and 30 % coverage. Consistent volume coverage correlates with a higher probability of identifying periodic orbits, whereas a diverging cross-sectional volume signifies instability in the corresponding mode.

In the same figure, perturbation along the antisymmetric stretch mode exhibits markedly different behaviour: the volume ratio oscillates around unity—remaining below 5–and the volume overlap consistently exceeds 88 %, typically above 97 %. These results demonstrate the stability of the tubular trajectory under this perturbation and indicate the relative ease of locating a pseudo-periodic orbit in this mode. This stability accounts for the pronounced peak observed in the corresponding power spectrum.

The symmetric stretch mode yields intermediate behaviour. In this case, the volume ratio reaches values as high as 86 without exhibiting a clear divergence trend. The volume overlap fraction remains consistently above 87.5 %, but displays more pronounced oscillations than in the antisymmetric stretch mode, while remaining less violent than those observed for the bending mode.

5.6 Fat manifolds and coupled vibrational modes

A tubular trajectory around a geodesic introduced in the previous section provides a higher-dimensional understanding of the stability of the trajectory perturbed by a certain mode direction. However, the tubular trajectory can be further generalised to understand the vibrational mode couplings. To do so, a machinery called *fat manifold* is used. The contents introduced in Chapter 5.6.1 and 5.6.2 suffice for the physics to be discussed in this manuscript. For a thorough and rigorous account of the subject, consult the work of Alexandre Vinogradov and Alessandro de Paris [140].

5.6.1 Definition of fat manifolds

Given the geodesic, metric, and for any s, an exponential map of $v \in V_s \subset T^{\perp}_{\gamma(s)}\mathcal{M}$ is defined as a solution of the geodesic equation

$$\frac{d^2\sigma^\mu}{d\tau^2} + \Gamma^\mu{}_{\nu\kappa}\frac{d\sigma^\nu}{d\tau}\frac{d\sigma^\kappa}{d\tau} = 0 \tag{5.113}$$

such that $\sigma(0) = \gamma(s)$ and $\dot{\sigma}(0) = v/\|v\|$. Note that V_s is an open and connected subset of $T^{\perp}_{\gamma(s)}\mathcal{M}$. Notationally, the exponential map is given as

$$\exp_{\gamma(s)}(v) = \sigma(\tau = \|v\|). \tag{5.114}$$

This definition then allows one to define a tubular neighbourhood.

Definition 5.4 (Tubular Neighborhood)**.** For a given family of neighbourhoods $\{V_s\}_s$, the *tubular neighbourhood* of radius $\epsilon(s)$ around the geodesic is:

$$U = \bigcup_{s\in[0,P]} \big\{\exp_{\gamma(s)}(v) : v \in V_s, \|v\| < \epsilon(s)\big\}. \tag{5.115}$$

This then allows one to give a natural definition of the projection map $\pi : U \to [0, P]$, where for any point $q \in U$, there exists an $s \in [0, P]$ and $v \in T^{\perp}_{\gamma(s)}\mathcal{M}$ such that $q = \exp_{\gamma(s)}(v)$. The projection is then defined by $\pi(q) = s$. Note that if one defines $\epsilon(0) < \rho(0)$, it is trivial that the tubular neighbourhood is a connected open subset of $\mathcal{T}(\gamma)$ with an initial tube radius $\rho(0)$.

For each parameter value $s \in [0, P]$ along the geodesic $\gamma(s)$, the perpendicular space $T^{\perp}_{\gamma(s)}\mathcal{M}$ consists of all vectors orthogonal to $\dot{\gamma}(s)$. Denote by $V_s \subset T^{\perp}_{\gamma(s)}\mathcal{M}$ an open, connected subset on which the exponential map is well-defined. The union of these perpendicular spaces is then

$$U^F = \bigcup_{s\in[0,P]} \{s\} \times V_s = \{(s, v) \colon s \in [0, P],\ v \in V_s\}. \tag{5.116}$$

Each element $(s, v) \in U^F$ consists of:

- a *parameter value* s indicating position along the geodesic, and
- a *perpendicular vector* v representing a possible vibrational deviation at that position.

The exponential map construction naturally defines a projection map:

$$\pi^F : U^F \to [0, P], \quad \pi^F(s, v) = s. \tag{5.117}$$

This projection simply extracts the parameter coordinate, "forgetting" the perpendicular vector information. For each fixed s, the level set $\pi^{F^{-1}}(s) = \{s\} \times V_s$ consists of all possible perpendicular vectors at parameter value s, and this level set has the natural structure of a vector space (since $V_s \subset T^{\perp}_{\gamma(s)}\mathcal{M}$ is a vector space).

The crucial connection between this parameter space U^F and actual trajectories in configuration space $\mathcal{M}$ is provided by the extended exponential map: $\Pi : U^F \to \mathcal{M}, \quad \Pi(s, v) = \exp_{\gamma(s)}(v)$ This map takes a ***fat point*** (s, v)–representing parameter value s with perpendicular deviation v - and produces the actual corresponding point in configuration space.

The image of Π gives us the tubular neighbourhood:

$$U = \Pi(U^F) = \bigcup_{s \in [0,P]} \left\{ \exp_{\gamma(s)}(v) : v \in V_s \right\} \tag{5.118}$$

This is the collection of all points in $\mathcal{M}$ that can be reached by taking perpendicular steps from the geodesic $\gamma(s)$.

The fat manifold can then be defined as shown below.

> **Definition 5.5** (Fat manifold)**.** The fat manifold is the parameter space U^F equipped with:
>
> 1. *The collection structure*: $U^F = (s, v) : s \in [0, P], v \in V_s \subset T^{\perp}_{\gamma(s)}\mathcal{M}$
> 2. *The projection map*: $\pi^F : U^F \to [0, P]$ given by $\pi^F(s, v) = s$
> 3. *The realisation map*: $\Pi : U^F \to \mathcal{M}$ given by $\Pi(s, v) = \exp_{\gamma(s)}(v)$

Each *level set* $\pi^{F^{-1}}(s)$ is called a *fat point* and represents all possible perpendicular deviations available at each parameter value s. The fat manifold thus encodes the complete "fattened" version of the geodesic trajectory, where each point along the trajectory is "thickened" to include all its perpendicular perturbations. Theorem 5.6 demonstrates this intuition.

Theorem 5.6. *$\mathcal{T}(\gamma)$, of Definition 5.3, is a fat manifold.*

Proof. Note that from Definition 5.3 that for a smooth curve $\sigma : [0, S] \to \mathcal{M}$ in an n-dimensional Riemannian manifold $(\mathcal{M}, g)$, the tubular trajectory is given by:

$$\mathcal{T}(\sigma) = \left\{ \exp_{\sigma(s)} \left(\sum_{i=1}^{n-1} v^i u_i(s) \right) : s \in [0, S], \sum_i \left(v^i\right)^2 \leqslant \rho(s)^2 \right\} \tag{5.119}$$

with $u_i(s)$ being a parallel-transported orthonormal frame of the normal bundle $\mathcal{T}^{\perp}_{\sigma(s)}\mathcal{M}$. Now, noticing from Definition 5.5, a fat manifold is the parameter space

$$U = \{(s, v) : s \in [0, S], v \in V_s\}, \tag{5.120}$$

equipped with three pieces of data:

- **Collection structure:** $U = \bigcup_{s\in[0,S]} \{s\} \times V_s$.
- **Projection map:** $\rho^F : U \to [0, S]$, $\rho^F(s, v) = s$.
- **Realisation map:** $\Pi : U \to \mathcal{M}$, $\Pi(s, v) = \exp_{\sigma(s)}(v)$.

Since $\mathcal{T}(\sigma) = \Pi(U)$ and Π is a diffeomorphism onto its image, these three structures transfer directly to $\mathcal{T}(\sigma)$:

- **Collection structure:** Points of $\mathcal{T}(\sigma)$ correspond bijectively to $(s, v) \in U$.
- **Projection map:** Every $x \in \mathcal{T}(\sigma)$ lies uniquely in $\exp_{\sigma(s)}(V_s)$, so $f : \mathcal{T}(\sigma) \to [0, S]$, $\rho(x) = s$, is well-defined and smooth.
- **Realisation map:** The exponential map realises each fat point in U as a point in $\mathcal{T}(\sigma)$.

Thus, $T(\sigma)$, together with the projection f and the realisation via $\exp$, satisfies all requirements of a fat manifold. □

5.6.2 Curves and their curvatures on a fat manifold

Definition 5.7 (Fat interval)**.** Let $I \subseteq \mathbb{R}$ be an interval and F be a finite-dimensional vector space, often referred to as a *fibre*. A *fat interval* I^F is the collection:

$$I^F = \{(\tau, w) : \tau \in I, w \in F\} \tag{5.121}$$

equipped with the projection map $\pi_I : I^F \to I$ defined by $\pi_I(t, w) = t$.

Definition 5.8 (Fat curve)**.** Let $U^F = \{(s, v) : s \in [0, P],\ v \in V_s \subset T^{\perp}_{\gamma(s)}\mathcal{M}\}$ be a fat manifold equipped with the projection $\pi^F : U^F \to [0, P]$, $\pi^F(s, v) = s$, and the realisation map $\Pi : U^F \to \mathcal{M}$, $\Pi(s, v) = \exp_{\gamma(s)}(v)$. Let $I \subset \mathbb{R}$ be an interval and F a finite-dimensional vector space; write $I^F = I \times F$ with projection $\pi_I : I^F \to I$, $\pi_I(\tau, w) = \tau$.

A *fat curve* (in U^F) over a base map $\mathrm{base} : I \to [0, P]$ is a smooth map

$$\gamma^F : I^F \longrightarrow U^F \tag{5.122}$$

such that the following square diagram commutes, i.e. $\pi^F \circ \gamma^F = \mathrm{base} \circ \pi_I$:

$$\begin{array}{ccc} I^F & \xrightarrow{\ \gamma^F\ } & U^F \\ {\scriptstyle \pi_I}\downarrow & & \downarrow{\scriptstyle \pi^F} \\ I & \xrightarrow{\ \mathrm{base}\ } & [0, P] \end{array}$$

Equivalently, there exists a smooth map $\xi : I \times F \to \bigcup_{s\in[0,P]} V_s$ with $\xi(\tau, w) \in V_{\mathrm{base}(\tau)}$ such that

$$\gamma^F(\tau, w) = \big(\mathrm{base}(\tau),\, \xi(\tau, w)\big). \tag{5.123}$$

The *realised fat curve in* $\mathcal{M}$ associated to γ^F is

$$\Gamma \;:=\; \Pi \circ \gamma^F \;:\; I^F \to \mathcal{M}, \qquad \Gamma(\tau, w) = \exp_{\gamma(\mathrm{base}(\tau))}\big(\,\xi(\tau, w)\,\big). \tag{5.124}$$

Note that $\circ$ is the binary operation representing the composition of two maps.

Given the fat manifold structure, one can now construct fat curves that describe how vibrational perturbations evolve along the geodesic trajectory. Consider a smooth curve $s(\tau)$ in the parameter interval $[0, P]$ representing evolution along the geodesic $\gamma(s) \subset \mathcal{M}$. For any vector $w \in F$ (where F is a finite-dimensional vector space), the *fat curve* is defined as:

$$\gamma^F : I^F \to U^F, \quad \gamma^F(\tau, w) = (s(\tau), v(\tau) + w) \tag{5.125}$$

where $v(\tau) \in T^{\perp}_{\gamma(s(t))}\mathcal{M}$ is a smooth family of perpendicular vectors that evolves along the geodesic according to the geometric structure of $\mathcal{M}$.

The fat curve $\gamma^F(\tau, w)$ represents:

1. **Base evolution**: Parameter $s(\tau)$ determines the position along the geodesic $\gamma(s)$.
2. **Vibrational perturbation**: Vector $v(\tau) + w$ represents the perpendicular deviation from the geodesic at parameter $s(\tau)$.

The actual trajectory in the configuration space is obtained through the realization map: $\Pi(\gamma^F(\tau, w)) = \exp_{\gamma(s(\tau))}(v(\tau) + w)$.

To analyse the geometry of fat curves, their tangent vectors are examined. For a fat curve $\gamma^F(\tau, w)$, the tangent vector at (τ_0, w_0) is

$$\frac{d}{d\tau}\gamma^F(\tau, w_0)\Big|_{\tau=\tau_0} = \Big(\frac{ds}{d\tau}(\tau_0), \frac{dv}{d\tau}(\tau_0)\Big) \tag{5.126}$$

This naturally decomposes into

$$\text{base component: } \frac{ds}{d\tau}(\tau_0) \quad \text{and} \quad \text{fibre component: } \frac{dv}{d\tau}(\tau_0). \tag{5.127}$$

Induced Metric Construction. The fat manifold U^F is equipped with a Riemannian metric $\tilde{g}$ by demanding that the realization map

$$\Pi : U^F \longrightarrow \mathcal{M} \tag{5.128}$$

preserves lengths infinitesimally. Concretely, for tangent vectors $X, Y \in T_{(s,v)}U^F$,

$$\tilde{g}_{(s,v)}(X, Y) = g_{\Pi(s,v)}\big(\Pi_* X, \Pi_* Y\big), \tag{5.129}$$

where g is the Jacobi metric on $\mathcal{M}$ and Π_* its differential.

Natural Horizontal–Vertical Decomposition. At each point $(s, v) \in U^F$, the tangent space splits orthogonally as

$$T_{(s,v)}U^F = H_{(s,v)} \oplus V_{(s,v)}. \tag{5.130}$$

Vertical space: $V_{(s,v)} = \ker(\pi^F{}_*) = \{(0, u) \mid u \in T^{\perp}_{\gamma(s)}\mathcal{M}\}$, representing pure fibre–vibrational motion.

Horizontal space: $H_{(s,v)} = V^{\perp_{\tilde{g}}}_{(s,v)}$, representing the evolution along the geodesic parameter.

The Riemannian metric $\tilde{g}$ on U^F determines a unique Levi-Cività connection. If $\Gamma(TU^F)$ denotes the space of smooth sections of the tangent bundle $TU^F = \bigcup_{r\in U^F} T_rU^F$, i.e., smooth maps $X : U^F \to TU^F$ such that $X(r) \in T_rU^F$ for each $r \in U^F$. Equivalently, $\Gamma(TU^F)$ is the space of smooth vector fields on U^F. The Levi-Cività connection is then the unique bilinear map

$$\tilde{\nabla} : \Gamma(TU^F) \times \Gamma(TU^F) \longrightarrow \Gamma(TU^F), \tag{5.131}$$

which is torsion-free and metric-compatible. That is, for any vector fields $X, Y, Z \in \Gamma(TU^F)$, it satisfies $\tilde{\nabla}_X Y - \tilde{\nabla}_Y X = [X, Y]$ (torsion-free condition) and $X\langle Y, Z\rangle_{\tilde{g}} = \langle \tilde{\nabla}_X Y, Z\rangle_{\tilde{g}} + \langle Y, \tilde{\nabla}_X Z\rangle_{\tilde{g}}$ (metric-compatibility), where $[X, Y]$ denotes the Lie bracket of the vector fields defined by $[X, Y]f = X(Yf) - Y(Xf)$ for an arbitrary smooth function f. It tells us how to differentiate vector fields along curves in U^F while preserving $\tilde{g}$.

With this in mind, one can then define a decomposition of curvature of the vector fields in terms of horizontal and vertical space contributions. This was first introduced to the world by Barret O'Neill in his work on Riemannian submersions in 1966: see [141].

Definition 5.9 (O'Neill Tensors)**.** Let X, Y be vector fields on U^F, and write

$$X = X^H + X^V, \quad Y = Y^H + Y^V, \tag{5.132}$$

for their orthogonal projections onto the horizontal and vertical subspaces.

The *O'Neill tensors*

$$\mathcal{A}, \mathcal{T} \colon \Gamma(TU^F) \times \Gamma(TU^F) \longrightarrow \Gamma(TU^F) \tag{5.133}$$

are defined by

$$\begin{aligned} \mathcal{A}_X Y &= \left(\tilde{\nabla}_{X^H} Y^V\right)^H \; + \; \left(\tilde{\nabla}_{X^V} Y^H\right)^V, \\ \mathcal{T}_X Y &= \left(\tilde{\nabla}_{X^H} Y^H\right)^V \; + \; \left(\tilde{\nabla}_{X^V} Y^V\right)^H, \end{aligned} \tag{5.134}$$

where the superscripts H, V denote horizontal and vertical components.

The tensor $\mathcal{A}$ encodes how horizontal and vertical motions mix. Its first component,

$$(\tilde{\nabla}_{X^H} Y^V)^H, \tag{5.135}$$

measures the way a horizontal push along the base causes a change in a vertical (vibrational) field, which is then projected back onto the horizontal subspace. Dually, the second component,

$$(\tilde{\nabla}_{X^V} Y^H)^V, \tag{5.136}$$

records how a pure vertical perturbation of the vibrational field induces a change in the horizontal direction and then is projected back to vertical.

By contrast, the tensor $\mathcal{T}$ captures the failure of the horizontal and vertical distributions to be integrable. Specifically,

$$(\tilde{\nabla}_{X^H} Y^H)^V \tag{5.137}$$

quantifies how two horizontal motions, when composed, *leak* into the vertical direction (i.e. generate vibrations), while

$$(\tilde{\nabla}_{X^V} Y^V)^H \tag{5.138}$$

measures how two vertical (vibrational) motions nevertheless produce a horizontal drift, altering the underlying trajectory.

Mode–Coupling Criterion. Two vibrational modes $\mathbf{e}_i, \mathbf{e}_j$ in the vertical bundle are said to be *coupled* exactly when

$$\left|\mathcal{A}_{\mathbf{e}_i}\mathbf{e}_j\right| \; + \; \left|\mathcal{T}_{\mathbf{e}_i}\mathbf{e}_j\right| > 0. \tag{5.139}$$

When modes are coupled, they collectively define a *fat submanifold* around the base trajectory $\gamma(s)$. Specifically, the coupled modes $\{\mathbf{e}_i\}$ span the vertical bundle $V_{(s,v)}$, and the exponential map $\exp_{\gamma(s)}(v)$ for $v \in V_{(s,v)}$ constructs a tubular neighborhood around $\gamma(s)$. The cross-sectional geometry of this tube at each parameter value s is determined by the span of the coupled modes, while the O'Neill tensors quantify how this cross-sectional shape evolves and deforms as the trajectory progresses.

In particular, the coupling between the trajectory direction $\partial/\partial s \in H$ and a vibrational mode $\mathbf{e}_i$ is quantified by

$$\left|\mathcal{A}_{\partial/\partial s}\mathbf{e}_i\right| + \left|\mathcal{T}_{\mathbf{e}_i}\partial/\partial s\right|. \tag{5.140}$$

This measures how motion along the trajectory ($\partial/\partial s$) induces changes in the vibrational mode $\mathbf{e}_i$, and conversely, how vibrational motion affects the trajectory direction. The resulting fat manifold exhibits geometric deformation—twisting, expansion, or contraction of the tubular cross-sections—precisely when this coupling is non-zero.

Finally, the *total coupling strength* at parameter s is defined by choosing an orthonormal basis $\{\mathbf{e}_i\}$ of $V_{(s,v)}$ and setting

$$\kappa_{\text{coupling}}(s) = \sum_{i,j}\Big(\left|\mathcal{A}_{\mathbf{e}_i}\mathbf{e}_j\right|^2 + \left|\mathcal{T}_{\mathbf{e}_i}\mathbf{e}_j\right|^2\Big). \tag{5.141}$$

This scalar quantity provides a comprehensive measure of mode coupling at each point along the trajectory and directly relates to the geometric deformation rate of the fat manifold's tubular structure.

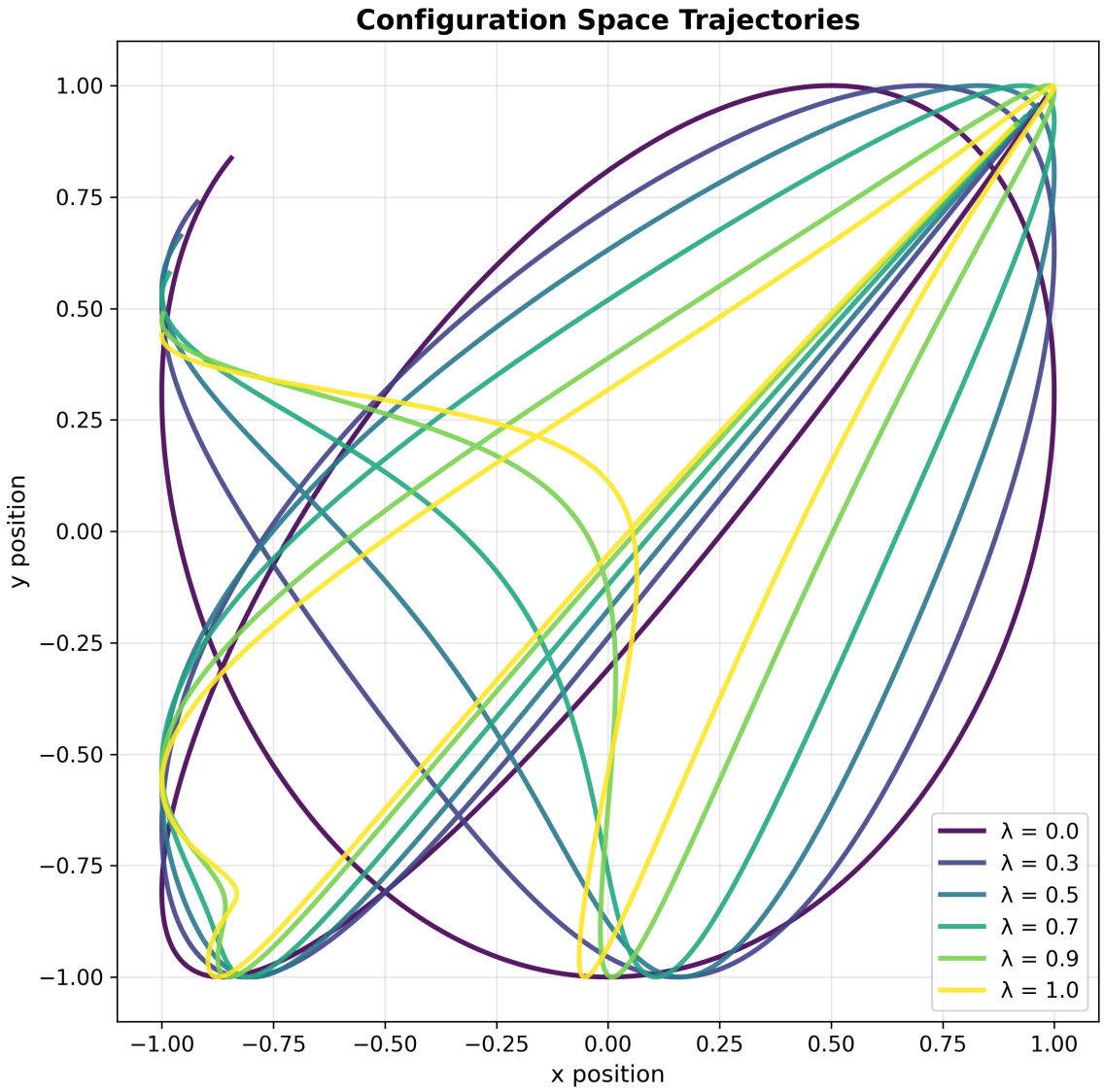


Figure 5.12. *Configuration space trajectories for the coupled harmonic oscillator system (Equation 5.142) with varying coupling strength $\lambda \in [0, 1]$. As λ increases, the trajectories transition from simple Lissajous curves ($\lambda = 0$) to complex multiply-wound patterns, reflecting the increasing mode coupling. Each trajectory serves as the base curve $\gamma(s)$ for constructing the corresponding fat manifold via the exponential map in the perpendicular directions spanned by the coupled vibrational modes.*

5.6.3 Illustrative example: coupled harmonic oscillators

The geometric framework of fat manifolds and O'Neill tensors can be illustrated through a simple two-dimensional coupled harmonic oscillator system with Hamiltonian

$$H = \frac{1}{2}(p_x^2 + p_y^2) + \frac{1}{2}(\omega_1^2 x^2 + \omega_2^2 y^2) + \lambda xy, \tag{5.142}$$

where λ represents the coupling strength between the two oscillatory modes. The corresponding Hamilton's equations are

$$\dot{x} = p_x, \quad \dot{y} = p_y, \quad \dot{p}_x = -\omega_1^2 x - \lambda y, \quad \dot{p}_y = -\omega_2^2 y - \lambda x. \tag{5.143}$$

For weak coupling ($\lambda \ll \omega_1, \omega_2$), the essential mode-mixing dynamics are captured by the approximate solution of the Hamilton's equations:

$$x(t) = \cos(\omega_1 t + \lambda \sin(\omega_2 t)), \quad y(t) = \cos(\omega_2 t + \lambda \sin(\omega_1 t)). \tag{5.144}$$

This ansatz exhibits frequency modulation characteristic of weakly coupled oscillators, where each mode's phase is modulated by the other mode's motion.

Geometric framework The configuration space trajectory $\gamma(t) = (x(t), y(t))$, with x and y given as in Equation (5.144), defines a curve in the two-dimensional configuration manifold. The velocity vector $\dot{\gamma}(t) = (\dot{x}(t), \dot{y}(t))$ determines the arc length element $ds = \|\dot{\gamma}(t)\|dt$. The associated unit tangent and normal vectors are

$$T(t) = \frac{\dot{\gamma}(t)}{\|\dot{\gamma}(t)\|}, \quad N(t) = \frac{(-\dot{y}(t), \dot{x}(t))}{\|\dot{\gamma}(t)\|}, \tag{5.145}$$

which form an orthonormal basis for the tangent space at each point along the trajectory.

O'Neill tensor analysis In the flat Euclidean configuration space, the Levi-Civita connection reduces to ordinary differentiation. The horizontal and vertical projections are defined with respect to the tangent space decomposition $T_{\gamma(t)}\mathbb{R}^2 = \text{span}\{T(t)\} \oplus \text{span}\{N(t)\}$. The relevant O'Neill tensor component is

$$\mathcal{A}_T N = (\nabla_T N)^H = \langle \nabla_T N, T \rangle T, \tag{5.146}$$

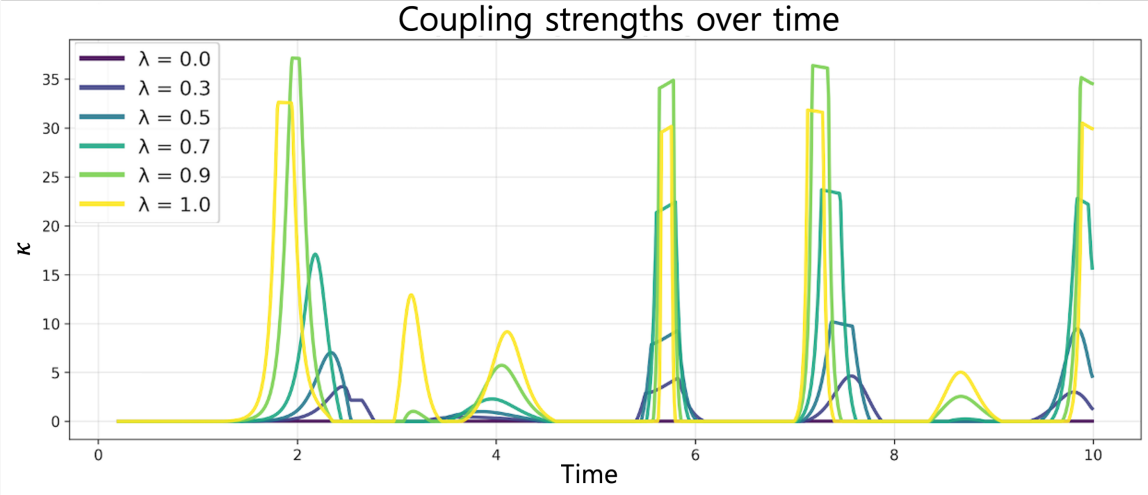


Figure 5.13. *Evolution of geometric coupling strength, $\kappa = \|\mathcal{A}_T N\|^2$, over time for the coupled harmonic oscillator system (Equation 5.142) with $\omega_1 = 1.0, \omega_2 = 1.2$ shown for coupling parameters $\lambda \in [0, 0.3, 0.5, 0.7, 0.9, 1.0]$. The periodic structure reflects the underlying oscillatory dynamics, with peak amplitudes scaling approximately as λ^2 for weak coupling. For $\lambda = 0$ (purple), the coupling constant vanishes identically to 0, confirming the absence of mode coupling in the decoupled system.*

where the superscript H denotes horizontal projection. The tensor $\mathcal{T}$ vanishes identically in two dimensions due to the absence of multiple vertical directions.

The connection $\nabla_T N = \frac{d}{dt} N(t)$ quantifies how the normal direction evolves due to mode coupling effects. Critically, when $\lambda = 0$, the O'Neill tensor vanishes throughout the entire trajectory, $\mathcal{A}_T N = 0$, reflecting the complete absence of mode coupling in the decoupled system. For $\lambda > 0$, the tensor captures the dynamical coupling between modes through the rotation of the normal vector field.

Coupling strength and physical interpretation The mode coupling strength is quantified by

$$\kappa_{\text{coupling}} = \|\mathcal{A}_T N\|^2 + \|\mathcal{T}_N N\|^2 = \|\mathcal{A}_T N\|^2, \tag{5.147}$$

where the second term vanishes in two dimensions. Figure 5.13 demonstrates the time evolution of κ_{coupling} for different coupling strengths.

Several key features emerge from this analysis:

- For the decoupled case ($\lambda = 0$), κ_{coupling} is identically zero, confirming that no mode mixing occurs when the coupling term is absent.
- For $\lambda > 0$, the coupling strength exhibits periodic behavior with frequency components related to ω_1 and ω_2, reflecting the underlying oscillatory dynamics.
- The peak coupling strength scales approximately as λ^2 for weak coupling, consistent with perturbative expectations.
- The temporal pattern reveals when modes are most strongly coupled, corresponding to specific phase-dependences between the oscillators.

The volume evolution rate of the tubular neighborhood is given by

$$\frac{dV}{ds} = -\text{tr}(\mathcal{A}_T) = -\langle \mathcal{A}_T N, N \rangle, \tag{5.148}$$

which vanishes when $\mathcal{A}_T N$ is purely horizontal (as in this example), indicating volume preservation. This reflects both the Hamiltonian nature of the dynamics and the specific geometric properties of the coupling.

The configuration space trajectories show the transition from simple Lissajous-like figures ($\lambda = 0$) to increasingly complex, multiply-wound patterns as the coupling strength increases. These geometric changes in the base trajectory directly correlate with the magnitude of the O'Neill tensor components, establishing a clear connection between classical trajectory geometry and mode coupling quantification through the fat manifold framework.

5.6.4 Carbon dioxide and coupled vibrational modes

In this section, the coupling between vibrational modes is analyzed within the fat manifold framework by focusing on regions where the coupling condition

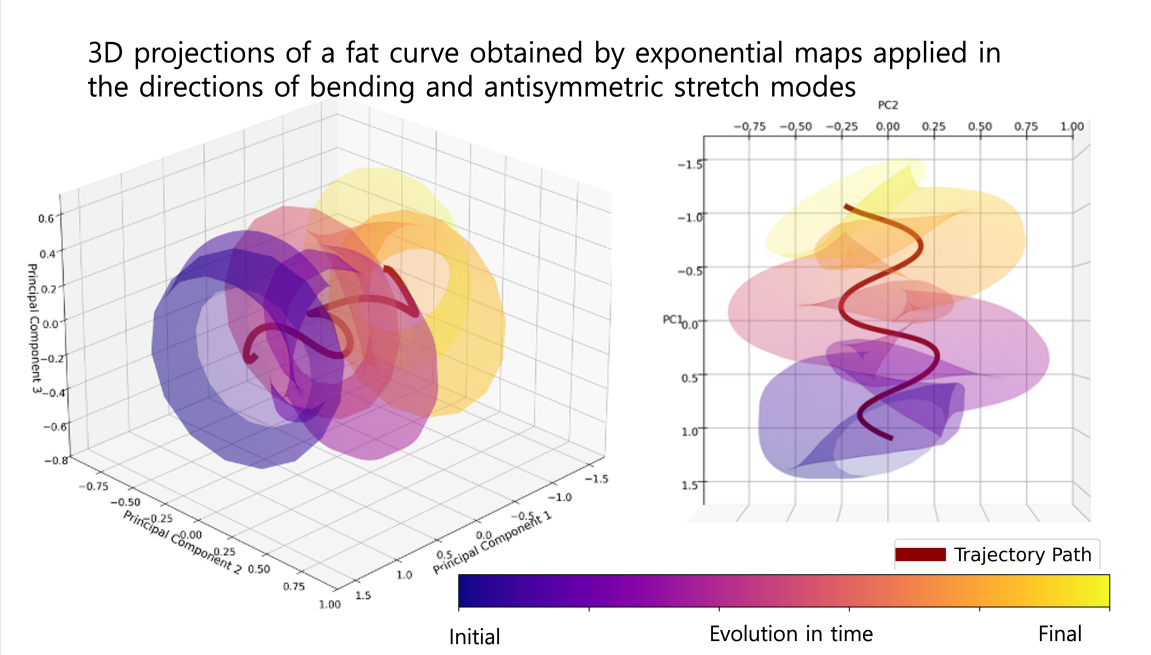


Figure 5.14. *3D projection of fat submanifold around a geodesic taken by the carbon dioxide molecule with the exponential map directions given by the initial bending and symmetric stretch mode directions*.

of Equation (5.140) is satisfied. The analysis considers trajectories within the tubular neighbourhood constructed via the exponential map $\exp_{\gamma(s)}(v)$ around the base geodesic $\gamma(s)$, where the horizontal and vertical space decomposition captures the distinction between trajectory evolution and vibrational motion.

When modes are strongly coupled, the exponential map generates a substantial tubular neighbourhood around the base trajectory, creating a genuine ”fat” manifold with non-trivial cross-sectional geometry. However, as the coupling strength diminishes and the total coupling parameter $\kappa_{\text{coupling}}(s)$ approaches zero, the tubular neighborhood contracts. In the limit of completely decoupled modes, the fat manifold degenerates: the perpendicular directions cease to contribute meaningfully to the dynamics, the tubular structure collapses, and the fat manifold reduces to the one-dimensional base trajectory $\gamma(s)$ itself. Thus, the ”fatness” of the manifold serves as a geometric indicator

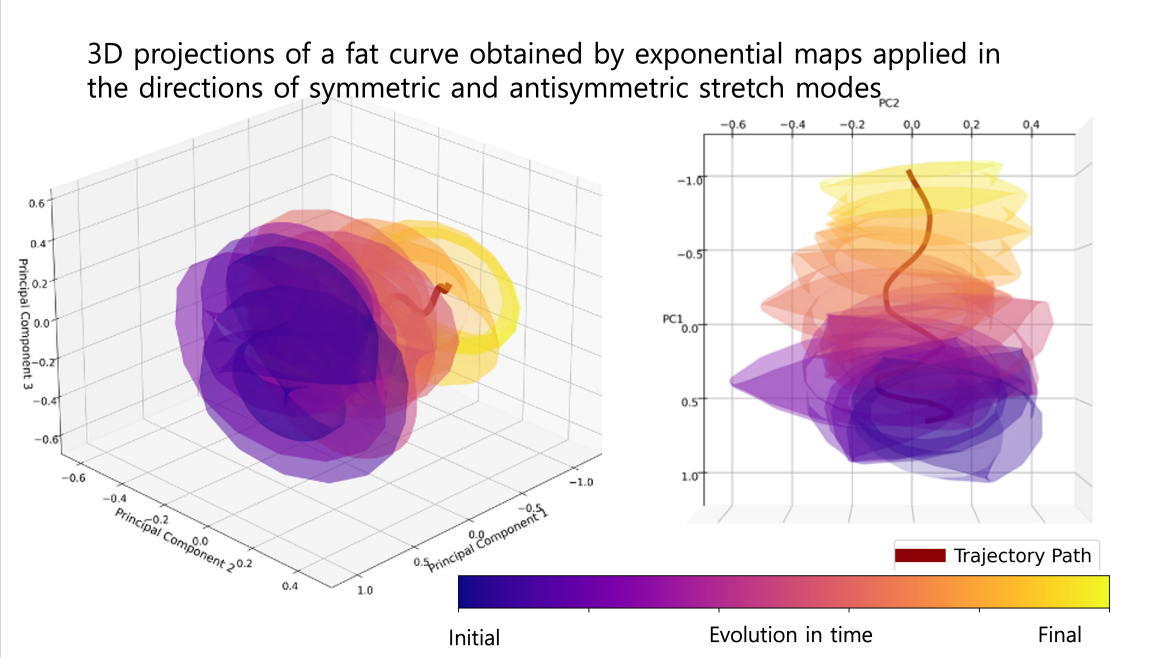


Figure 5.15. *3D projection of fat submanifold around a geodesic taken by the carbon dioxide molecule with the exponential map directions given by the initial antisymmetric stretch and symmetric stretch mode directions. The fat manifold exhibits well-defined periodic volume evolutions around a geodesic.*

of mode coupling strength—robust coupling sustains a substantial tubular neighborhood, while decoupled modes eliminate the transverse structure entirely.

Figures 5.14 and 5.15 present the three-dimensional projections of fat manifold envelopes surrounding the geodesic corresponding to the trajectory of carbon dioxide. The geodesic was obtained using the methods described in the preceding sections of this chapter. At time zero in both figures, the thermostat—initially set to $300\,\mathrm{K}$—was switched off. Thereafter, the system was propagated in the microcanonical (NVE) ensemble, enabling the analysis of vibrational mode couplings in the production stage of the molecular dynamics simulation. Figure 5.14 displays a relatively simple and regular fat manifold envelope structure, arising from the coupling between the bending and sym-

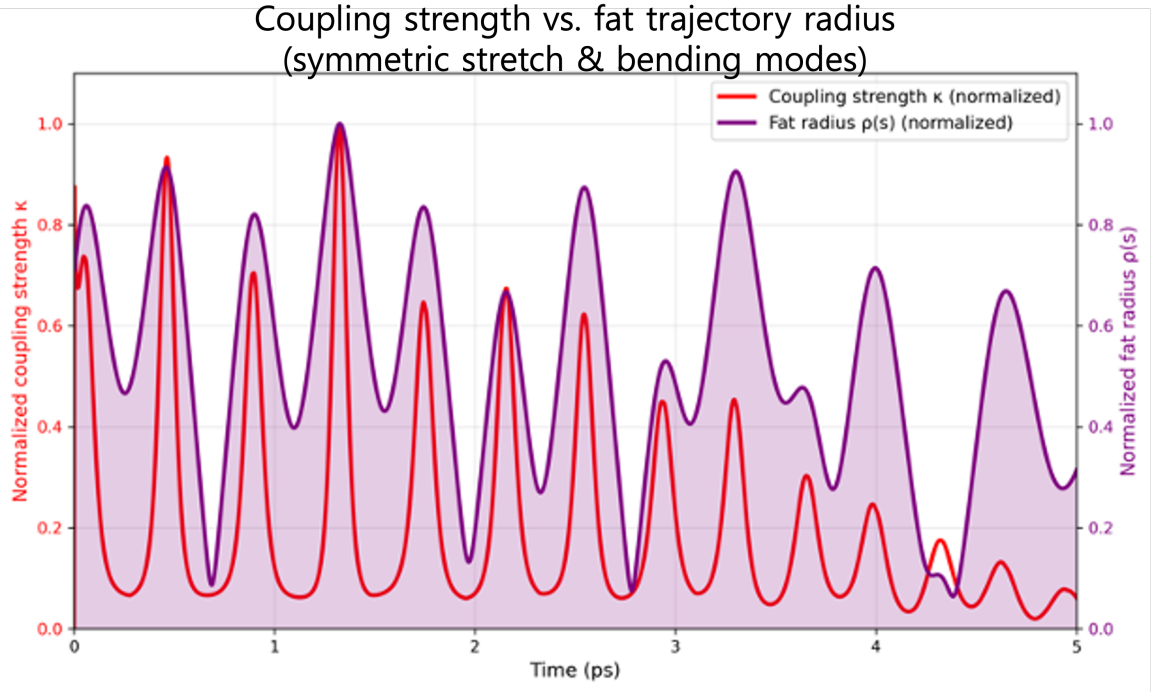


Figure 5.16. *Evolution of the normalised mode coupling between symmetric stretch and bending modes over time* (*red*) *and the normalised radius of the fat trajectory* (*purple*).

metric stretch modes. In contrast, Figure 5.15 demonstrates, at a glance, a more pronounced variation in the fat manifold structure.

Figures 5.16 and 5.17 compare temporal evolution of the coupling strengths to the radii of the envelope surrounding the geodesic.

Regarding the similarities between Figures 5.16 and 5.17, it is evident that the magnitude of the coupling strengths progressively decreases over time. This behaviour is expected because the thermostat was switched off at $t = 0$, when the couplings were maximised by thermal excitation. As the system relaxes and the effect of the initial temperature diminishes, the couplings weaken until equilibrium is reached. The alignment of the peaks with the fat trajectory radii further illustrates this relaxation process.

The most prominent differences between the two figures emerge beyond

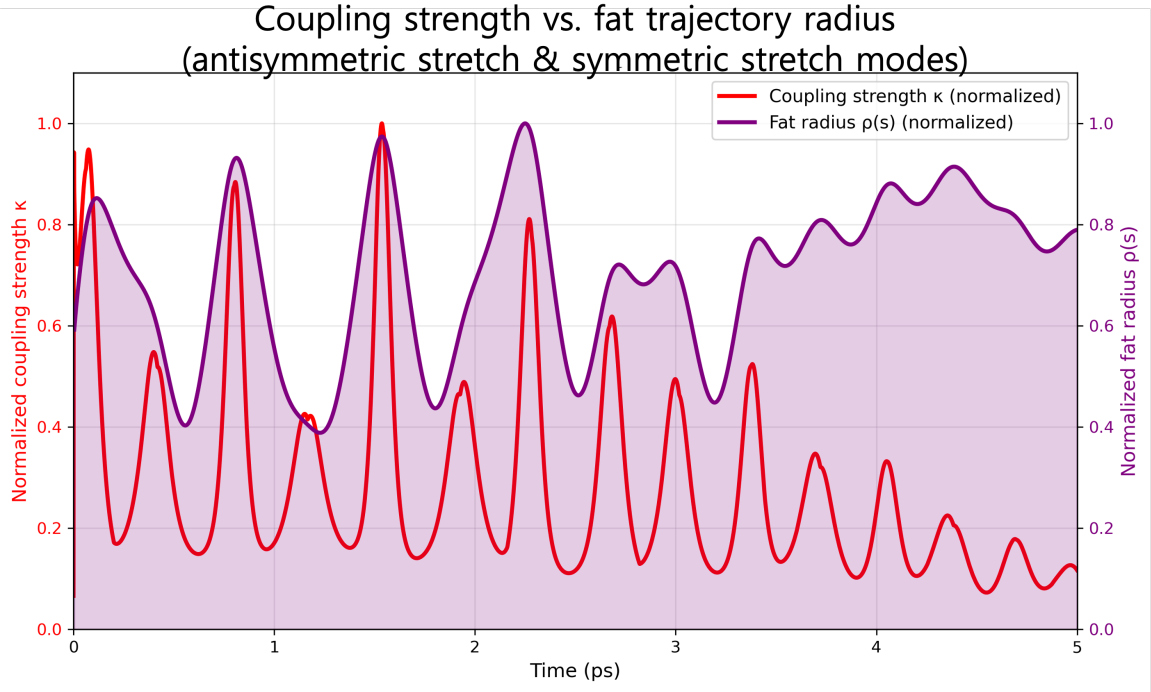


Figure 5.17. *Evolution of the normalised mode coupling between symmetric and antisymmetric stretch modes over time* (*red*) *and the normalised radius of the fat trajectory* (*purple*).

$4\,\mathrm{ps}$. In this time regime, the fat trajectory radius in Figure 5.16 exhibits no substantial changes in magnitude, although the peak heights gradually decrease. By contrast, the fat trajectory radius in Figure 5.17 appears to plateau toward a steady value, indicating that the coupling strength has diminished significantly. Figure 5.18 confirms this observation. In panel a), the magnitude of the time derivative of the fat trajectory radius decreases sharply, whereas panel b) also exhibits a decline, but the reduction is markedly less pronounced than in panel a). Together, these observations indicate that the coupling in Figure 5.16 dissipates more rapidly, while in Figure 5.17 the system retains residual coupling effects that persist even after the primary relaxation phase.

As the framework developed in this work is purely mathematical, the same analysis can be applied to modes that do not correspond to vibrational

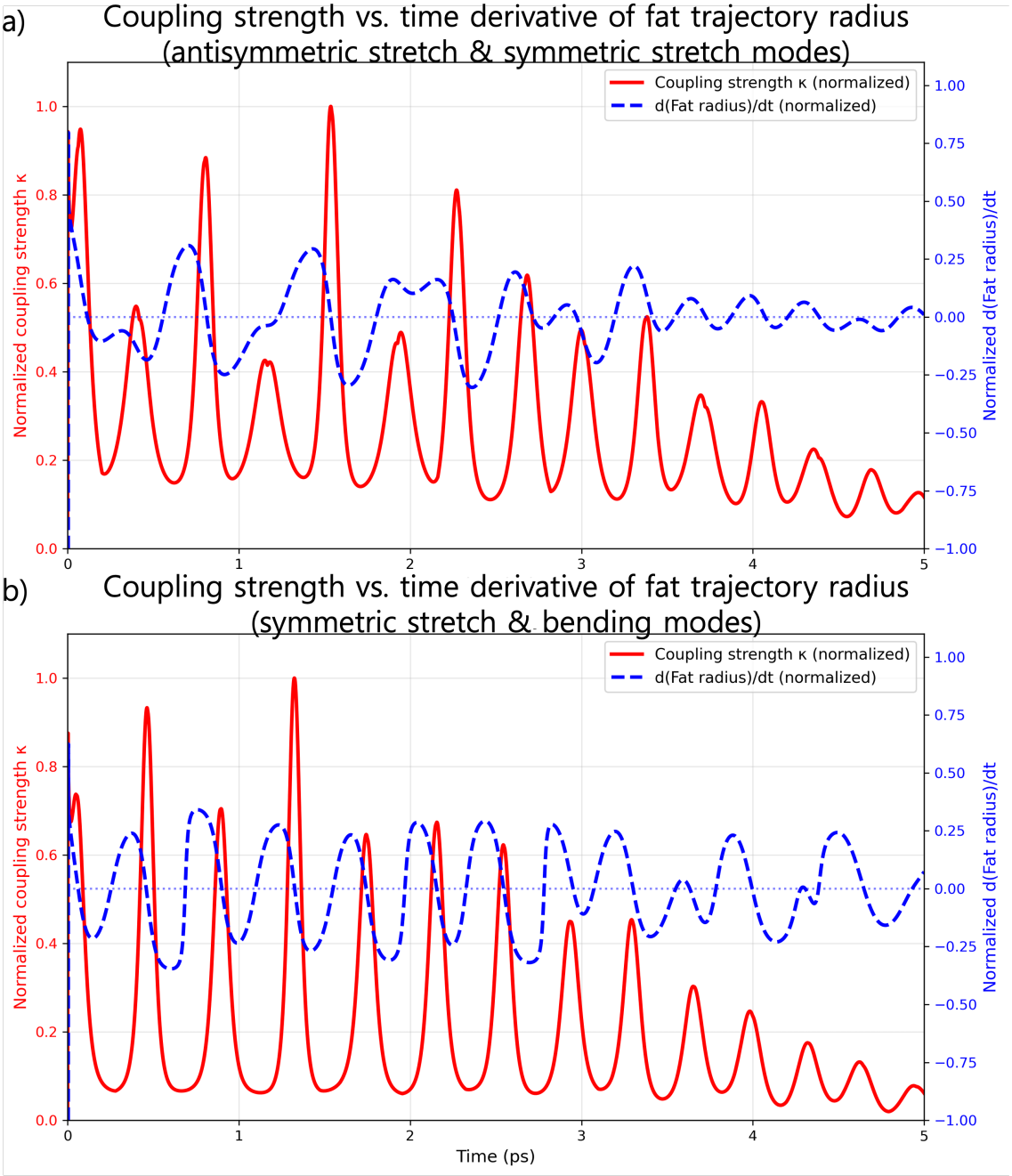


Figure 5.18. *Evolution of the time-derivatives of fat trajectory radii for a) case corresponding to Figure 5.16 and b) to Figure 5.17.*

modes. Figure 5.19 illustrates such a case for the two translational Ricci modes. Here, the undulation of the fat radius (panel a) shows no indication of convergence—neither in the persistence of peak shapes nor in the behaviour of the time derivative of the fat radius (panel b), which does not approach zero. This behaviour is physically reasonable, as translational motion should not exhibit damping or thermal equilibration in an isolated system; rather, it persists without attenuation in the absence of external forces or dissipative mechanisms.

The results in this section show that the fat manifold framework can quantify and track mode couplings over time using the fat trajectory radius and its time derivative. These outputs reveal both the strength and persistence of couplings, providing a clear basis for distinguishing rapidly relaxing interactions from those that remain over longer timescales.

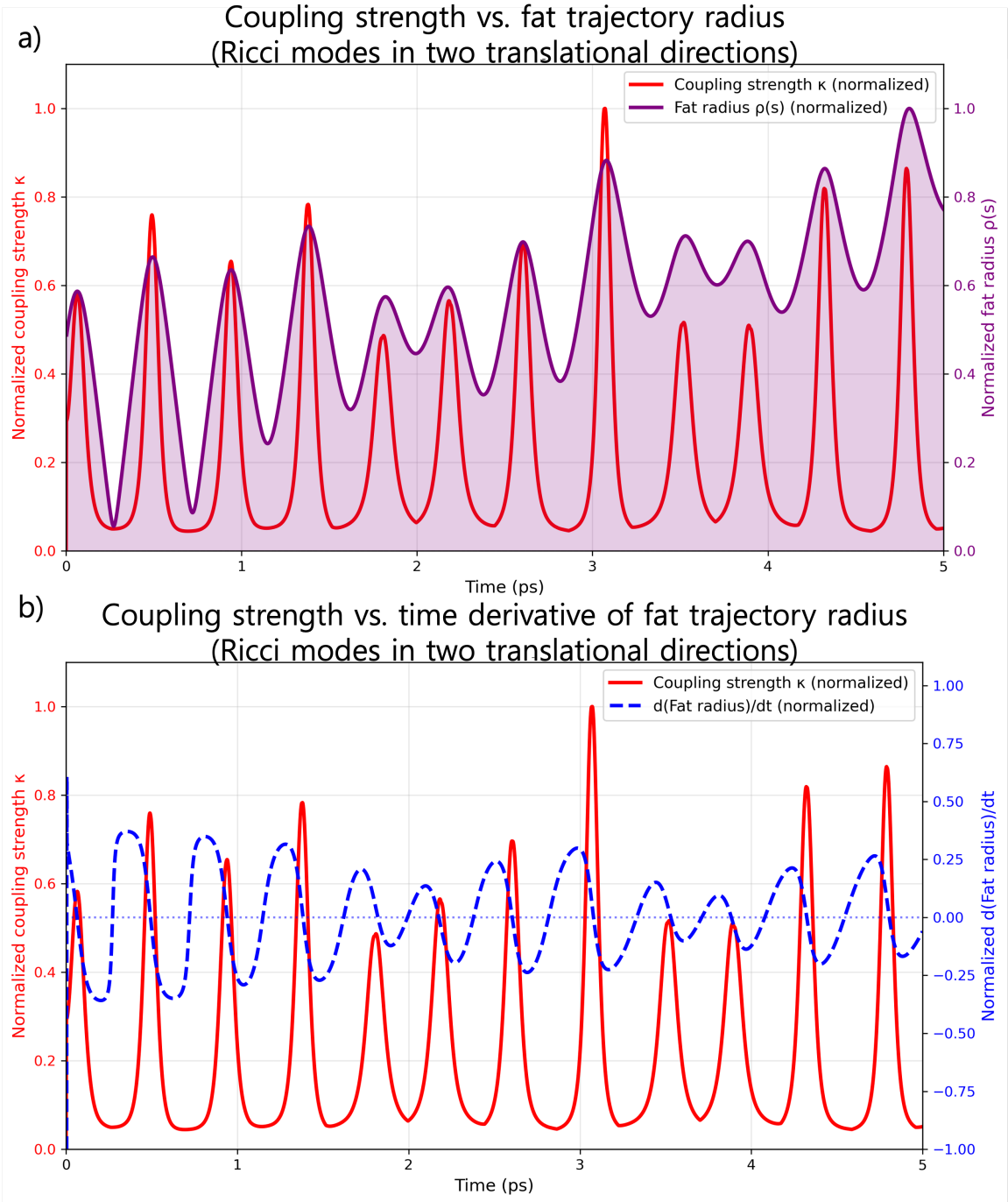


Figure 5.19. *Evolution of a) the radii and b) the time derivatives of the radii of the fat trajectory corresponding to the two translational Ricci modes. The coupling strength between the two translational modes is shown in red.*

CHAPTER 6

Concluding remarks

道尸掃尸星利望良古 彗星也白反也人是有叱多 後句 達阿羅浮去伊叱等邪 此也友物比所音叱兮叱只有叱故	*Someone*, *gazing at a star sweeping the path*, *exclaimed*, "*A comet*!" *Ah*, *but the moon had already risen far below*. *Look here—what kind of comet could it have been*?
융천사 融天師 혜성가 彗星歌	YUNGCHEONSA *Hyeseong-ga* (*Song of a Comet*)

This thesis set out to explore a series of questions concerning the geometrical and spectral understanding of molecular vibrations, as introduced in Chapter 1.1. In this concluding section, the previously posed questions are revisited in light of the results presented in the preceding chapters, with emphasis on the key insights gained, the challenges encountered, and the broader implications of the methodologies developed.

Q1. Is it possible to extract a physically meaningful decomposition of molecular spectral data without prior knowledge of the underlying chemistry or physics? Chapter 3 demonstrates that it is indeed possible to extract a physically meaningful decomposition of molecular spectral data without relying on prior chemical or physical assumptions. By treating the vibrational spectrum purely as a signal, the thesis develops an empirical mode decomposition framework that identifies spectral peaks through robust peak-finding

and smoothing algorithms, followed by a Lorentzian mixture fitting. This method isolates individual vibrational contributions based solely on the geometry and statistical features of the signal. Applications to both simple and complex molecules, including carbon dioxide and phenol, show that the extracted modes correspond closely to known vibrational structures, even without input from quantum calculations or symmetry considerations. This confirms the viability of a data-driven, model-agnostic approach to vibrational analysis.

Q2. Can molecular trajectories be isolated within a specific frequency band without compromising their physical integrity? Chapters 4.1 through 4.3 demonstrate that molecular trajectories can indeed be isolated within a specific frequency band without compromising their physical integrity. This is achieved through a Fourier integrator framework, where molecular motion is expressed in terms of its vibrational frequency components and selectively evolved within a user-defined band. By projecting both positions and forces onto a finite Fourier basis corresponding to target frequencies, the method enables band-limited simulations that preserve the symplectic structure and energy coherence of the system. Notably, the algorithm maintains physical realism by computing forces in real space and consistently mapping them back into Fourier space, ensuring that the simulated dynamics remain physically grounded.

Case studies involving carbon dioxide and phenol validate the approach: in each instance, restricting the simulation to low, mid, or high-frequency bands produced vibrational densities of states in close agreement with the relevant part of the full-spectrum simulations. Even in the presence of anharmonic couplings, the band-limited dynamics retain structural fidelity, with inter-mode interactions manifesting as identifiable off-diagonal correlations. These results confirm that selective vibrational analysis can be conducted without artificial distortion, offering both computational efficiency and interpretive clarity in molecular dynamics.

Q3. Can molecular trajectories be described usefully in geometrical terms? The results presented in Chapters 5.1 through 5.3 demonstrate that molecular trajectories can indeed be reformulated entirely in geometrical terms. By adopting the Jacobi metric—a Riemannian metric derived from the system's kinetic and potential energy—molecular dynamics are reinterpreted as geodesic flows on a compact Riemannian manifold.

This formulation allows the action principle from classical mechanics to be recast as a problem of arc-length extremisation, where the physical trajectory corresponds to a geodesic path on the energy landscape. Importantly, the configurational space of a molecule under this metric is shown to be compact with a well-defined boundary (the Hill region), ensuring that the system's trajectories remain bounded. This geometrisation not only provides a conceptually elegant framework but also offers numerical advantages, such as intrinsic energy conservation and stable integration.

Furthermore, the introduction of a quantum-uncertainty-based thermostat compatible with this geometric framework (Chapter 5.2) enables physically meaningful thermal sampling, offering a potential route out of the dilemma between non-ergodic deterministic dynamics and statistical models with uncontrolled time-correlation behaviour. The case study of carbon dioxide (Chapter 5.3) illustrates the practical viability of this approach: vibrational spectra obtained via thermostatted geodesic integration closely match those from conventional molecular dynamics, validating the geometric framework as both theoretically rigorous and computationally robust.

Q4. Can the internal vibrational motions of a molecule be understood as manifestations of quasi-periodicity? Chapters 5.4 through 5.6 of this thesis provide a compelling geometrical framework that links internal vibrational motions of molecules to quasi-periodic dynamics. Moving beyond the harmonic approximation and static normal mode analysis, the approach introduced here leverages the Ricci tensor—defined on the Riemannian manifold endowed

with the Jacobi metric—to generalise normal modes into smoothly evolving eigen-directions that track the intrinsic vibrational structure of the molecule throughout its trajectory. These Ricci eigendirections provide a coordinate-invariant and physically meaningful basis that remains valid even in highly anharmonic regimes.

To detect and characterise quasi-periodicity, this thesis employs two complementary tools: the holonomy angle, which quantifies geometric phase accumulation via parallel transport, and tubular trajectory analysis, which evaluates the volume overlap of geodesic neighbourhoods during recurrent motion.

In addition to these tools, this thesis develops the fat manifold framework, which quantifies the temporal evolution of mode couplings through the fat trajectory radius and its time derivative. These quantities provide direct, time-resolved measures of how vibrational interactions weaken or persist, offering both a visual and numerical basis for assessing the stability of mode interactions. By tracking the decay rates and persistence of couplings, the framework complements quasi-periodicity detection by identifying whether recurrent motions remain dynamically coherent over extended timescales or gradually lose synchronisation.

Together, these approaches reveal that even in thermally fluctuating systems, vibrational motions often recur along low-dimensional invariant structures in configuration space. Well-resolved spectral peaks are found to coincide with high quasi-periodicity scores, closed tubular trajectories, and persistent mode couplings identified through the fat manifold framework. This convergence of geometric, spectral, and dynamical indicators confirms that persistent vibrational modes are underpinned by near-periodic geodesic motion, and that their stability can be quantitatively assessed through the coupled evolution of fat trajectory radii. Thus, the internal vibrational dynamics of a molecule can be rigorously understood as quasi-periodic flows embedded in the curved geometry of the molecular energy landscape.

Efforts are currently underway to generalise the learning of potential energy surfaces (PES) within the Born–Oppenheimer (BO) approximation to the learning of full dynamical maps that go beyond BO. The present work offers a framework for representing an arbitrary dynamical system, potentially including beyond-BO quantum effects, in a way that is geometric, transparent, and easy to interrogate.

6.1 The shape of motion, the sound of structure

Molecular vibrations unfold not only in quantum states or numerical spectra, but through structured motions shaped by geometry and resonance. Trajectories traced as geodesics on energy landscapes, vibrations understood as quasi-periodic flows, and spectra treated as signals reveal a unified view where structure and motion reflect one another. Methods developed in this work—Ricci eigenmode tracking, monodromy of tubular trajectories, frequency band isolation, and empirical signal decomposition—highlight how geometry and signal intertwine. Within this interplay, motion gains shape and structure finds voice, offering a foundation for deeper explorations into the geometry of dynamics.

References

[1] K. Han, A. Boziki, A. Tkatchenko, and J. T. Berryman, “Tihi toolkit: A peak finder and analyzer for spectroscopic data,” *ACS omega*, vol. 9, no. 50, pp. 49 397–49 410, 2024.

[2] G. Keresztury, F. Billes, M. Kubinyi, and T. Sundius, “A density functional, infrared linear dichroism, and normal coordinate study of phenol and its deuterated derivatives: revised interpretation of the vibrational spectra,” *The Journal of Physical Chemistry A*, vol. 102, no. 8, pp. 1371–1380, 1998.

[3] J. Tyndall, *On the absorption and radiation of heat by gases and vapours*, *and on the physical connexion of radiation*, *absorption*, *and conduction*, 1861.

[4] “Ueber das verhältniss zwischen dem emissionsvermögen und dem absorptionsvermögen der körper für wärme and licht,” *Annalen der Physik und Chemie*, vol. 185, no. 2, pp. 275–301, 1860.

[5] N. Bohr, “On the constitution of atoms and molecules,” *Philosophical Magazine*, vol. 26, no. 1, pp. 1–25, 1913.

[6] A. Sommerfeld, “Zur quantentheorie der spektrallinien,” *Annalen der Physik*, vol. 356, no. 17, pp. 1–94, 1916.

[7] A. Kratzer, “Die ultraroten rotationsspektren der halogenwasserstoffe,” *Zeitschrift für Physik*, vol. 3, no. 5, pp. 289–307, 1920.

[8] M. Born and J. R. Oppenheimer, “Zur quantentheorie der molekeln,” *Annalen der Physik*, vol. 389, no. 20, pp. 457–484, 1927.

[9] R. T. Cygan, V. N. Romanov, and E. M. Myshakin, “Molecular simulation of carbon dioxide capture by montmorillonite using an accurate and flexible force field,” *The Journal of Physical Chemistry C*, vol. 116, no. 24, pp. 13 079–13 091, 2012.

[10] V. I. Arnol’d, *Mathematical methods of classical mechanics*. Springer Science & Business Media, 2013, vol. 60.

[11] J. Liouville, “Note sur la théorie de la variation des constantes arbitraires,” *Journal de mathématiques pures et appliquées*, vol. 3, pp. 342–349, 1838.

[12] E. Hairer, M. Hochbruck, A. Iserles, and C. Lubich, “Geometric numerical integration,” *Oberwolfach Reports*, vol. 3, no. 1, pp. 805–882, 2006.

[13] B. Leimkuhler and S. Reich, *Simulating Hamiltonian Dynamics*. Cambridge University Press, 2004.

[14] J. M. Sanz-Serna, “Symplectic integrators for hamiltonian problems: An overview,” *Acta Numerica*, vol. 1, pp. 243–286, 1992.

[15] R. D. Ruth, “A canonical integration technique,” *IEEE Transactions on Nuclear Science*, vol. 30, pp. 2669–2671, 1983.

[16] H. Yoshida, “Construction of higher order symplectic integrators,” *Physics Letters A*, vol. 150, no. 5–7, pp. 262–268, 1990.

[17] H. F. Trotter, “On the product of semi-groups of operators,” *Proceedings of the American Mathematical Society*, vol. 10, no. 4, pp. 545–551, 1959.

[18] M. Tuckerman, B. J. Berne, and G. J. Martyna, “Reversible multiple time scale molecular dynamics,” *The Journal of Chemical Physics*, vol. 97, pp. 1990–2001, 1992.

[19] T. d. J. Stuchi, “Symplectic integrators revisited,” *Brazilian Journal of Physics*, vol. 32, pp. 958–979, 2002.

[20] D. Cottrell and P. Tupper, “Energy drift in molecular dynamics simulations,” *BIT Numerical Mathematics*, vol. 47, no. 3, pp. 507–523, 2007.

[21] M. Sri Harish and P. K. Patra, “Temperature and its control in molecular dynamics simulations,” *Molecular Simulation*, vol. 47, no. 9, pp. 701–729, 2021.

[22] J. Wong-Ekkabut and M. Karttunen, “The good, the bad and the user in soft matter simulations,” *Biochimica et Biophysica Acta* (*BBA*)-*Biomembranes*, vol. 1858, no. 10, pp. 2529–2538, 2016.

[23] Z. Jia and B. Leimkuhler, “Molecular simulation in the canonical ensemble and beyond,” *ESAIM*: *Mathematical Modelling and Numerical Analysis*, vol. 41, no. 2, pp. 333–350, 2007.

[24] P. H. Hünenberger, “Thermostat algorithms for molecular dynamics simulations,” *Advanced computer simulation*: *Approaches for soft matter sciences I*, pp. 105–149, 2005.

[25] Y. Mori and Y. Okamoto, “Replica-exchange molecular dynamics simulations for various constant temperature algorithms,” *Journal of the Physical Society of Japan*, vol. 79, no. 7, Art. no. 074001, 2010.

[26] Z. Li, S. Xiong, C. Sievers, Y. Hu, Z. Fan, N. Wei, H. Bao, S. Chen, D. Donadio, and T. Ala-Nissila, “Influence of thermostatting on nonequilibrium molecular dynamics simulations of heat conduction in solids,” *The Journal of chemical physics*, vol. 151, no. 23, 2019.

[27] X. Yong and L. T. Zhang, “Thermostats and thermostat strategies for molecular dynamics simulations of nanofluidics,” *The Journal of chemical physics*, vol. 138, no. 8, 2013.

[28] G. Bussi, D. Donadio, and M. Parrinello, “Canonical sampling through velocity rescaling,” *The Journal of chemical physics*, vol. 126, no. 1, 2007.

[29] J. Hickman and Y. Mishin, “Temperature fluctuations in canonical systems: Insights from molecular dynamics simulations,” *Physical Review B*, vol. 94, no. 18, Art. no. 184311, 2016.

[30] W. G. Hoover, K. Aoki, C. G. Hoover, and S. V. De Groot, “Time-reversible deterministic thermostats,” *Physica D*: *Nonlinear Phenomena*, vol. 187, no. 1-4, pp. 253–267, 2004.

[31] M. E. Tuckerman, Y. Liu, G. Ciccotti, and G. J. Martyna, “Non-hamiltonian molecular dynamics: Generalizing hamiltonian phase space principles to non-hamiltonian systems,” *The Journal of Chemical Physics*, vol. 115, no. 4, pp. 1678–1702, 2001.

[32] M. Campisi, F. Zhan, P. Talkner, and P. Hänggi, “Reply to w. g. hoover [arxiv: 1204.0312 v2],” *arXiv preprint arXiv*:*1204.4412*, 2012.

[33] S. Queyroy, H. Nakamura, and I. Fukuda, “Numerical examination of the extended phase-space volume-preserving integrator by the nosé-hoover molecular dynamics equations,” *Journal of Computational Chemistry*, vol. 30, no. 12, pp. 1799–1815, 2009.

[34] W. G. Hoover, “Aspects of dynamical simulations, emphasizing nosé and nosé-hoover dynamics and the compressible baker map,” *arXiv preprint arXiv*:*1908.04379*, 2019.

[35] S. NOSÉ, “A molecular dynamics method for simulations in the,” *Journal Title*: *Molecular physics*, vol. 52, no. 2.

[36] W. G. Hoover, “Canonical dynamics: Equilibrium phase-space distributions,” *Phys*. *Rev*. *A*, vol. 31, no. 3, pp. 1695–1697, 1985.

[37] F. Legoll, M. Luskin, and R. Moeckel, “Non-ergodicity of the nosé–hoover thermostatted harmonic oscillator,” *Archive for rational mechanics and analysis*, vol. 184, no. 3, pp. 449–463, 2007.

[38] P. Langevin *et al*., “Sur la théorie du mouvement brownien,” *CR Acad*. *Sci*. *Paris*, vol. 146, no. 530-533, Art. no. 530, 1908.

[39] I. Prigogine and R. Balescu, “Sur la theorie moleculaire du mouvement brownien,” *Physica*, vol. 23, no. 1-5, pp. 555–568, 1957.

[40] B. Leimkuhler, E. Noorizadeh, and F. Theil, “A gentle stochastic thermostat for molecular dynamics,” *Journal of Statistical Physics*, vol. 135, no. 2, pp. 261–277, 2009.

[41] L. R. Bellet, “Ergodic properties of markov processes,” in *Open Quantum Systems II*: *The Markovian Approach*. Springer, 2006, pp. 1–39.

[42] G. A. Pavliotis, *Stochastic Processes and Applications*: *Diffusion Processes*, *the Fokker–Planck and Langevin Equations*, ser. Texts in Applied Mathematics. Springer, 2014, vol. 60.

[43] E. Weinan and J. C. Mattingly, “Ergodicity for the navier-stokes equation with degenerate random forcing: finite-dimensional approximation,” *Comm*. *Pure Appl*. *Math*, vol. 54, no. 11, pp. 1386–1402, 2001.

[44] R. Khasminskii, *Stochastic stability of differential equations*. Springer, 2012.

[45] S. P. Meyn and R. L. Tweedie, “Stability of markovian processes iii: Foster–lyapunov criteria for continuous-time processes,” *Advances in Applied Probability*, vol. 25, no. 3, pp. 518–548, 1993.

[46] M. Hairer and J. C. Mattingly, “Ergodicity of the 2d navier-stokes equations with degenerate stochastic forcing,” *Annals of Mathematics*, pp. 993–1032, 2006.

[47] E. B. Wilson, J. C. Decius, and P. C. Cross, *Molecular vibrations*: *the theory of infrared and Raman vibrational spectra*. Courier Corporation, 1980.

[48] S. Lang, *Linear algebra*. Springer Science & Business Media, 1987.

[49] G. Bussi and M. Parrinello, "Stochastic thermostats: comparison of local and global schemes," *Computer Physics Communications*, vol. 179, no. 1-3, pp. 26–29, 2008.

[50] M. Ceriotti, G. Bussi, and M. Parrinello, "Langevin equation with colored noise for constant-temperature<? format?> molecular dynamics simulations," *Physical review letters*, vol. 102, no. 2, Art. no. 020601, 2009.

[51] G. Grest, S. Nagel, A. Rahman, and T. Witten Jr, "Density of states and the velocity autocorrelation function derived from quench studies," *The Journal of Chemical Physics*, vol. 74, no. 6, pp. 3532–3534, 1981.

[52] A. Rahman, M. Mandell, and J. McTague, "Molecular dynamics study of an amorphous lennard-jones system at low temperature," *The Journal of Chemical Physics*, vol. 64, no. 4, pp. 1564–1568, 1976.

[53] D. Frenkel and B. Smit, *Understanding molecular simulation*: *from algorithms to applications*. Elsevier, 2023.

[54] A. Khintchine, "Korrelationstheorie der stationären stochastischen prozesse," *Mathematische Annalen*, vol. 109, no. 1, pp. 604–615, 1934.

[55] N. Wiener, "Generalized harmonic analysis," *Acta mathematica*, vol. 55, no. 1, pp. 117–258, 1930.

[56] W. Feller, *An introduction to probability theory and its applications*, *Volume 2*. John Wiley & Sons, 1991, vol. 2.

[57] P. Billingsley, *Probability and measure*. John Wiley & Sons, 2017.

[58] T. Schlick, *Molecular modeling and simulation*: *an interdisciplinary guide*. Springer, 2010, vol. 2.

[59] J. M. Haile, *Molecular Dynamics Simulation*: *Elementary Methods*. John Wiley & Sons, 1992.

[60] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," *Journal of computational chemistry*, vol. 25, no. 9, pp. 1157–1174, 2004.

[61] P. Hamm and M. Zanni, *Concepts and Methods of 2D Infrared Spectroscopy*. Cambridge University Press, 2011.

[62] J. L. McHale, *Molecular spectroscopy*. CRC Press, 2017.

[63] P. K. Mallick, *Fundamentals of Molecular Spectroscopy*. Springer, 2023.

[64] V. Vitale, J. Dziedzic, S. M.-M. Dubois, H. Fangohr, and C.-K. Skylaris, "Anharmonic infrared spectroscopy through the fourier transform of time correlation function formalism in onetep," *Journal of Chemical Theory and Computation*, vol. 11, no. 7, pp. 3321–3332, 2015.

[65] J. Guan *et al*., "Computational infrared and raman spectra by hybrid qm/mm techniques: a study on molecular and catalytic material systems," *Philosophical Transactions of the Royal Society A*, vol. 381, no. 2250, Art. no. 20220234, 2023.

[66] C. Banwell, *Fundamentals of Molecular Spectroscopy*, ser. European chemistry series. McGraw-Hill, 1972.

[67] D. C. Abbas, "Reduction of spurious baseline effects in nmr," *Review of Scientific Instruments*, vol. 50, no. 7, pp. 829–830, 1979.

[68] M. Sadeghi and F. Behnia, "Optimum window length of savitzky-golay filters with arbitrary order," *arXiv preprint arXiv*:*1808*.*10489*, 2018.

[69] M. Roy, V. R. Kumar, B. Kulkarni, J. Sanderson, M. Rhodes, and M. v. Stappen, “Simple denoising algorithm using wavelet transform,” *arXiv preprint nlin/0002028*, 2000.

[70] A. Savitzky and M. J. Golay, “Smoothing and differentiation of data by simplified least squares procedures.” *Analytical chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.

[71] A. Savitzky, “A historic collaboration,” *Analytical chemistry*, vol. 61, no. 15, pp. 921A–923A, 1989.

[72] E. Ostertagova and O. Ostertag, “Methodology and application of savitzky-golay moving average polynomial smoother,” *Global Journal of Pure and Applied Mathematics*, vol. 12, no. 4, pp. 3201–3210, 2016.

[73] M. Schmid, D. Rath, and U. Diebold, “Why and how savitzky–golay filters should be replaced,” *ACS Measurement Science Au*, vol. 2, no. 2, pp. 185–196, 2022.

[74] A. Lunardi, *Interpolation theory*. Springer, 2018, vol. 16.

[75] J. Bergh and J. Löfström, *Interpolation spaces*: *an introduction*. Springer Science & Business Media, 2012, vol. 223.

[76] P. C. Hansen, V. Pereyra, and G. Scherer, *Least squares data fitting with applications*. JHU Press, 2013.

[77] Z.-M. Zhang, S. Chen, and Y.-Z. Liang, “Baseline correction using adaptive iteratively reweighted penalized least squares,” *Analyst*, vol. 135, no. 5, pp. 1138–1146, 2010.

[78] S.-J. Baek, A. Park, Y.-J. Ahn, and J. Choo, “Baseline correction using asymmetrically reweighted penalized least squares smoothing,” *Analyst*, vol. 140, no. 1, pp. 250–257, 2015.

[79] A. Boziki, M. F. N., P. Fernandes, and A. Tkatchenko, “A journey with THeSeuSS: automated python tool for modeling IR and Raman vibrational spectra of molecules and solids,” vol. 15, no. 3, 2025.

[80] M. Fleischer and L. J. Cabri, “New mineral names,” *American Mineralogist*, vol. 61, no. 5-6, pp. 502–504, 06 1976.

[81] C. Milton, E. Dwornik, P. Estep-Barnes, R. Finkelman, A. Pabst, and S. Palmer, “Abelsonite, nickel porphyrin, a new mineral from the green river formation, utah,” *American Mineralogist*, vol. 63, no. 9-10, pp. 930–937, 1978.

[82] B. Lafuente, D. R. T., H. Yang, and N. Stone, *The power of databases*: *the RRUFF project*., T. Armbruster and R. M. Danisi, Eds. W. De Gruyter, 2015.

[83] D. C. Rapaport, *The art of molecular dynamics simulation*. Cambridge university press, 2004.

[84] E. Ditler and S. Luber, “Vibrational spectroscopy by means of first-principles molecular dynamics simulations,” *Wiley Interdisciplinary Reviews*: *Computational Molecular Science*, vol. 12, no. 5, Art. no. e1605, 2022.

[85] C. B. Moore, “Vibration→ vibration energy transfer,” *Advances in chemical physics*, pp. 41–83, 1973.

[86] G. S. Chirikjian and A. B. Kyatkin, *Engineering applications of noncommutative harmonic analysis*: *with emphasis on rotation and motion groups*. CRC press, 2000.

[87] Y. Katznelson, *An introduction to harmonic analysis*. Cambridge University Press, 2004.

[88] A. Adel and D. M. Dennison, “The infrared spectrum of carbon dioxide. part i,” *Physical Review*, vol. 43, no. 9, Art. no. 716, 1933.

[89] S. Civiš, M. Ferus, A. Knížek, S. Civiš, M. Ferus, and A. Knížek, “Carbon dioxide and the effects on climate,” *The Chemistry of CO2 and TiO2*: *From Breathing Minerals to Life on Mars*, pp. 1–7, 2019.

[90] H. C. Lau, U. Faul, J. X. Mitrovica, D. Al-Attar, J. Tromp, and G. Garapić, “Anelasticity across seismic to tidal timescales: a self-consistent approach,” *Geophysical Journal International*, Art. no. ggw401, 2016.

[91] K. Alvin, K. Park, and L. Peterson, “Extraction of undamped normal modes and nondiagonal modal damping matrix from damped system realization parameters,” in *34th Structures*, *Structural Dynamics and Materials Conference*, 1993, Art. no. 1653.

[92] K. Matsuo, Y. Miyazato, and H.-D. Kim, “Shock train and pseudo-shock phenomena in internal gas flows,” *Progress in aerospace sciences*, vol. 35, no. 1, pp. 33–100, 1999.

[93] F. Vanherpe, L. Olivas Duarte, and P. Lafon, “Sound vs. pseudo-sound contributions to the wind noise,” in *18th AIAA/CEAS Aeroacoustics Conference* (*33rd AIAA Aeroacoustics Conference*), 2012, Art. no. 2207.

[94] F. Michelot and J. Moret-Bailly, “Approche algébrique des spectres vibrationnels des molécules polyatomiques,” *Journal de physique*, vol. 48, no. 1, pp. 51–72, 1987.

[95] T. J. Macke and D. A. Case, “Modeling unusual nucleic acid structures.” ACS Publications, 1998.

[96] D. A. Case *et al*., *Amber 2023*. University of California, San Francisco, 2023.

[97] A. Molina, P. Smereka, and P. M. Zimmerman, “Exploring the relationship between vibrational mode locality and coupling using constrained optimization,” *The Journal of Chemical Physics*, vol. 144, no. 12, 2016.

[98] A. Semparithi and S. Keshavamurthy, “Intramolecular vibrational energy redistribution as state space diffusion: Classical-quantum correspondence,” *The Journal of chemical physics*, vol. 125, no. 14, 2006.

[99] M. Hassani, P. H. Shaon, C. J. Mallon, T. Shi, J. N. Monzy, E. E. Fenlon, D. M. Leitner, and M. J. Tucker, “Modulating vibrational energy redistribution in highly conjugated systems,” *The Journal of Chemical Physics*, vol. 162, no. 15, 2025.

[100] S. Karmakar, P. K. Yadav, and S. Keshavamurthy, “Stable chaos and delayed onset of statisticality in unimolecular dissociation reactions,” *Communications Chemistry*, vol. 3, no. 1, Art. no. 4, 2020.

[101] C. Froeschlé, M. Guzzo, and E. Lega, “Graphical evolution of the arnold web: from order to chaos,” *Science*, vol. 289, no. 5487, pp. 2108–2110, 2000.

[102] K. N. Swamy and W. L. Hase, “Search for quasiperiodic motion in vibrationally excited formaldehyde formed by s1→ so internal conversion,” *Chemical Physics Letters*, vol. 92, no. 4, pp. 371–378, 1982.

[103] A. Bach, J. M. Hostettler, and P. Chen, “Quasiperiodic trajectories in the unimolecular dissociation of ethyl radicals by time-frequency analysis,” *The Journal of chemical physics*, vol. 123, no. 2, 2005.

[104] L. Casetti, M. Pettini, and E. Cohen, “Geometric approach to hamiltonian dynamics and statistical mechanics,” *Physics Reports*, vol. 337, no. 3, pp. 237–341, 2000.

[105] S. C. Farantos, “Chemical dynamics: A periodic orbits approach,” *Time-Dependent Quantum Molecular Dynamics*, pp. 27–43, 1992.

[106] S. C. Farantos, “Exploring molecular vibrational motions with periodic orbits,” *International Reviews in Physical Chemistry*, vol. 15, no. 2, pp. 345–374, 1996.

[107] A. Karamatskou and H. Kleinert, “Geometrization of the schrödinger equation: Application of the maupertuis principle to quantum mechanics,” *International Journal of Geometric Methods in Modern Physics*, vol. 11, no. 08, Art. no. 1450066, 2014.

[108] P. Maraner, “On the jacobi metric for a general lagrangian system,” *Journal of Mathematical Physics*, vol. 60, no. 11, Art. no. 112901, 11 2019.

[109] J. Jost and J. Jost, *Riemannian geometry and geometric analysis*. Springer, 2008, vol. 42005.

[110] I. Horenko and M. Weiser, “Adaptive integration of molecular dynamics,” *Journal of computational chemistry*, vol. 24, no. 15, pp. 1921–1929, 2003.

[111] L. Spörkel and W. Thiel, “Adaptive time steps in trajectory surface hopping simulations,” *The Journal of Chemical Physics*, vol. 144, no. 19, 2016.

[112] H. C. Andersen, “Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations,” *Journal of computational Physics*, vol. 52, no. 1, pp. 24–34, 1983.

[113] L. Verlet, “Computer” experiments” on classical fluids. i. thermodynamical properties of lennard-jones molecules,” *Physical review*, vol. 159, no. 1, Art. no. 98, 1967.

[114] G. Dahlquist and Å. Björk, “Numerical methods,” 1974.

[115] R. Elber, A. P. Ruymgaart, and B. Hess, “Shake parallelization,” *The European Physical Journal Special Topics*, vol. 200, no. 1, pp. 211–223, 2011.

[116] R. Montgomery *et al*., "Who's afraid of the hill boundary?" *SIGMA. Symmetry, Integrability and Geometry: Methods and Applications*, vol. 10, Art. no. 101, 2014.

[117] L. Di Cairano, M. Gori, and M. Pettini, "Coherent riemannian-geometric description of hamiltonian order and chaos with jacobi metric," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 29, no. 12, 2019.

[118] H. Seifert, "Periodische bewegungen mechanischer systeme," *Mathematische Zeitschrift*, vol. 51, no. 2, pp. 197–216, 1948.

[119] B.-L. Hu and Y. Zhang, "Uncertainty relation at finite temperature," *Physical Review Letters*, 2993.

[120] D. A. McQuarrie, *Statistical Mechanics*. Harper & Row, 2000.

[121] F. Tama and Y.-H. Sanejouand, "Conformational change of proteins arising from normal mode calculations," *Protein engineering*, vol. 14, no. 1, pp. 1–6, 2001.

[122] A. Amadei, A. Linssen, B. L. de Groot, D. Van Aalten, and H. Berendsen, "An efficient method for sampling the essential subspace of proteins," *Journal of Biomolecular Structure and Dynamics*, vol. 13, no. 4, pp. 615–625, 1996.

[123] T. Keyes, "Instantaneous normal mode approach to liquid state dynamics," *The Journal of Physical Chemistry A*, vol. 101, no. 16, pp. 2921–2930, 1997.

[124] B. Chow and D. Knopf, *The Ricci Flow: An Introduction: An Introduction*. American Mathematical Soc., 2004, vol. 1.

[125] M. P. Do Carmo, *Differential geometry of curves and surfaces: revised and updated second edition*. Courier Dover Publications, 2016.

[126] K. S. Thorne, C. W. Misner, and J. A. Wheeler, *Gravitation*. Freeman San Francisco, 2000.

[127] L. Di Cairano, M. Gori, G. Pettini, and M. Pettini, "Hamiltonian chaos and differential geometry of configuration space–time," *Physica D*: *Nonlinear Phenomena*, vol. 422, Art. no. 132909, 2021.

[128] Y. Saad, *Numerical methods for large eigenvalue problems*: *revised edition*. SIAM, 2011.

[129] P. Lindqvist, "A note on the nonlinear rayleigh quotient," *LECTURE NOTES IN PURE AND APPLIED MATHEMATICS*, pp. 223–223, 1994.

[130] Y. Saad, *Iterative methods for sparse linear systems*. SIAM, 2003.

[131] E. R. Davidson, "Theiterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices," *J. Comput. Phys*, vol. 17, pp. 87–94, 1975.

[132] A. Guichardet, "On rotation and vibration motions of molecules," in *Annales de l'IHP Physique théorique*, vol. 40, no. 3, 1984, pp. 329–342.

[133] J. E. Marsden, R. Montgomery, and T. S. Rațiu, *Reduction*, *symmetry*, *and phases in mechanics*. American Mathematical Soc., 1990, vol. 436.

[134] J. H. Hannay, "Angle variable holonomy in adiabatic excursion of an integrable hamiltonian," *Journal of Physics A*: *Mathematical and General*, vol. 18, no. 2, Art. no. 221, 1985.

[135] S. C. Farantos, "Periodic orbits in biological molecules: Phase space structures and selectivity in alanine dipeptide," *The Journal of chemical physics*, vol. 126, no. 17, 2007.

[136] W. H. Press, *Numerical recipes 3rd edition*: *The art of scientific computing*. Cambridge university press, 2007.

[137] E. Fehlberg, *Low-order classical Runge-Kutta formulas with stepsize control and their application to some heat transfer problems*. National aeronautics and space administration, 1969, vol. 315.

[138] A. Iserles, H. Z. Munthe-Kaas, S. P. Nørsett, and A. Zanna, "Lie-group methods," *Acta numerica*, vol. 9, pp. 215–365, 2000.

[139] J. Q. Gallier and J. Quaintance, *Differential geometry and lie groups*. Springer, 2020, vol. 12.

[140] A. De Paris and A. M. Vinogradov, *Fat manifolds and linear connections*. World Scientific, 2009.

[141] B. O'Neill, "The fundamental equations of a submersion." *Michigan Mathematical Journal*, vol. 13, no. 4, pp. 459–469, 1966.

[142] N. Balakrishnan and V. B. Nevzorov, *A primer on statistical distributions*. John Wiley & Sons, 2004.

[143] C. Forbes, M. Evans, N. Hastings, and B. Peacock, *Statistical distributions*. John Wiley & Sons, 2011.

[144] A. Pick, B. Zhen, O. D. Miller, C. W. Hsu, F. Hernandez, A. W. Rodriguez, M. Soljačić, and S. G. Johnson, "General theory of spontaneous emission near exceptional points," *Optics express*, vol. 25, no. 11, pp. 12 325–12 348, 2017.

[145] I. Shomroni, O. Bechler, S. Rosenblum, and B. Dayan, "Demonstration of weak measurement based on atomic spontaneous emission," *Physical review letters*, vol. 111, no. 2, Art. no. 023604, 2013.

[146] M. C. Newstein, "Spontaneous emission in the presence of a prescribed classical field," *Physical Review*, vol. 167, no. 1, Art. no. 89, 1968.

[147] G. Pagnini and R. Saxena, "A note on the voigt profile function," *arXiv preprint arXiv*:*0805*.*2274*, 2008.

[148] H. O. Di Rocco, D. I. Iriarte, and J. Pomarico, “General expression for the voigt function that is of special interest for applied spectroscopy,” *Applied Spectroscopy*, vol. 55, no. 7, pp. 822–826, 2001.

[149] C. R. Henderson, “Best linear unbiased estimation and prediction under a selection model,” *Biometrics*, pp. 423–447, 1975.

[150] C. Chatfield, *Introduction to multivariate analysis*. Routledge, 2018.

[151] A. C. Aitken, “Iv.—on least squares and linear combination of observations,” *Proceedings of the Royal Society of Edinburgh*, vol. 55, pp. 42–48, 1936.

[152] B. H. Baltagi and B. H. Baltagi, *Econometric analysis of panel data*. Springer, 2008, vol. 4.

[153] T. D. Stanley and H. Doucouliagos, “Neither fixed nor random: weighted least squares meta-analysis,” *Statistics in medicine*, vol. 34, no. 13, pp. 2116–2127, 2015.

[154] Q. Geng, X. Hao, and Y. Wang, “Forecasting the volatility of crude oil futures: A time-dependent weighted least squares with regularization constraint,” *Journal of Forecasting*, vol. 43, no. 2, pp. 309–325, 2024.

[155] Y. Zhou and X. Fang, “A mixed weighted least squares and weighted total least squares adjustment method and its geodetic applications,” *Survey review*, vol. 48, no. 351, pp. 421–429, 2016.

[156] M. Khatibinia, M. J. Fadaee, J. Salajegheh, and E. Salajegheh, “Seismic reliability assessment of rc structures including soil–structure interaction using wavelet weighted least squares support vector machine,” *Reliability Engineering & System Safety*, vol. 110, pp. 22–33, 2013.

[157] H. White, “A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity,” *Econometrica*: *Journal of the Econometric Society*, pp. 817–838, 1980.

[158] L. Röseler *et al*., “The replication database: Documenting the replicability of psychological science,” *Journal of Open Psychology Data*, vol. 12, no. 1, 2024.

[159] G. S. Kimeldorf and G. Wahba, “A correspondence between bayesian estimation on stochastic processes and smoothing by splines,” *The Annals of Mathematical Statistics*, vol. 41, no. 2, pp. 495–502, 1970.

[160] G. Kimeldorf and G. Wahba, “Some results on tchebycheffian spline functions,” *Journal of mathematical analysis and applications*, vol. 33, no. 1, pp. 82–95, 1971.

[161] Y.-J. Kim, “Smoothing spline regression: Scalable computation and cross validation,” Ph.D. dissertation, Purdue University, 2003.

[162] C. Gu, D. M. Bates, Z. Chen, and G. Wahba, “The computation of generalized cross-validation functions through householder tridiagonalization with applications to the fitting of interaction spline models,” *SIAM Journal on Matrix Analysis and Applications*, vol. 10, no. 4, pp. 457–480, 1989.

APPENDIX A

Useful distributions

Three commonly-used distributions in spectroscopic signal processing are introduced in this chapter: Gaussian, Lorentzian and Voigt distributions.

A.1 Gaussian

A Gaussian, or a normal, distribution with mean μ and standard deviation σ has its probability density function defined as below.

$$G(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{A.1}$$

This distribution is normally denoted as $\mathcal{N}(\mu,\sigma^2)$.

A.2 Lorentzian

The probability density function of a Lorentzian, or a Cauchy, distribution is given as

$$C(x;x_0,\gamma) = \frac{1}{\pi}\frac{\gamma}{(x-x_0)^2+\gamma^2} \tag{A.2}$$

where x_0 is the median and γ is the median absolute deviation. [142] One of the key characteristics of this distribution is that it does not have a mean. [143]

A.3 Voigt

Spontaneous emission of fermionic system in vacuum yields a spectrum that can be described as a sum of Lorentzian profiles. [144–146] However, atomic

motions and inhomogeneous fields broadens the spectral profiles making the signal to look more like a sum of Gaussians. To capture both of these physical effects simultaneously, physicists use what is known as a Voigt profile (or Voigt distribution), which combines the characteristics of both Gaussian and Lorentzian distributions.

A Voigt distribution is defined via the following expression:

$$V(x;\mu,x_0,\sigma,\gamma)=\int_{-\infty}^{\infty}G(y;\sigma)C(x-y;x_0,\gamma)dy. \tag{A.3}$$

Here, the variable x is normalised using the Gaussian width $\omega_G=\sqrt{2}\sigma$ so that the parameters can be omitted.

Using the dimensionless variable $x'=x/\omega_G$, the Voigt function can be expressed in a simplified form:

$$V(x)=\frac{1}{\sqrt{\pi}\omega_G}K(x',y), \tag{A.4}$$

where $y=\omega_L/\omega_G$ and the kernel function $K(x,y)$ is given by the following expression. [147]

$$K(x,y)=\frac{y}{\pi}\int_{-\infty}^{+\infty}\frac{e^{-\xi^2}}{(x-\xi)^2+y^2}d\xi \tag{A.5}$$

This kernel function can be alternatively represented as:

$$K(x,y)=\frac{1}{\sqrt{\pi}}\int_{0}^{+\infty}e^{-y\xi-\xi^2/4}\cos(x\xi)d\xi \tag{A.6}$$

which can be re-written in computable form as

$$K(x,y)=\sum_{n=0}^{\infty}(-1)^n\left\{\frac{1}{\Gamma(n+1)}{}_1F_1\left(\frac{2n+1}{2},\frac{1}{2};y^2\right)-\frac{2a}{\Gamma(\frac{2n+1}{2})}{}_1F_1\left(n+1,\frac{3}{2};y^2\right)\right\}x^{2n}, \tag{A.7}$$

where ${}_1F_1(\alpha,\beta;z)$ denotes the confluent hypergeometric function. [148]

APPENDIX B

Least squares algorithm

The least squares method is a standard approach to approximate solutions to overdetermined systems, i.e. systems with the number of equations exceeding the number of unknowns. Essentially, the method is interested in solving systems of linear equations of the form

$$Ax = b \tag{B.1}$$

where $A \in \mathbb{R}^{m \times n}$ is the matrix determining the nature of the problem, $x \in \mathbb{R}^n$ is the vector of unknowns, and $b \in \mathbb{R}^m$ is the data observed for $m, n \in \mathbb{N}$ with $m > n$. The least squares problems are defined as in Definition B.1

Definition B.1 (Least squares method)**.** Given a linear system of the form given in Equation B.1, the residual is defined as $r = Ax - b$. The *least squares solution* x^* is the solution that satisfies

$$x^* = \min_x \|Ax - b\|^2 = \min_x \|r(x)\|^2 \tag{B.2}$$

where $\|x\| = \sqrt{x \cdot x}$ is the Euclidean L^2 norm. Finding this x^* is the *least squares problem* corresponding to the given system, and the parameters encoded in x^* are called the *estimator* of the problem. Differentiating $\|r(x)\|^2$ with respect to x and setting the result to zero gives an equation called *normal equation* that gives x^* once solved.

If the matrix $A^T A$ is invertible, the solution can be found immediately as

$$A^T Ax = A^T b \implies x^* = \left(A^T A\right)^{-1} A^T b \tag{B.3}$$

and the parameters encoded in x^* are called the *ordinary least squares estimator*.

B.1 Simple and weighted linear regressions

Suppose a set of data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ is given, and want to fit a line $y = mx + c$. The residue of the problem, according to Definition B.1, is $r = y - (mx + c)$. Hence, the least squares problem becomes the optimisation problem of the following:

$$\|r\|^2 = (y - (mx + c)) \cdot (y - (mx + c)) \tag{B.4}$$

$$= y \cdot y - 2my \cdot x - 2cy \cdot y + m^2 x \cdot x + 2cmx \cdot 1 + 2 + c^2, \tag{B.5}$$

then asserting $\partial\|r\|/\partial c = 0$ and $\partial\|r\|/\partial m = 0$ gives

$$c = \frac{\sum_i^n y_i \sum_i^n {x_i}^2 - \sum_i^n x_i y_i \sum_i^n x_i}{n \sum_i^n {x_i}^2 - \left(\sum_i^n x_i\right)^2} \tag{B.6}$$

$$m = \frac{n \sum_i^n x_i y_i - \sum_i^n y_i \sum_i^n x_i}{n \sum_i^n x_i^2 - \left(\sum_i^n x_i\right)^2}. \tag{B.7}$$

The parameters m and c provide the linear estimator for the data given by the model $\hat{y} = mx + c$, which is unbiased and minimises the root mean squared error (RMSE), where RMSE is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} = \sqrt{r \cdot r} \tag{B.8}$$

for $\hat{y}_i = mx_i + c$. Note that the unbiasedness is true since the mean of $\hat{y}$ equals the mean of y by construction. [149]

The data are not always evenly distributed with respect to the x_i-values. When the data points are not uniformly spread, it may be beneficial to introduce a set of weights w_i for each datapoint (x_i, y_i). These weights account for the varying importance or reliability of different datapoints. In this case, the least-squares problem is modified to minimize the weighted sum of squared residuals, given by: [150]

$$J = \sum_{i=1}^{n} w_i (y_i - c - mx_i)^2. \tag{B.9}$$

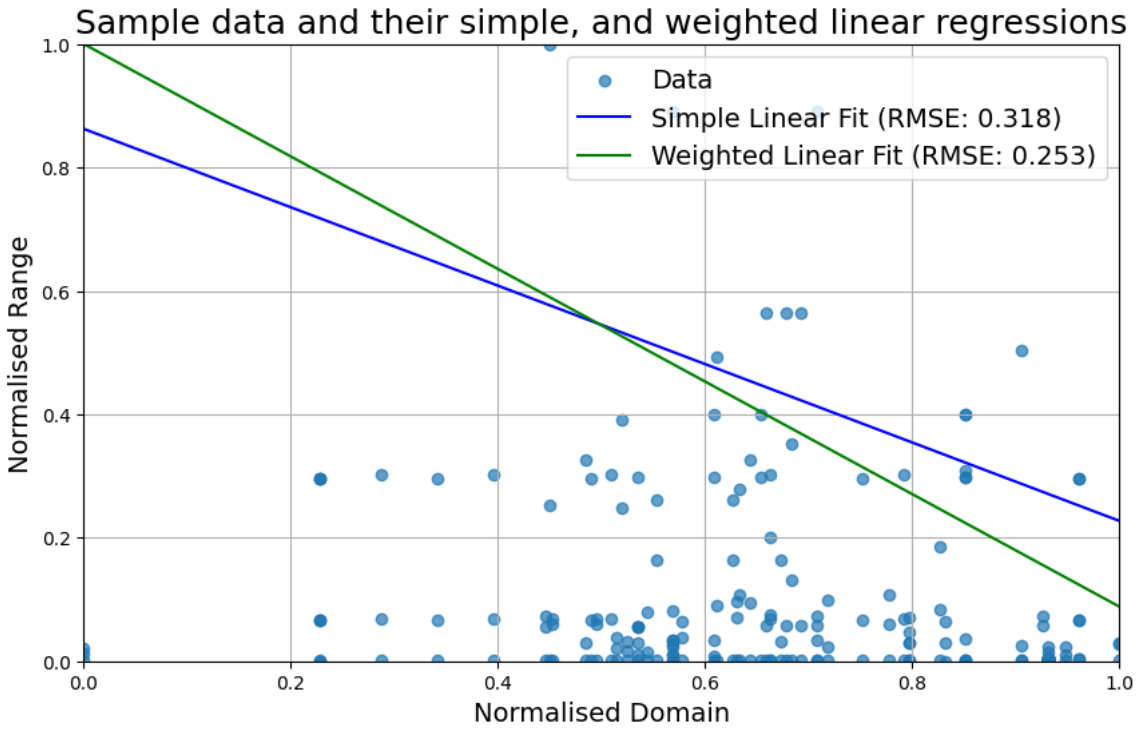


Figure B.1. *Comparison of simple and weighted least squares regression fits on psychology research data. The simple linear fit (blue) results in a higher RMSE of 0.318, while the weighted linear fit (green), which accounts for the variable residuals, achieves a lower RMSE of 0.253. The scatter plot shows data points normalized between 0 and 1 for both the domain and range.*

The fitting parameters minimizing this quantity are given as

$$c = \frac{\sum_i^n w_i y_i \sum_i^n w_i {x_i}^2 - \sum_i^n w_i x_i \sum_i^n w_i x_i y_i}{\sum_i^n w_i \sum_i^n w_i x_i^2 - \left(\sum_i^n x_i\right)^2}, \tag{B.10}$$

$$m = \frac{\sum_i^n w_i \sum_i^n w_i x_i y_i - \sum_i^n w_i y_i \sum_i^n w_i x_i}{\sum_i^n w_i \sum_i^n w_i x_i^2 - \left(\sum_i^n x_i\right)^2}. \tag{B.11}$$

One of the most intuitive weights that corrects the statistical deviations is

$$w_i = \frac{1}{{\sigma_i}^2} \tag{B.12}$$

as this ensures that the fit is unbiased and RMSE is minimized. [151]

However, the values of σ_i are typically unknown to the experimenter since the residuals may exhibit varying levels of spread. [152] This requires the use

of problem-specific weight models to address the differing variance across observations. [153–156]

One of the σ_i values robust of the varying levels of residuals can be defined iteratively via the following equation. [157]

$$\sigma_i{}^2 = (y_i - \hat{y}_i)^2 = r \cdot r \tag{B.13}$$

The Figure B.1 shows the performance of simple and weighed least-squared fittings of a data from a psychology research work. [158]

B.2 Penalised least squares

The methods introduced in Section 3.2, arPLS and airPLS, are variants of what is known as *pernalised least squares* (PLS) method, which minimizes the penalised least-squared functional

$$J^{\text{PLS}} = \sum_{i=1}^{n} w_i (y_i - f(x_i))^2 + \lambda \rho[f] \tag{B.14}$$

for $\lambda \in \mathbb{R}$, the penalty parameter, and a quadratic functional ρ determining the local curvature profiles of the function f in the given domain. This method was first made to understand how to fit with a polynomial or spline functions using the least-squares methods by George Kimeldorf and Grace Wahba in 1970 [159, 160], but its robustness and scalability. For instance, the cubic smoothing for univariate regressions the functional ρ can take the following form:

$$\rho[f] = \int_a^b \left(\frac{d^2 f}{dx^2} \right)^2 dx \tag{B.15}$$

where (a, b) is the interval containing all given datapoints. [161] The penalty parameter λ can be either empirically chosen, or optimized via likelihood functions. [162]

The influence of the penalty parameter λ is demonstrated in Figure B.2, which shows various PLS fits to a noisy dataset using different values of λ.

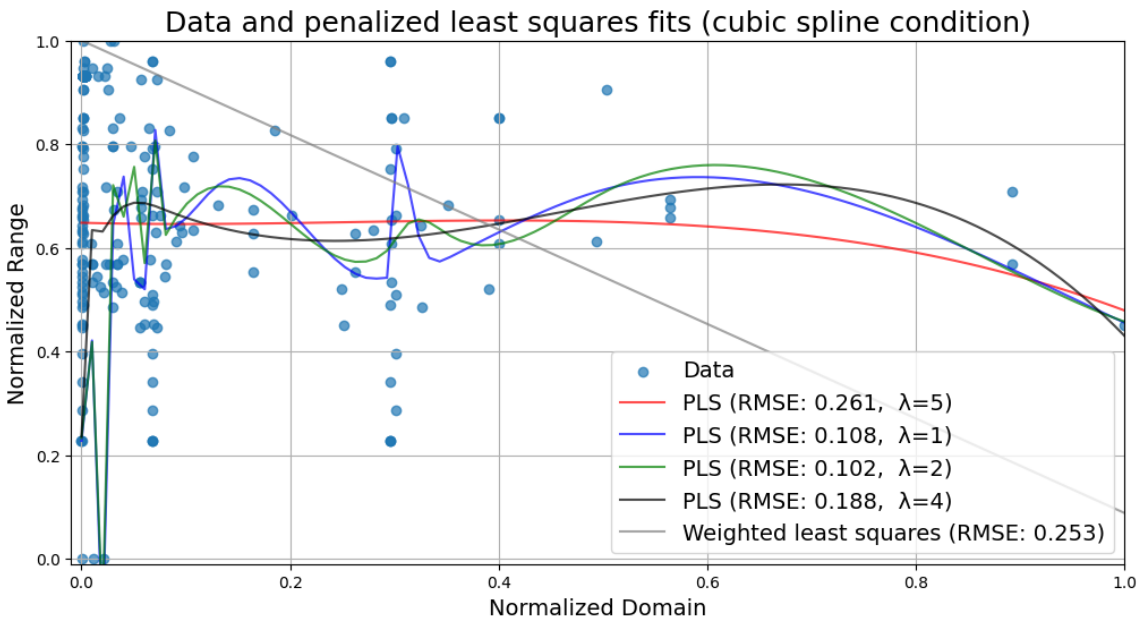


Figure B.2. *Comparison of penalized least squares (PLS) fits with different penalty parameters (λ) applied to noisy data. The fits demonstrate the trade-off between smoothness and accuracy: smaller values of λ (1,2) produce more flexible fits with lower RMSE but may overfit local variations, while larger values (4,5) yield smoother curves at the cost of higher RMSE. A standard weighted least squares fit is shown for comparison. All fits use a cubic spline condition for the penalty functional.*

When $\lambda = 1$ or $\lambda = 2$, the resulting fits (blue and green curves) closely follow the local variations in the data, achieving lower RMSE values of approximately 0.108 and 0.102, respectively. However, these fits may be capturing noise rather than the underlying trend. Increasing λ to 4 or 5 (black and red curves) produces progressively smoother fits that are less sensitive to local fluctuations, though at the cost of higher RMSE values (0.188 and 0.261).

For comparison, a standard weighted least squares fit (gray curve) is shown, which exhibits less control over local smoothness and results in an RMSE of 0.253. This highlights a key advantage of the PLS method: by adjusting λ, one can explicitly balance the trade-off between fidelity to the data (minimizing

residuals) and smoothness of the solution (minimizing the penalty term). This flexibility makes PLS particularly suitable for baseline correction problems, where the goal is to capture smooth underlying trends while being robust to noise and outliers.